Original Paper

# TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos

Wenwen Kong[1*], MD; Shijie Song[2*], MSc; Yuxiang Chris Zhao[3], PhD; Qinghua Zhu[2], PhD; Ling Sha[4], MD

[1]Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China
[2]School of Information Management, Nanjing University, Nanjing, China
[3]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China
[4]Nanjing Jiangning Maternal and Child Health and Family Planning Service Center, Nanjing, China
[*]these authors contributed equally

**Corresponding Author:**
Shijie Song, MSc
School of Information Management
Nanjing University
Xianlin Ave 163
Nanjing, 210023
China
Phone: 1 15951973800
Email: ssong@smail.nju.edu.cn

## Abstract

**Background:** Diabetes has become one of the most prevalent chronic diseases, and many people living with diabetes use social media to seek health information. Recently, an emerging social media app, TikTok, has received much interest owing to its popularity among general health consumers. We notice that there are many videos about diabetes on TikTok. However, it remains unclear whether the information in these videos is of satisfactory quality.

**Objective:** This study aimed to assess the quality of the information in diabetes-related videos on TikTok.

**Methods:** We collected a sample of 199 diabetes-related videos in Chinese. The basic information presented in the videos was coded and analyzed. First, we identified the source of each video. Next, 2 independent raters assessed each video in terms of the completeness of six types of content (the definition of the disease, symptoms, risk factors, evaluation, management, and outcomes). Then, the 2 raters independently assessed the quality of information in the videos, using the DISCERN instrument.

**Results:** In regard to the sources of the videos, we found 6 distinct types of uploaders; these included 3 kinds of individual users (ie, health professionals, general users, and science communicators) and 3 types of organizational users (ie, news agencies, nonprofit organizations, and for-profit organizations). Regarding content, our results show that the videos were primarily about diabetes management and contained limited information on the definition of the disease, symptoms, risk factors, evaluation, and outcomes. The overall quality of the videos was acceptable, on average, although the quality of the information varied, depending on the sources. The videos created by nonprofit organizations had the highest information quality, while the videos contributed by for-profit organizations had the lowest information quality.

**Conclusions:** Although the overall quality of the information in the diabetes videos on TikTok is acceptable, TikTok might not fully meet the health information needs of patients with diabetes, and they should exercise caution when using TikTok as a source of diabetes-related information.

*(J Med Internet Res 2021;23(9):e30409)* doi: 10.2196/30409

**KEYWORDS**

diabetes; information quality; infodemiology; social media; short video apps; TikTok

## Introduction

Diabetes has become one of the most prevalent chronic diseases throughout the world. According to a recent report by the International Diabetes Federation, the estimated global prevalence of diabetes in people aged 20-79 years reached 493 million in 2019, accounting for 9.3% of the total world population [1]. Crude estimates of diabetes prevalence are 13.0%

for US adults [2] and 12.8% for Chinese adults [3]. This high prevalence of diabetes results in huge financial burdens and losses for societies. In 2019, diabetes and related complications led to approximately 4.2 million deaths globally and resulted in US $760 billion of health expenditures [1]. Therefore, there is a pressing need to take action in managing diabetes.

Individuals living with diabetes can actively manage their chronic condition. Early studies suggest that intensive blood glucose control can greatly reduce the risk of complications from the disease [4-6]. However, effective blood glucose control is not an easy task among people living with diabetes. Adequate glycemic controls require a constellation of actions, such as a customized diet, exercise plans, regular self-assessments of blood glucose levels, and optimized medication [7]. According to one study, in 2013, only 25.8% of patients with diabetes had received treatment in China, and only 39.7% of those treated had adequate blood glucose control [8]; such low treatment and adherence rates may be associated with people's limited knowledge of the disease [9]. People living with diabetes often have diverse needs for information regarding their chronic condition, such as basic information on diabetes and the effectiveness of treatment options, on the sequelae of diabetes, blood glucose control, etc [10]. Nevertheless, they usually encounter many difficulties finding relevant and easy-to-understand information on their conditions [11].

Emerging internet technologies provide opportunities for better health communication and patient education. The internet has shifted the role of patients from passive information recipients to active information seekers [12]. General health information consumers use social media platforms (eg, discussion forums, microblogs, and group chatting) to seek both instrumental advice and emotional support [13-15]. Patients with diabetes who actively use social media for information, according to recent evidence, are associated with having lower glycated hemoglobin values [16]. A possible explanation for this is that social media provides patients with many opportunities to gain health knowledge, thereby increasing patient activation (ie, the ability and willingness that equip patients to take active action in managing their health care) [17]. Therefore, it is essential to utilize social media for better health communication for managing diabetes conditions.

Despite the considerable benefits of social media, its use for health communication has some limitations. In the literature, the quality of the information is the most extensively mentioned concern [18]. The possibility of encountering faulty health information on social media increases risks for patients, who may make health decisions on the basis of inaccurate information [19]. The quality of unmoderated information poses challenges for both patients and health care providers. On one hand, patients need to be able to distinguish high-quality information sources from low-quality ones [20,21]; one the other, health professionals and institutions are expected to respond to and combat health misinformation to protect the public [22]. Therefore, it is important to examine the quality of health information on social media.

Recently, an emerging short video app, TikTok, has attracted the interest of health care researchers [23]. During the COVID-19 pandemic, health-related videos on TikTok were widely viewed and shared. For example, COVID-19–related videos on the app have been watched approximately 93.1 billion times [24]. TikTok contains many videos about diabetes; however, their quality remains unstudied. To address this gap, this study aimed to systematically assess the quality of the information in diabetes-related videos on TikTok.

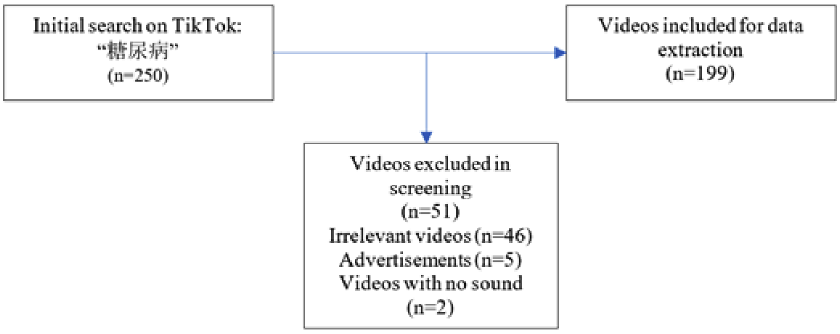
# Methods

## Search Strategy and Data Extraction

Using the keyword "糖尿病" ("diabetes" in Chinese), we searched TikTok during the period from January 20-25, 2021, and we retrieved the first 250 videos delivered by TikTok's recommended sorting process. We included the videos directly related to diabetes and excluded videos on other topics, commercial advertisements, and videos with no sound. After the screening, we obtained 199 videos for further data extraction and analysis (Figure 1).

We retrieved and extracted the basic information for each video, including the URL, publication date, name of the uploader, type of uploader (individual vs organization), uploaders' verification status, length of the video, number of times it was shared, and number of "likes" and comments it received. The extracted data were recorded in Excel (Microsoft Inc).

**Figure 1.** Video screening procedure.



## Measures

We measured 2 aspects of diabetes-related videos on TikTok: their content and the quality of their information. First, we adopted the coding schema proposed by Goobie et al [25] to rate the quality of six types of content: the definition of the disease, signs/symptoms, risk factors, evaluation, management, and outcomes. Two raters assessed each video independently and scored how sufficiently the video addressed each of the content types on a 3-item scale: 0 points (no content), 1 point (some content), 2 points (extensive content).

To rate the quality of the information, we adopted the DISCERN instrument. According to a systematic review [26], since its publication in 1998, DISCERN has been one of the most widely adopted instruments for assessing the quality of health information. It is a brief questionnaire that enables its users to assess the quality of health information concerning treatment choices [27]. The instrument consists of 16 questions, with response choices based on a 5-point scale, ranging from 1=poor to 5=good. These 16 questions are divided into 3 sections. The first 8 concern the reliability of the publication, such as whether its aims are clear and whether it is relevant, balanced, and unbiased. The scores for this section indicate whether the publication can be trusted as a source of information for choosing a treatment for a particular disease. The 7 questions in the second section focus on the details of treatment choices, such as whether the publication describes how each treatment works and explains its risks and benefits. The scores for this section reflect the quality of the publication's information about treatment choices (including self-care). The third section consists of 1 final question, based on all the previous ones; it asks users to rate the overall quality of the publication as a source of information about treatment choices. Of note is the fact that, although the original DISCERN instrument was designed for evaluating written publications, it has been widely used for assessing health-related videos. For example, researchers have used it to evaluate YouTube videos informing patients about treatments for cancer [28,29] and diabetes [30].

### Rating process

Two authors (WK and LS) worked on the rating tasks; both are certified physicians who work at endocrinology departments at 2 local hospitals. The raters independently scored each video for its coverage of the 6 types of content and applied the 16 questions of the DISCERN instrument. Interrater reliability was assessed with SPSS (version 22, IBM Corp). The interrater reliability for each of the 6 items relating to video content ranged from 0.813 to 0.981, and all of the reliability coefficients are highly significant at an error margin of 0.1%. The interrater reliability for each of the 16 items of the DISCERN instrument ranges from 0.898 to 0.991, and all of the reliability coefficients are highly significant at an error margin of 0.1%. These results indicate satisfactory interrater reliability.

## Results

### Video Sources

We identified 2 primary sources of the videos: individual and organizational users. We further identified 3 types of video creators among individual users: health professionals, science communicators, and general users. Among organizational users, we identified three types of sources: news agencies, nonprofit organizations, and for-profit organizations.

The results suggest that individual users published most of the videos (n=156, 78.4%). Among individual users, health professionals contributed the most videos (n=138, 69.3%), followed by general users (n=12, 6.0%), and science communicators (n=6, 3.0%). We noted that only 43 videos were uploaded by organizational users, and these accounted for 21.6% of the videos in our sample. Among organizational users, news agencies contributed the most videos (n=31, 15.6%), followed by nonprofit organizations (n=7, 3.5%) and for-profit organizations (n=5, 2.5%) (Table 1).

In our sample, the shortest video lasts only 13 seconds, while the longest lasts 407 seconds. On average, the videos are approximately 1 minute long. All videos were published after 2019. The earliest video had been on TikTok for 589 days, while the latest one was published 3 days prior to the day we collected the data. The videos in the sample received 2.75 million "likes" and 157,700 comments and were shared 305,200 times. Table 2 shows the characteristics of the videos, described by the median numbers across different sources.

**Table 1.** Descriptions of video sources.

| Source type | Source description | Videos, n (%) |
|---|---|---|
| **Individual users (n=156)** | | |
| Health professionals | Individuals who identify themselves as health professionals (eg, doctors and nurses) | 138 (69.3) |
| General users | General users (eg, general health consumers) | 12 (6.0) |
| Science communicators | Individuals who are engaged in scientific communication (eg, popular science writers) | 6 (3.0) |
| **Organizational users (n=43)** | | |
| News agencies | News agencies and the press | 31 (15.6) |
| Nonprofit organizations | Organizations operated for collective, public, or social benefit and public hospitals | 7 (3.5) |
| For-profit organizations | Organizations that pursue commercial interests | 5 (2.5) |

**Table 2.** Characteristics of the videos across sources (median numbers).

| Source type | Length of video (seconds), median | Days on TikTok, median | "Likes," median | Comments, median | Times shared, median |
|---|---|---|---|---|---|
| Health professionals (n=138) | 50.5 | 131 | 4191 | 164 | 802.5 |
| General users (n=12) | 135.5 | 161 | 3089.5 | 78.5 | 891 |
| Science communicators (n=6) | 43 | 135 | 24,000 | 593 | 4966 |
| News agencies (n=31) | 43 | 90 | 9917 | 167 | 3505 |
| Nonprofit organizations (n=7) | 52 | 200 | 23,000 | 194 | 3023 |
| For-profit organizations (n=5) | 63 | 317 | 33,000 | 405 | 6759 |

## Video Content

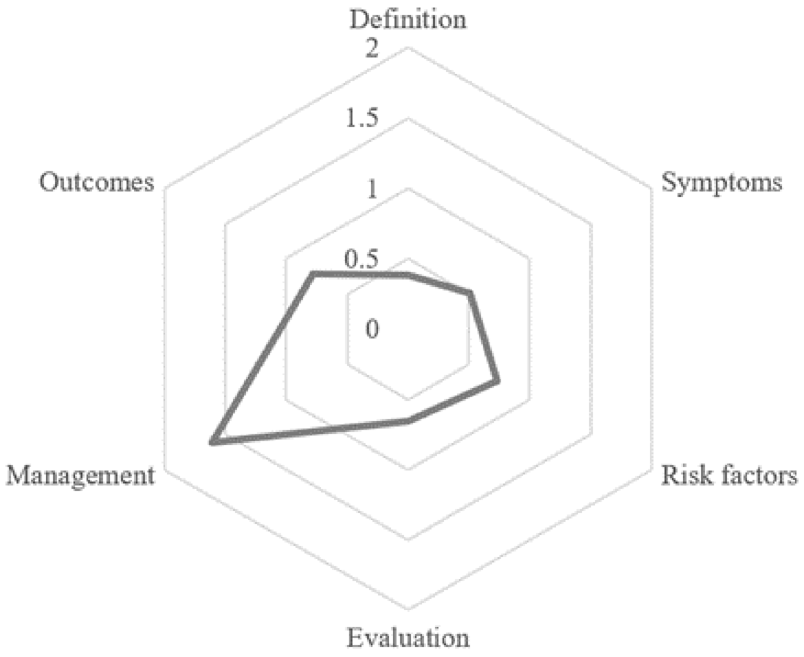
It was not our intention to exclude any type of diabetes during searching and screening. However, we found that most of the videos were about type 2 diabetes mellitus (n=193, 97%). We identified only 4 (2%) videos about gestational diabetes mellitus and 2 (1%) videos about type 1 diabetes mellitus.

Moreover, we averaged the scores of the 2 raters for each aspect of the video content and obtained scores that ranged over the full 5-point scale, from "no content" to "extensive content." The results show that more than half of the videos contain little or no content on the definition of the disease, symptoms, or evaluations of diabetes. Overall, 46.2% of the videos contain little or no content on diabetes-related risk, and 66.8% have some or more information on outcomes. Management of diabetes was the most frequent topic in the sample. Overall, 67.8% of the videos sufficiently introduced diabetes management (Table 3). The overall scores for all the videos are given in Figure 2.

**Table 3.** Completeness of video content.

| Content | Definition, n (%) | Symptoms, n (%) | Risk factors, n (%) | Evaluation, n (%) | Management, n (%) | Outcomes, n (%) |
|---|---|---|---|---|---|---|
| No content (0 points) | 113 (56.8) | 115 (57.8) | 79 (39.7) | 100 (50.3) | 14 (7.0) | 54 (27.1) |
| Little content (0.5 points) | 27 (13.6) | 9 (4.5) | 13 (6.5) | 8 (4.0) | 3 (1.5) | 12 (6.0) |
| Some content (1 point) | 52 (26.1) | 47 (23.6) | 66 (33.2) | 54 (27.1) | 44 (22.1) | 115 (57.8) |
| Most content (1.5 points) | 7 (3.5) | 14 (7.0) | 17 (8.5) | 4 (2.0) | 3 (1.5) | 5 (2.5) |
| Extensive content (2 points) | 0 (0) | 14 (7.0) | 24 (12.1) | 33 (16.6) | 135 (67.8) | 13 (6.5) |

**Figure 2.** Completeness of video content.



## Information Quality

Our results suggest that the general quality of the diabetes information videos on TikTok is acceptable. Overall, the videos published by the nonprofit organizations had the highest DISCERN scores, followed by those published by the health professionals and news agencies. The videos published by the for-profit organizations had the lowest total DISCERN scores, followed by those of the science communicators and general TikTok users. The mean numbers for the whole instrument indicate significant differences across the video sources, at α=.01 (Table 4).

Regarding reliability, videos published by the nonprofit organizations had the highest scores, while those from the for-profit organizations had the lowest scores. Our results suggest that nonprofit organizations, news agencies, and individual health professionals also contributed videos with relatively high reliability. The differences in reliability across the different video sources are significant, at α=.01.

In regard to treatment choices, diabetes-related videos on TikTok were of medium to low quality. Nonprofit organizations and health professionals contributed higher-quality videos on treatment choices than other sources; however, the differences are not significant.

**Table 4.** DISCERN scores of diabetes-related TikTok videos by source.

| Video source | Reliability of the videos (items 1-8)[a], mean (SD) | Quality of treatment choices (items 9-15)[b], mean (SD) | Overall information quality (item 16)[c], mean (SD) | Total DISCERN scores[d], mean (SD) |
|---|---|---|---|---|
| Health professionals (n=138) | 28.10 (3.59) | 16.37 (5.00) | 3.26 (0.67) | 47.74 (7.71) |
| General users (n=12) | 25.38 (3.19) | 13.42 (3.63) | 2.58 (0.51) | 41.38 (6.14) |
| Science communicators (n=6) | 25.00 (3.21) | 15.25 (5.29) | 3.08 (0.20) | 43.33 (8.23) |
| News agencies (n=31) | 28.48 (3.76) | 16.02 (5.14) | 3.23 (0.71) | 47.73 (7.39) |
| Nonprofit organizations (n=7) | 29.14 (2.25) | 18.00 (3.40) | 3.50 (0.50) | 50.64 (4.61) |
| For-profit organizations (n=5) | 24.20 (30.50) | 13.20 (5.03) | 2.60 (0.55) | 40.00 (7.11) |

[a] *P*=.005 (1-way analysis of variance).

[b] *P*=.23 (1-way analysis of variance).

[c] *P*=.004 (1-way analysis of variance).

[d] *P*=.009 (1-way analysis of variance).

## Discussion

### Principal Findings

This study systematically evaluated the information quality of diabetes-related videos on TikTok. According to a recent systematic review [13], the use of social media as a source of information is gaining in popularity among patients with diabetes. The various social media channels provide patients with a convenient means to seek medical knowledge and get social support [31]. While traditional social platforms (eg, Facebook, Twitter, and Instagram) have been widely investigated as channels of diabetes-related health communication [32], the role of emerging, mobile-based apps in disseminating diabetes knowledge is not yet fully understood. Our results reveal that TikTok is a powerful platform for disseminating diabetes-related information. The 199 diabetes videos examined in our study received 2.75 million “likes” and were commented on and shared thousands of times, which indicates that TikTok is a promising channel for health communication.

We identified 2 main categories of video uploaders (ie, individual and organizational users), each containing several more specific types of users. Individual users included health professionals, science communicators, general TikTok users; organizational users comprised news agencies, nonprofit organizations, and for-profit organizations. Health professionals contributed the most videos, while the for-profit organizations contributed the least. Many prior studies have suggested that health professionals and organizations can utilize social media for effective health communication and public health promotion [33,34]. Our study revealed that health professionals in China have been actively engaged in promoting diabetes knowledge via TikTok; however, nonprofit health organizations use this emerging video-based channel less frequently.

In terms of video content, the study found 3 types of imbalances. First, most of the videos were about type 2 diabetes mellitus, while very few videos discussed gestational diabetes mellitus and type 1 diabetes mellitus. Second, most of the videos were about disease management, but few fully addressed other aspects of content, such as the definition and symptoms of the disease, risk factors, evaluation, and outcomes. Third, when many videos are generally reliable, these videos were of average to fair quality concerning treatment choices. Prior studies suggest that patients with diabetes have various health information needs, including a need for information about treatment, course of the disease, abnormalities in glucose metabolism, progression of diabetes through their life cycle, coping techniques, and prevention [35,36]. Moreover, these information needs vary, depending on the type of diabetes mellitus. For example, young people with type 1 diabetes mellitus may be particularly interested in “diabetes through the life circle” [35]. Given the observed imbalances in video content, we suspect that current diabetes-related videos on TikTok cannot fully meet patients’ information needs. Therefore, we call for more pertinent videos to address patients’ comprehensive information needs.

Our study found that the quality of information in the videos differed with the type of source. Videos published by nonprofit organizations had the highest quality, while those from the for-profit organizations had the lowest quality. This finding is consistent with those of prior studies, which suggest that government-sponsored platforms are more likely to publish high-quality information than for-profit organizations [25].

Unfortunately, the videos contributed by the nonprofit organizations account for a mere portion of the total corpus of diabetes-related videos on TikTok. We suggest that government departments and public hospitals contribute more high-quality material and leverage the power of this social media channel to promote public health. Given the large variations in information quality from the different sources, we also suggest that patients exercise caution when using TikTok to obtain diabetes-related information.

### Limitations and Future Directions

The findings of this study should be viewed in light of several limitations. First, the study looked only at the quality of diabetes information, not the quality of communication. For example, we observed that the communication modalities varied largely in the TikTok videos. Some videos used rich materials (eg, illustrative images or cases and visual data) to communicate the information, while some were based on plain narratives. Unfortunately, the instruments employed in this paper targeted information quality and overlooked the quality of communication. We call for more empirical studies in the future to investigate the communication quality of diabetes-related videos on TikTok. Second, the videos included in our study were in Chinese; therefore, the findings cannot be applied to diabetes-related videos on TikTok in other languages (eg, English). We encourage future researchers to assess the information quality of diabetes-related videos in other languages to obtain deeper insight into quality issues with diabetes-related videos on TikTok. Third, there are many instruments for assessing the quality of health-related information, such as DISCERN, JAMA benchmarks, and the HONcode principles. This study employed the DISCERN instrument because it has proved effective for assessing the quality of videos on other platforms and apps (eg, YouTube). However, we encourage more studies using a variety of instruments to triangulate the validity of these findings in the future.

### Conclusions

This study assessed the information quality of 199 diabetes-related videos on the short video mobile app TikTok. The results show that the videos primarily addressed diabetes management but contained limited information on other types of content, such as the definition and symptoms of the disease, risk factors, evaluation, and outcomes. The overall quality of the diabetes videos was found to be acceptable on average, although it varied significantly, depending on the type of source. We conclude that the health information needs of patients with diabetes might not be fully met by watching TikTok videos, and patients should exercise caution when using TikTok for diabetes-related information.

## Acknowledgments

We thank the anonymous reviewers and editors for their helpful comments on the manuscript. The study was supported by the National Science Foundation of China (72074112, 72174083) and the Key Projects of Philosophy and Social Sciences Research of Chinese Ministry of Education (19JZD021).

## Conflicts of Interest

None declared.

## References

1. IDF Diabetes Atlas 2019. International Diabetes Federation. URL: https://www.diabetesatlas.org/en/resources/ [accessed 2021-08-16]
2. National Diabetes Statistics Report, 2020. Atlanta, GA. Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services. 2020. URL: https://www.cdc.gov/diabetes/data/statistics-report/index.html [accessed 2021-08-16]
3. Li Y, Teng D, Shi X, Qin G, Qin Y, Quan H, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. BMJ 2020 Apr 28;369:m997 [FREE Full text] [doi: 10.1136/bmj.m997] [Medline: 32345662]
4. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet 1998 Sep 12;352(9131):837-853. [doi: 10.1016/S0140-6736(98)07019-6] [Medline: 9742976]
5. Dluhy RG, McMahon GT. Intensive glycemic control in the ACCORD and ADVANCE trials. N Engl J Med 2008 Jun 12;358(24):2630-2633. [doi: 10.1056/NEJMe0804182] [Medline: 18539918]
6. ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2560-2572. [doi: 10.1056/NEJMoa0802987] [Medline: 18539916]
7. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
8. Xu Y, Wang L, He J, Bi Y, Li M, Wang T, 2010 China Noncommunicable Disease Surveillance Group. Prevalence and control of diabetes in Chinese adults. JAMA 2013 Sep 04;310(9):948-959. [doi: 10.1001/jama.2013.168118] [Medline: 24002281]
9. Bains SS, Egede LE. Associations between health literacy, diabetes knowledge, self-care behaviors, and glycemic control in a low income population with type 2 diabetes. Diabetes Technol Ther 2011 Mar;13(3):335-341 [FREE Full text] [doi: 10.1089/dia.2010.0160] [Medline: 21299402]

10. Weymann N, Härter M, Dirmaier J. Information and decision support needs in patients with type 2 diabetes. Health Informatics J 2016 Mar;22(1):46-59 [FREE Full text] [doi: 10.1177/1460458214534090] [Medline: 24916569]
11. Lee DYL, Armour C, Krass I. The development and evaluation of written medicines information for type 2 diabetes. Health Educ Res 2007 Dec;22(6):918-930. [doi: 10.1093/her/cym048] [Medline: 17947249]
12. Nordfeldt S, Ängarne-Lindberg T, Berterö C. To use or not to use--practitioners' perceptions of an open web portal for young patients with diabetes. J Med Internet Res 2012 Nov 09;14(6):e154 [FREE Full text] [doi: 10.2196/jmir.1987] [Medline: 23137767]
13. Elnaggar A, Ta Park V, Lee SJ, Bender M, Siegmund LA, Park LG. Patients' Use of Social Media for Diabetes Self-Care: Systematic Review. J Med Internet Res 2020 Apr 24;22(4):e14209 [FREE Full text] [doi: 10.2196/14209] [Medline: 32329745]
14. Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res 2014 Jan 10;16(1):e12 [FREE Full text] [doi: 10.2196/jmir.2630] [Medline: 24413148]
15. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
16. Oser SM, Stuckey HL, Parascando JA, McGinley EL, Berg A, Oser TK. Glycated Hemoglobin Differences Among Blog-Reading Adults With Type 1 Diabetes Compared With Those Who Do Not Read Blogs: Cross-Sectional Study. JMIR Diabetes 2019 Apr 02;4(2):e13634 [FREE Full text] [doi: 10.2196/13634] [Medline: 30938693]
17. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent Surfing on Social Health Networks is Associated With Increased Knowledge and Patient Health Activation. J Med Internet Res 2016 Aug 10;18(8):e212 [FREE Full text] [doi: 10.2196/jmir.5832] [Medline: 27511272]
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
19. Song S, Zhang Y, Yu B. Interventions to support consumer evaluation of online health information credibility: A scoping review. Int J Med Inform 2021 Jan;145:104321. [doi: 10.1016/j.ijmedinf.2020.104321] [Medline: 33202372]
20. Song S, Zhao Y, Song X, Zhu Q. The Role of Health Literacy on Credibility Judgment of Online Health Misinformation. 2019 Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); June 10-13, 2019; Xi'an. [doi: 10.1109/ichi.2019.8904844]
21. Zhang Y, Song S. Older Adults' Evaluation of the Credibility of Online Health Information. 2020 Presented at: CHIIR '20: Conference on Human Information Interaction and Retrieval; March 14-18, 2020; Vancouver, BC. [doi: 10.1145/3343413.3377997]
22. Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021 Apr;148:104375. [doi: 10.1016/j.ijmedinf.2021.104375] [Medline: 33461008]
23. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users' intention to continue the use of TikTok. INTR 2021 Jun 24;ahead-of-print(ahead-of-print):1-23. [doi: 10.1108/intr-10-2020-0593]
24. Ostrovsky AM, Chen JR. TikTok and Its Role in COVID-19 Information Propagation. J Adolesc Health 2020 Nov;67(5):730 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.039] [Medline: 32873499]
25. Goobie GC, Guler SA, Johannson KA, Fisher JH, Ryerson CJ. YouTube Videos as a Source of Misinformation on Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc 2019 May;16(5):572-579. [doi: 10.1513/AnnalsATS.201809-644OC] [Medline: 30608877]
26. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
27. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
28. ReFaey K, Tripathi S, Yoon JW, Justice J, Kerezoudis P, Parney IF, et al. The reliability of YouTube videos in patients education for Glioblastoma Treatment. J Clin Neurosci 2018 Sep;55:1-4. [doi: 10.1016/j.jocn.2018.07.001] [Medline: 30075980]
29. Radonjic A, Fat Hing NN, Harlock J, Naji F. YouTube as a source of patient information for abdominal aortic aneurysms. J Vasc Surg 2020 Feb;71(2):637-644. [doi: 10.1016/j.jvs.2019.08.230] [Medline: 31611104]
30. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube Useful as a Source of Health Information for Adults With Type 2 Diabetes? A South Asian Perspective. Can J Diabetes 2018 Aug;42(4):395-403.e4. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
31. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]

32. Gabarron E, Larbi D, Dorronzoro E, Hasvold PE, Wynn R, Årsand E. Factors Engaging Users of Diabetes Social Media Channels on Facebook, Twitter, and Instagram: Observational Study. J Med Internet Res 2020 Sep 29;22(9):e21204 [FREE Full text] [doi: 10.2196/21204] [Medline: 32990632]
33. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What People "Like": Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
34. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
35. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. Syst Rev 2018 Feb 14;7(1):27 [FREE Full text] [doi: 10.1186/s13643-018-0690-0] [Medline: 29444711]
36. Kuske S, Schiereck T, Grobosch S, Paduch A, Droste S, Halbach S, et al. Diabetes-related information-seeking behaviour: a systematic review. Syst Rev 2017 Oct 24;6(1):212 [FREE Full text] [doi: 10.1186/s13643-017-0602-8] [Medline: 29065919]

---

*Edited by G Eysenbach; submitted 13.05.21; peer-reviewed by A Alasmari, M Lim, R Fang; comments to author 04.06.21; revised version received 21.07.21; accepted 07.08.21; published 01.09.21*

*Please cite as:*
*Kong W, Song S, Zhao YC, Zhu Q, Sha L*
*TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos*
*J Med Internet Res 2021;23(9):e30409*
*URL: https://www.jmir.org/2021/9/e30409*
*doi: 10.2196/30409*
*PMID:*

Original Paper

# COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation

Saranya Sankaranarayanan*, MS; Jagadheshwar Balan*, MS; Jesse R Walsh, PhD; Yanhong Wu, PhD; Sara Minnich, BA; Amy Piazza, MLS; Collin Osborne, MS; Gavin R Oliver, MS; Jessica Lesko, MA; Kathy L Bates, MBA; Kia Khezeli, PhD; Darci R Block, PhD; Margaret DiGuardo, MD; Justin Kreuter, MD; John C O'Horo, MPH, MD; John Kalantari, PhD; Eric W Klee, PhD; Mohamed E Salama, MD; Benjamin Kipp, PhD; William G Morice, MD, PhD; Garrett Jenkinson, PhD

Mayo Clinic, Rochester, MN, United States

*these authors contributed equally

**Corresponding Author:**
Garrett Jenkinson, PhD
Mayo Clinic
200 1st St SW
Rochester, MN, 55905
United States
Phone: 1 507 293 9457
Email: Jenkinson.William@mayo.edu

## Abstract

**Background:** COVID-19 is caused by the SARS-CoV-2 virus and has strikingly heterogeneous clinical manifestations, with most individuals contracting mild disease but a substantial minority experiencing fulminant cardiopulmonary symptoms or death. The clinical covariates and the laboratory tests performed on a patient provide robust statistics to guide clinical treatment. Deep learning approaches on a data set of this nature enable patient stratification and provide methods to guide clinical treatment.

**Objective:** Here, we report on the development and prospective validation of a state-of-the-art machine learning model to provide mortality prediction shortly after confirmation of SARS-CoV-2 infection in the Mayo Clinic patient population.

**Methods:** We retrospectively constructed one of the largest reported and most geographically diverse laboratory information system and electronic health record of COVID-19 data sets in the published literature, which included 11,807 patients residing in 41 states of the United States of America and treated at medical sites across 5 states in 3 time zones. Traditional machine learning models were evaluated independently as well as in a stacked learner approach by using AutoGluon, and various recurrent neural network architectures were considered. The traditional machine learning models were implemented using the AutoGluon-Tabular framework, whereas the recurrent neural networks utilized the TensorFlow Keras framework. We trained these models to operate solely using routine laboratory measurements and clinical covariates available within 72 hours of a patient's first positive COVID-19 nucleic acid test result.

**Results:** The GRU-D recurrent neural network achieved peak cross-validation performance with 0.938 (SE 0.004) as the area under the receiver operating characteristic (AUROC) curve. This model retained strong performance by reducing the follow-up time to 12 hours (0.916 [SE 0.005] AUROC), and the leave-one-out feature importance analysis indicated that the most independently valuable features were age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen level, and serum iron level. In the prospective testing cohort, this model provided an AUROC of 0.901 and a statistically significant difference in survival (*P*<.001, hazard ratio for those predicted to survive, 95% CI 0.043-0.106).

**Conclusions:** Our deep learning approach using GRU-D provides an alert system to flag mortality for COVID-19–positive patients by using clinical covariates and laboratory values within a 72-hour window after the first positive nucleic acid test result.

***(J Med Internet Res 2021;23(9):e30157)*** doi: 10.2196/30157

**KEYWORDS**

COVID-19; mortality; prediction; recurrent neural networks; missing data; time series; deep learning; machine learning; neural network; electronic health record; EHR; algorithm; development; validation

## Introduction

COVID-19 is caused by the SARS-CoV-2 virus and is suspected to be of zoonotic origin, with spillover from bats or pangolins into humans in Wuhan, China [1,2]. COVID-19 has become one of the largest public health emergencies of the past century with over 203 million confirmed cases and 4.3 million deaths as of August 2021 according to the World Health Organization [3]. The COVID-19 pandemic has overwhelmed global medical supply chains, hospitals, and economies, which has led governments to respond with varying policies, including mask mandates and travel restrictions [4,5]. At times, hospitals and health care workers have become so overburdened with patients with COVID-19 that they have been forced to ration care, raising logistical and ethical concerns [6].

The clinical course of COVID-19 is diverse with most individuals experiencing mild or asymptomatic disease, but many patients develop life-threatening diseases, including features such as cytokine storms, thrombotic complications, or severe acute respiratory syndrome requiring mechanical ventilation or extracorporeal membrane oxygenation [7]. A major medical challenge is therefore to reliably triage patients according to their risk for severe disease. Age is consistently observed to be a predominant risk factor for severe disease [7], but deaths are not limited to older adults and the majority of older patients survive COVID-19 [7]. Other comorbidities and laboratory test values are expected to be capable of further individualizing and enhancing mortality prediction. Recent studies investigating statistical and machine learning (ML) models for mortality prediction have confirmed that detailed evaluation of medical records can facilitate further stratification of patients [8-12].

A systematic review of 147 published or preprint prediction models found consistent problems with inherent biases in the data sets investigated or created in all such studies, ultimately concluding that “we do not recommend any of these reported prediction models for use in current practice” [12]. Clinical practices differ in the nature of their observational electronic health record (EHR) data set, patient population, clinical practices, and electronic record or laboratory ordering practices. Correspondingly, the literature review conducted at the outset of this study indicated that the existing prediction models were likely unsuited to our clinical setting without essentially starting afresh by retraining, validating, and testing predictions.

We describe Mayo Clinic’s experience assembling, what is to our knowledge, the largest reported COVID-19 database for mortality prediction and using this database to create a system for COVID-19 mortality prediction, tailored to a unique patient population. Despite the biases inherent to it, because this large and growing database represents a health care system spanning 5 states and 3 time zones over a study window greater than 11 months, our model is expected to be the least confounded and the most generalized COVID-19 mortality predictor published to date. We report the successful development and validation of a state-of-the-art ML model to provide mortality prediction shortly after confirmation of SARS-CoV-2 infection in this Mayo Clinic patient population and discuss in detail the various logistical and scientific challenges involved in the early deployment of such a system in a rapidly changing pandemic environment.

## Methods

### Study Design

This work required the development of a data set and the subsequent modeling of the resultant cohort. After data collection and cleaning, 2 broad classes of algorithms were considered to model this data. The first approach ignores the time series nature of the underlying data and applies traditional ML classifiers. The second approach explicitly models the time series data while dealing with the missing-not-at-random (MNAR) values by using specialized recurrent neural networks (RNNs). Both types of modeling methods were run independently and compared using cross-validation, and a single winning model was selected for prospective performance validation.

### EHR and Laboratory Information System Observational Cohort Data Collection

This study adheres to a research protocol approved by the Mayo Clinic Institutional Review Board. Data were retrospectively collected after March 1, 2020 on COVID-19–positive individuals presenting to a Mayo Clinic site or health system, while excluding patients without research consent or from European Union countries covered by the general data protection regulation law. We restricted our focus to 11,807 patients with a positive COVID-19 nucleic acid test result on or before January 27, 2021 and at least one non-COVID test result. Although the data collection system is deployed and ongoing, the January cutoff was selected for this study to provide sufficient cohort size while allowing a minimum of 3 weeks of follow-up to accurately establish survival status. Mayo Clinic’s EHR and laboratory information system (LIS) contain data from each of its 3 campuses (Rochester, Minnesota; Jacksonville, Florida; Scottsdale, Arizona) as well as the surrounding health system sites spanning 5 states (MN, IA, WI, FL, and AZ). Although the EHR contains clinically reportable laboratory results, many of these can only be reported within defined ranges, which can result in qualitative text values rather than the raw numeric measurements. Because many ML algorithms typically work better with quantitative rather than qualitative results, we used the LIS to gather such laboratory testing measurements and the EHR to gather the remaining variables. The EHR data were queried from an underlying Db2 database (IBM Corp), and the LIS data were queried from an SQL database (Microsoft).

### Multivariate Time-Series Data With Missingness

The clinical covariates collected were age, sex, height, weight, Charlson comorbidity score, temperature, blood pressure, respiratory rate, oxygen saturation ($SpO_2$) levels, and diagnoses of chronic kidney disease or diabetes mellitus. Furthermore, we included laboratory test values from a basic metabolic panel, complete blood counts, and some less routine test results of relevance to COVID-19, as determined by scientific literature and physician collaborators. Table 1 details the features collated into our database. In Multimedia Appendix 1, we provide a detailed breakdown of these clinical covariates and laboratory values in our cohort (Table S1 of Multimedia Appendix 1) as well as the cohort's geographic distribution (Figure S1 of Multimedia Appendix 1). Differentiating between missing data and absence of a condition is not possible from EHR diagnostic codes, particularly for patients treated in an outpatient setting. Therefore, we focused mainly on the Charlson comorbidity index [13], which is populated in our EHR when there is a recorded medical history during a "patient encounter" in the EHR. Thus, this variable is available and can be assigned a value corresponding to no comorbidities, which is distinct from missingness in the case of no recorded medical history in the EHR. However, owing to their emphasis within the literature, we also included chronic kidney disease [9] and diabetes mellitus [14] as independent comorbidity variables using their ICD-10 (International Classification of Diseases) codes while acknowledging that these variables conflate missingness with lack of a condition.

Clinical covariates such as pre-existing conditions, height, and weight were sampled infrequently, whereas heart rate and $SpO_2$ were recorded every 15 minutes for inpatients in our EHR, and other laboratory tests were intermediate in terms of frequency. Therefore, to deal with these multiscale time series measurements, we used the laboratory measurements as the starting point to define our sampling time points. For the variables of sex, age, weight, height, diabetes mellitus, chronic kidney disease, and Charlson comorbidity index, we encoded these variables to exist at the first time point only; in our top performing RNN models, we observed no difference in performance using this strategy when compared to repeating the observations at each time point. For the frequently observed variables of blood pressure systole, blood pressure diastole, temperature, pulse, respiratory rate, and $SpO_2$, we computed the minimum and maximum measurements for each calendar day and appended these to each laboratory time point during those dates; if no laboratory time point existed on a given day, we created a new one at noon using these minimum and maximum values. We considered time points within ±72 hours of each patient's first positive polymerase chain reaction (PCR) result and performed a sensitivity analysis on the length of the patient follow-up after this positive test result.

**Table 1.** Feature measurements collected.

| Abbreviation | Description (units or levels) |
|---|---|
| sex | Sex (male or female) |
| age | Age at time of polymerase chain reaction–positive test result (years) |
| weight | Weight (kg) |
| height | Height (cm) |
| PCR | SARS-CoV-2 nucleic acid test (+ or –) |
| SERO | SARS-CoV-2 serology antibody test (+ or –) |
| BASAA | Basophil count test ($10^9$/L) |
| EOSAA | Eosinophil count test ($10^9$/L) |
| HCT | Hematocrit test (%) |
| HGB | Hemoglobin test (g/dL) |
| LYMAA | Lymphocyte count test ($10^9$/L) |
| MCV | Mean corpuscular volume test (fL) |
| MONAA | Monocyte count test ($10^9$/L) |
| NEUAA | Neutrophil count test ($10^9$/L) |
| PLTC | Platelet count test ($10^9$/L) |
| RBC | Red blood cell count test ($10^{12}$/L) |
| RDW | Red cell distribution width test (%) |
| WBC | White blood cell count test ($10^9$/L) |
| CRP | C-reactive protein test (mg/L) |
| D-DIMER | D-dimer test (ng/mL) |
| FERR | Ferritin test (mg/L) |
| IL6 | Interleukin-6 test (pg/mL) |
| TRPS | Troponin T test (ng/L) |
| FIBTP | Fibrinogen test (mg/dL) |
| LD | Lactate dehydrogenase test (U/L) |
| IRON | Serum iron test (mg/dL) |
| TIBC | Total iron binding capacity test (mg/dL) |
| SAT | Percent iron saturation test (%) |
| TRSFC | Transferrin test (mg/dL) |
| BUN | Blood urea nitrogen test (mg/dL) |
| CHL | Chloride test (mmol/L) |
| GLU | Glucose test (mg/dL) |
| CALC | Calcium test (mg/dL) |
| CREA | Creatinine test (mg/dL) |
| POTA | Potassium test (mmol/L) |
| ALB | Albumin test (g/dL) |
| BICA | Bicarbonate test (mmol/L) |
| SODI | Sodium test (mmol/L) |
| BILI | Bilirubin test (mg/dL) |
| BPsystole | Blood pressure systole (mm Hg) |

| Abbreviation | Description (units or levels) |
|---|---|
| BPdiastole | Blood pressure diastole (mm Hg) |
| Temp | Temperature (°C) |
| Pulse | Heart rate (1/min) |
| Resp | Respiratory rate (1/min) |
| $SpO_2$ | Oxygen saturation (%) |
| Charlson | Charlson comorbidity index (10-year survival probability) |
| CKD | Chronic kidney disease (+ or –) |
| DM | Diabetes mellitus (+ or –) |

## Time-Flattened ML Models

Time series data were flattened/encoded to a fixed length list of features by carry forward imputation (ie, selection of the most recently observed covariate values), ensuring compatibility with traditional ML models. Specifically, after the data are flattened in this fashion, it forms a tabular prediction task suitable for any canonical supervised classification algorithm. The recently published [15] Python-based automated ML tool AutoGluon-Tabular (v0.2.0) was utilized to enable standardized and reproducible ensemble stacking of many model classes (eg, deep neural networks, LightGBM boosted trees, CatBoost boosted trees, Random Forests, Extremely Randomized Trees, XGBoost, and k-Nearest Neighbors).

AutoGluon-Tabular models were fit to our tabular data frames using the “AutoGluon.TabularPrediction.fit” function using all the default parameters except eval_metric='roc_auc'. After running the fit function, access to each individual model created by AutoGluon was achieved by the “get_model_names” method on the resulting prediction object. This then allowed us to pass the specified model to the “predict_proba” method’s optional “model” argument for each of the following model types: KNeighborsUnif, KNeighborsDist, NeuralNetFastAI, LightGBMLarge, NeuralNetMXNet, RandomForestGini, ExtraTreesGini, RandomForestEntr, ExtraTreesEntr, LightGBM, XGBoost, LightGBMXT, CatBoost, WeightedEnsemble_L2. Hereon, we refer to WeightedEnsemble_L2 as the “AutoGluon” model since this was the output of the “predict_proba” method when no single model type was specified.

For relatively static features such as height, weight, or Charlson comorbidity index, we would expect the time-flattened models to be at no disadvantage, whereas the more frequently measured data such as laboratory values or blood pressure will lose information, particularly about trends in the covariates. For instance, 2 individuals with a fever of 39°C recorded in the most recent observation would be treated the same even if one had a sustained high fever and the other had a brief downward trending spike. Of course, there are many potential degrees of freedom to capture more information in the flattened data; one could define a fixed number of the most recent observations or fit a line through the observations over time and pass the slope and intercept as features to the classifier. However, ultimately, the choice to flatten the time series is a choice of convenience and one that attempts to leverage the extensive research efforts devoted to tabular prediction, and therefore, we study here only the last observation carried forward modeling, since proper modeling efforts should account for the time series structure in the EHR data. We next look at models of this form.

## RNN Time Series Models

As the second approach, we implemented the modified gated recurrent unit (GRU) binary classification models proposed by Che et al [16] that are capable of accounting for the MNAR patterns within EHR data, and we adopt their notation. For a given patient, we have $D = 54$ variables and a given time series of $T$ time points can be represented as a $T \times D$ matrix $X$ whose rows $x_t \in \mathbb{R}D$, $t= 1, \ldots, T$ represent the $t$-th observation with $D$ variables $x_t^d$, $d = 1, \ldots, D$. Accompanying each observation $x_t$ is a time stamp $s_t \in \mathbb{R}$, which starts at time 0, $s_1 = 0$ and a binary masking vector $m_t \in \{0, 1\}^D$ with $m_t^d$ taking value 1 when $x_t^d$ is observed and 0 otherwise. From these values, we can compute the time intervals.

$$\delta_t^d = \begin{cases} s_t - s_{t-1} + \delta_{t-1}^d, & \text{if } t > 1, m_{t-1}^d = 0 \\ s_t - s_{t-1}, & \text{if } t > 1, m_{t-1}^d = 1 \\ 0, & \text{otherwise} \end{cases} \quad \textbf{(1)}$$

With these definitions, we can look at various modifications to the standard GRU architecture whose $j$-th hidden unit has a reset gate $r_t^j$ and update gate $r_t^j$ with hidden state $h_t^j$ at time $t$ and update the equations.

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad \textbf{(2)}$$
$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad \textbf{(3)}$$
$$\tilde{h}_t = \tanh(W x_t + U(r \odot h_{t-1}) + b) \quad \textbf{(4)}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad \textbf{(5)}$$

With matrices $W_z$, $W_r$, $W$, $U_z$, $U_r$,$U$ and vectors $b_z$, $b_r$, $b$ composed of model parameters, $\odot$ is the Hadamard product, and $\sigma(\cdot)$ is the elementwise sigmoid function. Before modifying the architecture, there are 3 methods to use the GRU above to handle missing data: in “GRU-Mean,” missing values are imputed by their means in the training data; in “GRU-Forward,” missing values are imputed by their last observed value; and in “GRU-Simple,” we simply concatenate the $x_t$, $m_t$, and $\delta_t$ variables into a single observation vector $x$ ‘$t$ and pass this through the GRU equations above. The GRU-D method uses trainable decay weights.

$$\gamma_t = \exp\{-\max(0, W_\gamma \delta_t + b_\gamma)\} \qquad \textbf{(6)}$$

With $W_\gamma$ and $b_\gamma$ being trainable model parameters. The observations are then replaced by the update.

$$\hat{x}_t^d = m_t^d x_t^d + (1 - m_t^d)(\gamma_{x_t}^d x_{t'}^d + (1 - \gamma_{x_t}^d)\tilde{x}^d \qquad \textbf{(7)}$$

where $x_{t'}^d$ is the last observed value of the *d*-th variable and $\tilde{x}^d$ is the empirical mean of the *d*-th variable in the training data. The modified GRU update equations for GRU-D become the following.

$$r_t = \sigma(W_r \hat{x}_t + U_r \hat{h}_{t-1} + V_r m_t + b_r) \qquad \textbf{(8)}$$
$$z_t = \sigma(W_z \hat{x}_t + U_z \hat{h}_{t-1} + V_z m_t + b_z) \qquad \textbf{(9)}$$
$$\tilde{h}_t = \tanh(W \hat{x}_t + U(r \odot \hat{h}_{t-1}) + V m_t + b) \qquad \textbf{(10)}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \qquad \textbf{(11)}$$
$$h_{t-1} = \gamma_{h_t} \odot h_{t-1} \qquad \textbf{(12)}$$

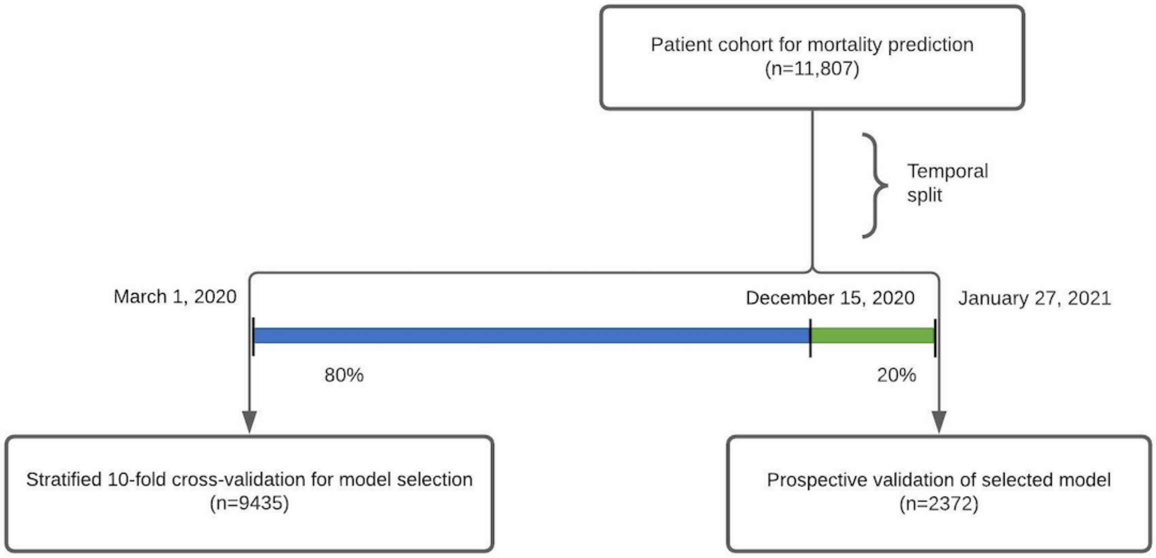
where $V_z$, $V_r$, $V$ are new model parameters to directly handle the masking vector $m_t$ in the model.

Our implementation of the above equations in Python is a slightly modified version of the code available on the GRU-D paper's [16] GitHub repository. For the core RNN algorithms, we only edited the original GRU-D code, where required, to be compatible with the more recent versions of tensorflow.keras (version 2.1.0) and numpy (version 1.19.2) used in our high-performance computing cluster environment. We selected the specific RNN algorithm by setting the "--model" argument to be "GRUforward," "GRU0," "GRUsimple," and "GRUD" for GRU-forward, GRU-mean, GRU-simple, and GRU-D, respectively. We utilized the default hyperparameters of the algorithm; however, in our testing, we found that increasing the batch size from 32 to 256 facilitated faster training of the algorithms. Therefore, a batch size of 256 is the only nondefault hyperparameter selection made in our implementation of the RNN algorithms.

### Temporal Cohort Split

As depicted in the CONSORT (Consolidated Standards of Reporting Trials) diagram of Figure 1, patients who first tested positive for COVID-19 from March 1, 2020 through December 15, 2020 (9435/11,807, 79.9%) were assigned to a model selection cohort, whereas patients who first tested positive for COVID-19 from December 16, 2020 through January 27, 2021 (2372/11,807, 20.1%) were used as a prospective testing cohort for the final algorithm. All experiments in the model selection cohort were performed using an identical 10-fold stratified cross-validation using binary classification with the positive class defined as death within 21 days of the first positive PCR test result. Only the single best performing model was evaluated on the prospective cohort after being fit against the entire model selection cohort.

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) diagram demonstrating the temporal split of our cohort for model selection and prospective validation.



## Results

### Model Selection

In Figure 2 and Table 2, we compared the results of our various models by using cross-validation area under the receiver operator characteristic (AUROC) curve in the training cohort. Although not in a statistically significant way, we recapitulated the findings of Che et al [16], discovering that the GRU-D model has the highest average cross-validation AUROC curve among all other standard variants of GRU modeling in time series with missing values. In addition, GRU-Simple has higher average cross-validation AUROC curve than the GRU-Forward and GRU-Mean, and the most notable difference underlying these categories is the inclusion of missingness indicators as features to GRU-Simple, which could indicate the value of MNAR patterns in the classification task. GRU-D's biologically inspired architecture attempts to make even more efficient use of this information and exceeds the performance of all the tested RNN methods. AutoGluon, which only had access to the last measurement of each variable, showed strong performance despite this limitation. In Table 2, each individual AutoGluon model was also benchmarked (those with suffix "-AutoGluon"), along with the final ensemble estimate (labeled simply as "AutoGluon"). Although GRU-D ultimately outranked AutoGluon, each method's performance fell within the other's standard error intervals. AutoGluon's automated hyperparameter tuning and model stacking may indicate that GRU-D could benefit from the addition of hyperparameter search. However, this process may risk overfitting this cross-validation data set, and thus, we selected GRU-D with the default settings rather than attempting to further improve the cross-validation AUROC curve via hyperparameter optimization.

**Figure 2.** Receiver operating characteristic curves for the 18 models evaluated. AUROC: area under the receiver operating characteristic.

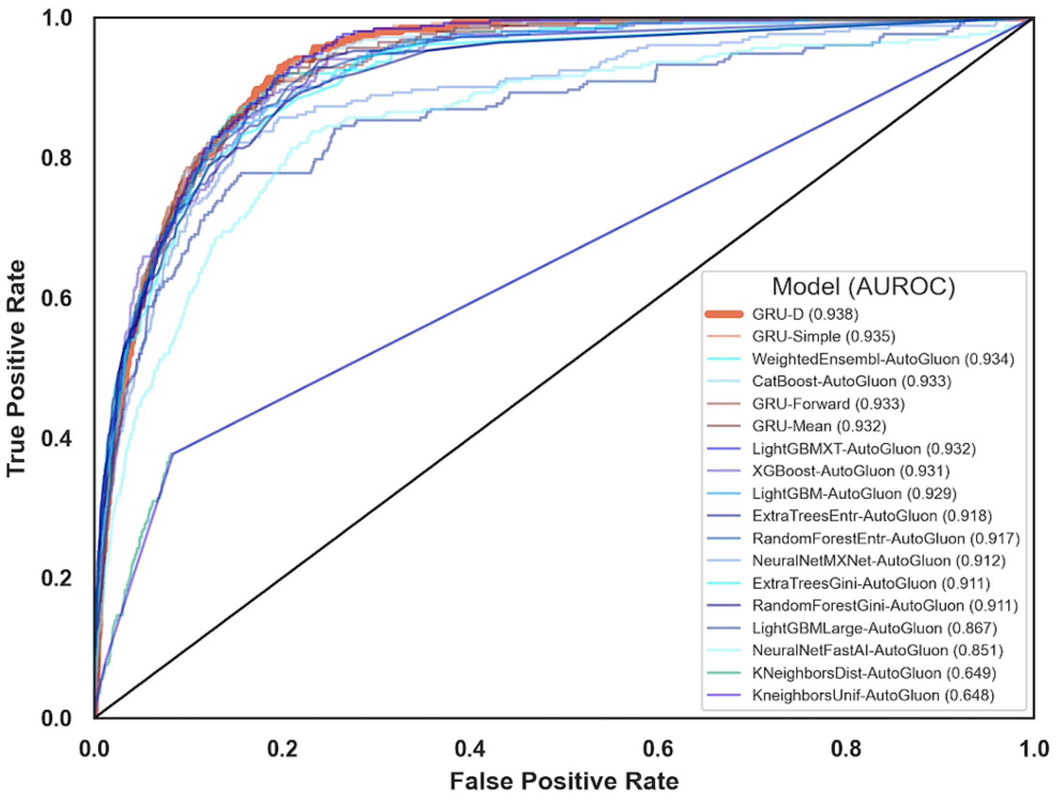


**Table 2.** Modeling results sorted by performance.

| Model | Area under the receiver operator characteristic curve (SE) |
|---|---|
| KNeighborsUnif-AutoGluon | 0.648 (0.011) |
| KNeighborsDist-AutoGluon | 0.649 (0.011) |
| NeuralNetFastAI-AutoGluon | 0.858 (0.013) |
| LightGBMLarge-AutoGluon | 0.867 (0.014) |
| NeuralNetMXNet-AutoGluon | 0.907 (0.008) |
| RandomForestGini-AutoGluon | 0.911 (0.007) |
| ExtraTreesGini-AutoGluon | 0.911 (0.009) |
| RandomForestEntr-AutoGluon | 0.917 (0.008) |
| ExtraTreesEntr-AutoGluon | 0.918 (0.007) |
| LightGBM-AutoGluon | 0.929 (0.007) |
| XGBoost-AutoGluon | 0.931 (0.006) |
| LightGBMXT-AutoGluon | 0.931 (0.005) |
| GRU-Mean | 0.932 (0.005) |
| GRU-Forward | 0.933 (0.006) |
| CatBoost-AutoGluon | 0.933 (0.005) |
| AutoGluon | 0.934 (0.005) |
| GRU-Simple | 0.935 (0.004) |
| GRU-D | 0.938 (0.004) |

## Length of Time Series

Clearly, we would expect availability of more time series data to result in improved model performance. To determine if predictions could be made utilizing data prior to 72 hours of a patient's first positive PCR test result, we assessed the performance of GRU-D when we restricted the time series to 12, 24, 48, and 72 hours of follow-up after the first positive PCR test result. The results in Table 3 demonstrate that although we lose performance when predicting earlier in the patient's disease, we are still able to provide accurate predictions even using data within the same day (12 hours of follow-up) that a patient tests positive for COVID.

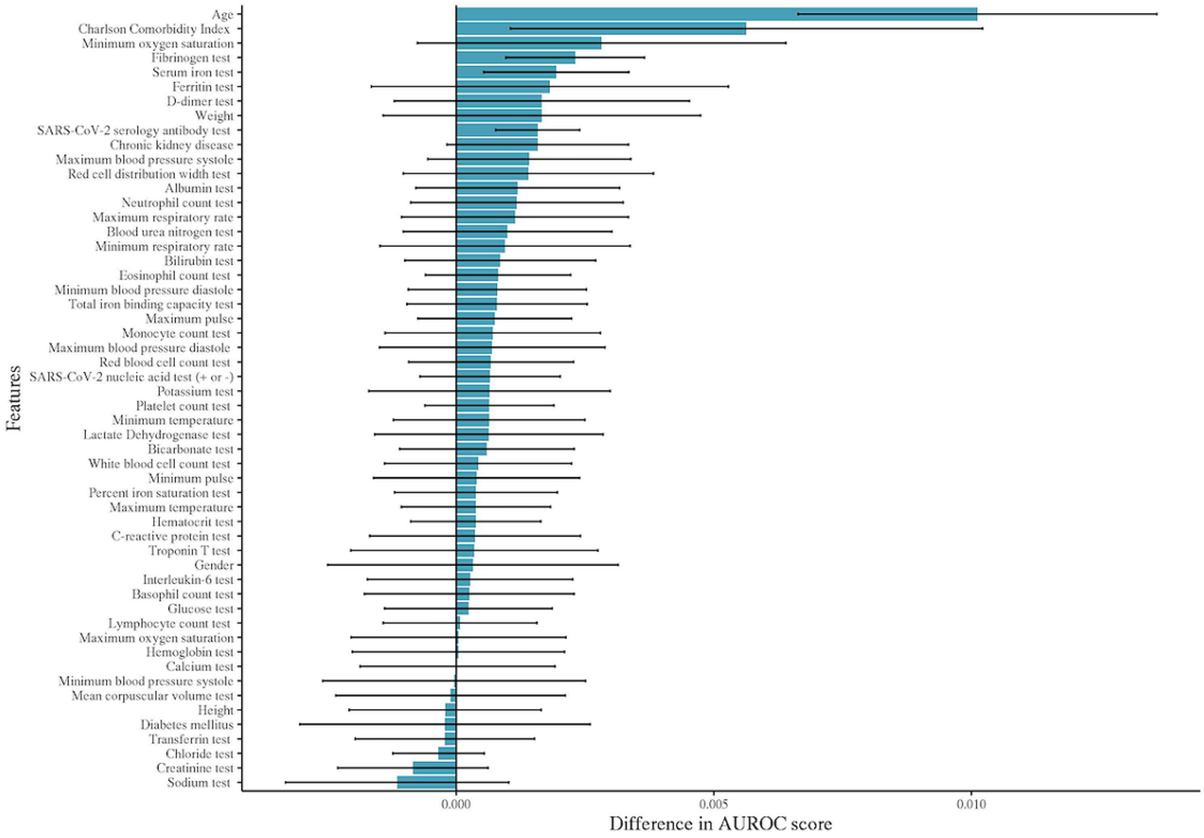
**Table 3.** GRU-D performance versus length of time series.

| Follow-up after positive finding for polymerase chain reaction | Area under the receiver operator characteristic curve (SE) |
|---|---|
| 12 h | 0.916 (0.005) |
| 24 h | 0.919 (0.006) |
| 48 h | 0.925 (0.005) |
| 72 h | 0.938 (0.004) |

## MNAR as an Asset and Feature Importance

To demonstrate the fact that MNAR data can improve model predictions by GRU-D, we generated a synthetic data set with laboratory test values replaced by Bernoulli coin flips. Therefore, the only valuable information contained within this data set's laboratory values is the missing data patterns that can be viewed as encoding clinical suspicion or concern. For instance, the D-dimer laboratory value is ordered less frequently than other tests, and therefore, its presence alone can be informative of clinical concern for thrombotic events. Our results found that randomizing the laboratory values resulted in an AUROC curve of 0.904 (0.006), which indicates that the laboratory values in aggregate contributed 0.034 to the AUROC score (since this is the drop in performance compared to the model with actual laboratory values). We ran a further experiment omitting the laboratory values entirely, which produced a lower AUROC of 0.890 (0.006). Therefore, the missing patterns alone contributed 0.014 to the AUROC. To contextualize this finding, we dropped each feature individually from the model, assessed the decrease in the AUROC score, and summarized the top 10 features in the decreasing order of the difference in the AUROC score (Figure 3). We note here that the drop due to missing patterns exceeds the drop due to removing any single variable from the analysis, making the MNAR pattern one of the most valuable pieces of information available to GRU-D. In Multimedia Appendix 1, we show a detailed error analysis of our model using these top 10 features. The fact that age and Charlson comorbidity index are the most significant contributors to mortality prediction is consistent with the well-known risk factors for COVID mortality. The findings of the fibrinogen test, serum iron test, and ferritin test were the 3 most important laboratory values in our models. The presence of chronic kidney disease, weight, serology test, and $SpO_2$ were the clinical covariates that also ranked in the top 10 variables by importance. Interestingly, height had low importance, indicating that BMI may not be as effective as weight itself in mortality prediction. However, a limitation of this drop-one-feature variable importance is that a low-ranking feature such as height cannot be said to be irrelevant, just that any information it carries is redundant within other features.

**Figure 3.** Feature importance in the GRU-D recurrent neural network model as defined by the average drop in the area under the receiver operator characteristic curve (with 95% CI) when each feature is individually removed from the analysis. The top 5 features are seen to be age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen levels, and serum iron levels. AUROC: area under the receiver operator characteristic.

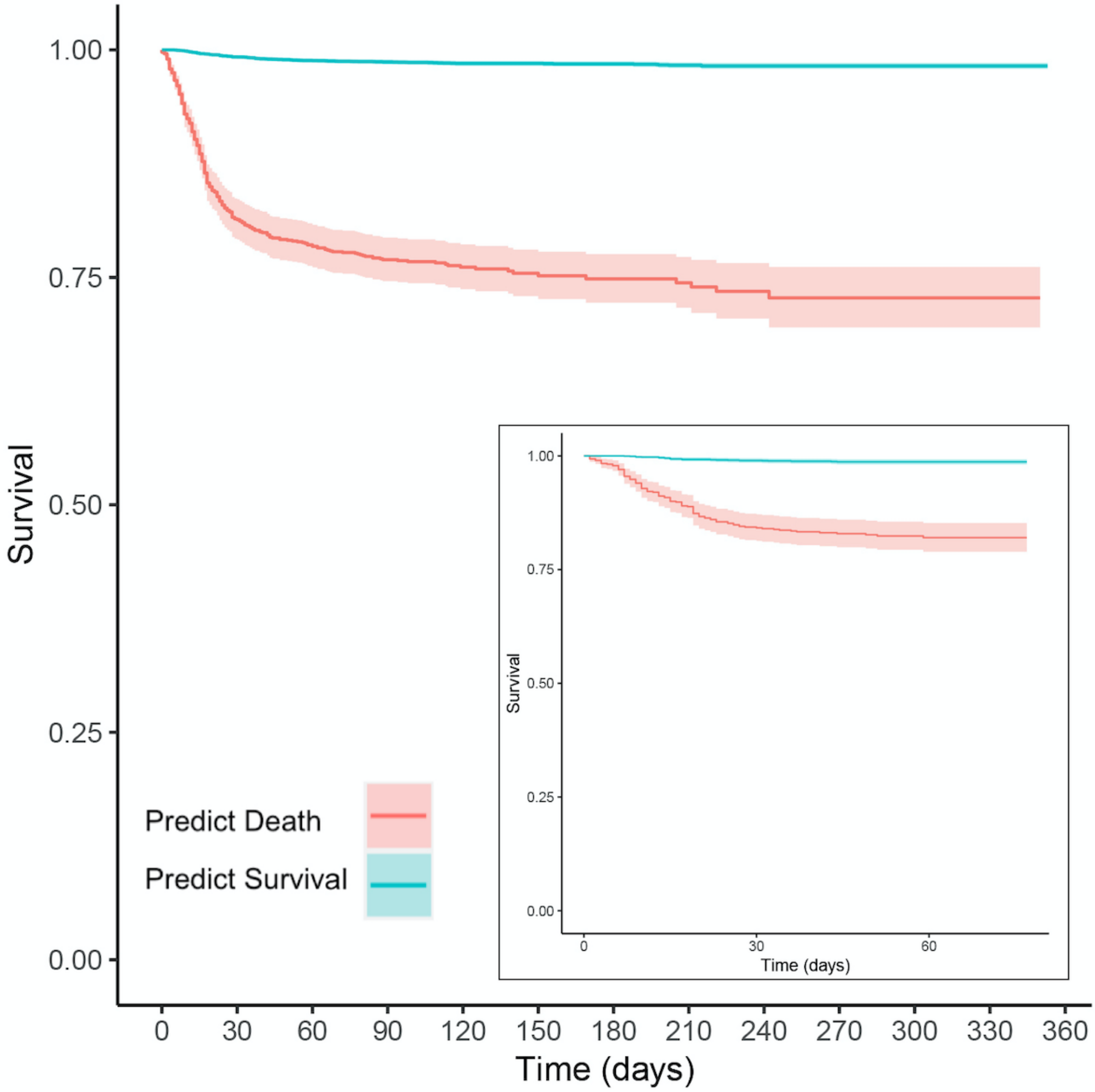


## Prospective Validation and Survival Analysis

To demonstrate the efficacy of our proposed mortality prediction, we performed a Kaplan-Meier analysis using the survival R library [17]. Specifically, we chose a decision boundary on the GRU-D model's ROC curve, which provided a specific delineation of high-risk and low-risk groups of patients. In our cross-validation cohort, binary classification provides accuracy of 89% (95% CI 88%-90%), recall of 80% (95% CI 74%-85%), precision of 17% (95% CI 15%-19%), and a negative predictive value of 99% (95% CI 99%-100%). Furthermore, although the precision is somewhat low with numerous false positives, we see among the survivors over twice the rate of mechanical ventilation or extracorporeal membrane oxygenation when they are predicted to die by GRU-D (Fisher exact test *P*<.001, odds ratio 2.1, 95% CI 1.8-2.5). We validated this performance in our prospective testing cohort, finding an AUROC of 0.901, accuracy of 78% (95% CI 76%-79%), recall of 85% (95% CI 77%-91%), precision of 14% (95% CI 12%-17%), and a negative predictive value of 99% (95% CI 99%-100%).

Our Kaplan-Meier analysis results in Figure 4 demonstrate the statistically significant stratification provided by our ML model in both the cross-validation and prospective testing experiments. Building a Cox Proportional Hazards model for our prediction in the cross-validation cohort provides a statistically significant difference in survival between the 2 groups (*P*<.001 for the likelihood ratio, logrank, and Wald tests), with a prediction of survival having a substantially improved hazard ratio of 0.053 (95% CI 0.043-0.066). We validated this finding in the prospective testing cohort with a statistically significant difference in survival between the 2 groups (*P*<.001 for the likelihood ratio, logrank, and Wald tests), with a prediction of survival having a substantially improved hazard ratio of 0.067 (95% CI 0.043-0.106).

**Figure 4.** Kaplan-Meier survival curves for the GRU-D stratified populations in the cross-validation cohort (main figure) and the prospective test cohort (inset), where teal is the prediction of low risk of death and red is the prediction of high risk. Both figures have 95% CIs visualized for the teal and red curves, although the teal confidence bands are tight due to our large sample sizes.



## Discussion

### Study Overview

In this study, we collected and processed over 50 laboratory and clinical covariates in a population of nearly 12,000 Mayo Clinic patients who tested positive for SARS-CoV-2 by PCR. In this large and geographically diverse data set, we found that the GRU-D RNN could provide state-of-the-art mortality prediction. This performance remained strong even in a held-out test set that mimics how a deployed system would be trained retrospectively and then prospectively utilized in a clinically evolving pandemic setting.

### Principal Results

Our cross-validation experiments summarized in Table 2 indicated that the top performing model to predict mortality in our cohort was the GRU-D RNN. We thus selected the GRU-D method to predict the mortality of patients with COVID-19 and prospectively found an AUROC of 0.901, accuracy of 78% (95% CI 76%-79%), recall of 85% (95% CI 77%-91%), precision of 14% (95% CI 12%-17%), negative predictive value of 99% (95% CI 99%-100%), and a statistically significant difference in survival (*P*<.001, hazard ratio for those predicted to survive, 95% CI 0.043-0.106). As can be expected in prospective validation, we observed a modest drop in AUROC although most of the performance characteristics were close to their original cross-validation estimates, that is, the negative predictive value was largely unchanged, while precision and accuracy showed minor decreases with the recall showing modest improvements.

We chose a prospective/retrospective split in time since this is the most realistic way to assess the potential performance of a system if launched clinically, because it would be trained on data up until its go-live date and then run prospectively in a potentially evolving pandemic environment. Notably, the cutoff date for the 80/20 split creating the prospective test set was December 15, 2020, which is the day after the first COVID vaccine received the United States Food and Drug Administration approval, meaning that our prospective cohort represented a distinctly different clinical environment compared to the period in which the model was trained. The relatively minor loss of performance in prospective validation shows the robustness of the modeling herein, but the observed loss of performance also demonstrates the need for continued retraining/validation of such a model during a constantly evolving pandemic.

## Limitations

The application and deployment of ML methods in clinical practice require concerted care and diligence. One may be inclined to interpret the high negative predictive value of our prediction algorithm as an indication that the best use of the algorithm in practice is as a screening mechanism to discharge patients who are not at risk in order to conserve resources for higher-risk individuals. However, such a conclusion illustrates a pitfall of using a correlative prediction algorithm to make causal conclusions. The algorithm is highly confident that under the current standards of care at Mayo Clinic, these individuals are not likely to succumb to their illness; this is quite distinct from asserting that it is safe to reduce the care for these patients. Arriving at this latter conclusion would likely require a randomized controlled trial, and given the much lower survival rate published in the New York City data set [11] where medical systems were overcapacity, it seems unlikely that reducing care from those who survived in our cohort would have been a safe measure. Because the Mayo Clinic health systems have not been overcapacity, our mortality predictions should be viewed as representing patient stratification when full clinical support is available.

Therefore, we conclude that the algorithm is better deployed as an alert system that flags only those patients it deems as high risk to provide the treating physician with an additional data point that aims to summarize the many covariates and the laboratory values routinely available. In this context, the algorithm has had abundant experience in the provider's system, effectively "seeing" all patients with COVID-19 that have attended Mayo Clinic and conveying these lessons to physicians who could not have gained such experience personally.

A web interface to this model may allow for widespread usage but given the complexity and error-prone nature of users providing the high dimensional time-series measurements with correct units, the system is better suited for integration within the EHR/LIS infrastructure. We are now exploring the details of deployment of such a GRU-D alert system, which involves discussions with physicians to assess numerous implementation details, for example, deciding whether the alerts would be passive EHR/chart-based flags or a direct page to the frontline clinical provider. Passive chart alerts are less intrusive to existing workflows (ie, a direct page interrupts a physician while tending to other patients) but also provide less-immediate feedback. Additionally, active alerts could also be sent to a triage group to consider if evaluation is needed (for example, from the registered respiratory therapist) rather than interrupting bedside clinicians. Furthermore, for either type of alert, there is the question of prescribing a bedside assessment or leaving it to provider discretion, which is again a matter of balancing disruption of the workflow with the likelihood of missing a critical event. There will not be a universally appropriate implementation for all hospital systems owing to staffing and procedural differences. However, since our algorithm predicts overall COVID-19 mortality and is not tailored to flag imminent events such as cardiopulmonary arrest, it may be appropriate to consider less intrusive chart alerts without prescribed bedside follow-ups.

We have also seen nuances in the challenges and opportunities presented by MNAR data. In the context of traditional statistical inference and imputation, MNAR data is a worst-case scenario so challenging that many practical applications effectively ignore the reality and proceed with algorithms designed for the missing completely at random or missing at random settings. A diligent statistician making this decision may perform a sensitivity analysis under a very limited set of assumed MNAR mechanisms to provide some assurances regarding the robustness of the chosen imputation or analytical strategy [18]. However, here we have demonstrated that classification problems can be quite distinct in this regard. Specifically, if the missing data mechanism is tightly coupled to the ultimate prediction task, it is entirely possible for MNAR data to be an asset rather than an impediment. One can construct a context where the class label is so tightly linked to the missing data mechanism that the patterns of missingness provide more discriminative power than the underlying values themselves (see Multimedia Appendix 2) [19]. In LIS systems, the number of potential laboratory tests that could be ordered at any time is astronomical, and it is unlikely that a practicing physician will ever order a "complete observation" of every test available on a single patient at every point in time. Instead, tests are ordered based on reasonable clinical suspicion that a test might return an abnormal result. From a prognostication point of view, this clinical suspicion is an enormously valuable piece of information that will almost never be captured in a structured data field in the EHR. If an algorithm cannot build off of this clinical suspicion as a starting point, it is also likely that its conclusions may appear to be a "step behind" the ordering clinician. Instead, an algorithm should learn what it can from the MNAR data patterns (here partly encoding clinical suspicion) in addition to the final value returned by the laboratory test.

We also note some of the real-world challenges that are faced when attempting to deploy such an alert system into clinical practice. First, in the retrospective experimental design followed here and by other papers in the literature, the time series data are constructed using the time of sample collection since this is the most biologically accurate way to represent the data and build predictive models. However, in practice, if there can be delays in the turnaround for certain tests, this will either result in delayed predictions (so that the deployed testing data match its retrospective training counterpart) or result in biased predictions when delayed laboratory test results are treated as missing. Therefore, although 72 hours is early in the course of illness, it is crucial that we have demonstrated reasonable performance even when only considering data collected on the same day as the first positive PCR result, because a real-world delay of 48 hours on certain laboratory test values may occur during a global pandemic, and thus, it is critical that the system can still provide accurate and timely predictions even when laboratory test results are delayed. Additionally, with vaccines now being delivered, the models presented herein should be considered as mortality predictions for an unvaccinated individual, and in practice, a vaccinated individual will be expected to be at low risk for mortality based on the clinical trials data.

Another challenge in dealing with LIS data comes from nonstandardization of test coding prior to reporting to the EHR. In a multisite system, the same laboratory test may have multiple test codes to account for the different ordering facilities or variability in local billing regulations. This creates the potential for discrepancies in the values stored within the underlying database such as differing units of measure. Substantial effort is therefore devoted to linking the LIS results to the EHR to ensure consistency across test codes and complete coverage of results in the EHR. The COVID-19 pandemic has created added complexity due to the rapidly evolving and continuously updating availability of COVID-19 nucleic acid and antibody tests. Therefore, effective data collection and deployment of ML methodologies necessitates extensive team-based laboratory and medical expertise to ensure that data aggregation and modeling efforts can be rapidly modified to suit the changing nature of the underlying data set. Scalability also presents practical challenges. This is illustrated by a scenario in which internal workflows began to fail due to limitations in the number of query results being returned by Tableau, necessitating that SQL queries take place on a high-performance computing cluster using a Python/Pandas toolchain. Although these logistical challenges may be of limited academic interest, they are important to document, as such barriers have been a greater impediment to rapid real world deployment than more traditional topics in the ML literature such as the identification of appropriate classification algorithms.

## Comparison With Prior Work

For context, in Table 4, we summarize some of the largest published COVID-19 mortality studies and specifically, the cohorts analyzed and the most relevant features identified. When smaller cohorts see insufficient numbers of deaths for direct mortality prediction, studies tend to focus on the prediction of severe outcomes. For instance, in a cohort of 123 patients with COVID-19 in Vulcan Hill Hospital, China, in the study of Pan et al [20], the mortality classifier based on XGBoost yielded an AUC of 0.86-0.92. Likewise, in a cohort of 372 Chinese cases (99.7% cohort survival rate), Gong et al [9] found that the following variables provided an AUROC of 0.85. Similarly, in a study of 375 patients with COVID-19 conducted by Ko et al [21], the mortality prediction model based on XGBoost had 92% accuracy. In a study of 398 COVID-19–positive patients by Abdulaal et al [22], 86% accuracy was achieved (95% CI 75%-93%). In a large study of 2160 cases over 54 days from 3 hospitals in Wuhan, China with sufficient cases to assess mortality (88% cohort survival rate), Gao et al [8] reported 0.92-0.98 as the AUROC using an ensemble classifier. Furthermore, Vaid et al [11] used 4098 inpatient cases over 68 days in New York City (83% cohort survival rate) to achieve an AUROC of 0.84-0.88 in mortality prediction. Kim et al [23] studied 4787 patients and their XGBoost-based classifier demonstrated an AUC of 0.88-0.89 (95% CI 0.85-0.91) in predicting the need for intensive care, which is distinct from mortality prediction. Bolourani et al [24] studied 11,525 patients to achieve an AUROC of 0.77 in predicting respiratory failure within 48 hours of admission, which is also distinct from mortality prediction, based on data from the emergency department by using an XGBoost model.

The dramatically different cohort mortality rates and the associated predictive accuracies may be in part due to the differing straining of the local health care systems at the time of study (both Wuhan and New York City experienced waves of patients that at different times overwhelmed the health care infrastructure), and the relatively geographically narrow nature of each of these data sets underscores why it is unlikely that these mortality predictions would extend directly to our patient population in a health care system spanning 3 time zones and multiple locales unrepresented in the literature.

**Table 4.** Summary of the related studies.

| Study | Patients (n) | Model/algorithm | Cohort survival | Prediction | Area under the receiver operating characteristic curve | Feature importance |
|---|---|---|---|---|---|---|
| Pan et al [20] | 123 | XGBoost | 52.8% | Mortality | 0.86-0.92 | Lymphocyte percentage, prothrombin time, lactate dehydrogenase, total bilirubin, eosinophil percentage, creatinine, neutrophil percentage, and albumin level |
| Gong et al [9] | 372 | Nomogram | 99.7% | Severity | 0.85 (95% CI 0.790-0.916) | Higher lactate dehydrogenase, C-reactive protein, red blood cell distribution width, direct bilirubin, blood urea nitrogen, and lower albumin |
| Ko et al [21] | 375 | XGBoost | 98.1% | Mortality | —,[a] accuracy of 92% | Not assessed |
| Abdulaal et al [22] | 398 | Artificial neural network | — | Mortality | —, accuracy of 86% (95% CI 75%-93%) | Altered mentation, dyspnea, age, collapse, gender, and cough |
| Shi et al [10] | 487 | Custom risk score calculation | 100% | Severity | — | Advanced age, presence of hypertension, and being male |
| Gao et al [8] | 2160 | Ensemble model based on logistic regression, gradient-boosted decision tree, neural network, and support vector machine | 88% | Mortality | 0.92-0.98 | Consciousness, chronic kidney disease, lymphocyte counts, sex, sputum, blood urea nitrogen, respiratory rate, oxygen saturation, D-dimer, number of comorbidities, albumin, age, fever, and platelet count |
| Vaid et al [11] | 4098 | XGBoost | 83% | Mortality | 0.84-0.88 | Age, anion gap, C-reactive protein, lactate dehydrogenase, oxygen saturation, blood urea nitrogen, ferritin, red cell distribution width, and diastolic blood pressure. |
| Kim et al [23] | 4787 | XGBoost | — | Need for intensive care | 0.88-0.89 | Activities of daily living, age, dyspnea, body temperature, sex, and underlying comorbidities |
| Bolourani et al [24] | 11,525 | XGBoost | — | Predicting respiratory failure | 0.77 | Invasive mode of oxygen delivery being a nonrebreather mask, emergency severity index values of 1 and 3, maximum respiratory rate, maximum, oxygen saturation, Black race, age on admission, eosinophil percentage, serum sodium level, and serum lactate level. |
| This study | 11,807 | GRU-D | 95.4% | Mortality | 0.938 cross-validation; 0.901 prospectively | Figure 2, top 5: age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen level, and serum iron level |

[a] Not available.

As indicated in Table 4, this study represents the largest cohort collected for mortality prediction in COVID-19, and the GRU-D algorithm shows state-of-the-art performance. Notably, many papers selected models based on XGBoost, which also showed strong cross-validation performance in our data. However, Table 2 demonstrates that XGBoost was not even in the top 5 algorithms that we assessed. Additionally, in agreement with Gao et al [8], we find that ensemble algorithms such as AutoGluon can provide stronger performance, although as noted previously, the GRU-D algorithm ended up ranked most highly in our cross-validation experiments.

## Conclusions

We have aggregated and analyzed one of the largest multistate COVID-19 EHR databases for mortality prediction. Using this database, we have trained and prospectively validated a highly effective ML algorithm using the GRU-D neural network architecture to predict the mortality of patients with COVID-19 shortly after their first positive PCR test result.

## Acknowledgments

We thank the Advanced Diagnostics Laboratory, Department of Laboratory Medicine and Pathology, and the Center for Individualized Medicine at Mayo Clinic for funding this research. We thank Dr Nicholas Chia for his insightful discussions and feedback on this manuscript. We are grateful to the thorough and constructive comments of the anonymous reviewers whose feedback greatly improved our paper.

## Conflicts of Interest

JCO has equity in Doximity, Inc, and has received small grants from Nference, Inc, and personal fees from Bates College. None of these are related to the current paper.

## Multimedia Appendix 1

Error analysis.
[DOCX File , 496 KB-Multimedia Appendix 1]

## Multimedia Appendix 2

Monte Carlo missing-not-at-random simulation.
[DOCX File , 592 KB-Multimedia Appendix 2]

## References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
2. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
3. WHO Coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-08-11]
4. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
5. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet 2020 Apr 11;395(10231):1225-1228 [FREE Full text] [doi: 10.1016/S0140-6736(20)30627-9] [Medline: 32178769]
6. Rosenbaum L. Facing Covid-19 in Italy - Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. N Engl J Med 2020 May 14;382(20):1873-1875. [doi: 10.1056/NEJMp2005492] [Medline: 32187459]
7. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
8. Gao Y, Cai G, Fang W, Li H, Wang S, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020 Oct 06;11(1):5033 [FREE Full text] [doi: 10.1038/s41467-020-18684-2] [Medline: 33024092]
9. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]
10. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020 Mar 18;24(1):108 [FREE Full text] [doi: 10.1186/s13054-020-2833-7] [Medline: 32188484]
11. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
12. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
13. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]

14. Kumar A, Arora A, Sharma P, Anikhindi SA, Bansal N, Singla V, et al. Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. Diabetes Metab Syndr 2020;14(4):535-545 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.044] [Medline: 32408118]
15. Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, et al. AutoGluon-Tabular: robust and accurate AutoML for structured data. 2020 Presented at: 7th ICML Workshop on Automated Machine Learning; July 12-18; Virtual Location URL: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_7.pdf
16. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. Sci Rep 2018 Apr 17;8(1):6085 [FREE Full text] [doi: 10.1038/s41598-018-24271-9] [Medline: 29666385]
17. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. New York, USA: Springer; 2000:7-13.
18. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Stat Med 2020 Sep 20;39(21):2815-2842. [doi: 10.1002/sim.8569] [Medline: 32419182]
19. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
20. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. J Med Internet Res 2020 Nov 11;22(11):e23128 [FREE Full text] [doi: 10.2196/23128] [Medline: 33035175]
21. Ko H, Chung H, Kang WS, Park C, Kim DW, Kim SE, et al. An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model. J Med Internet Res 2020 Dec 23;22(12):e25442 [FREE Full text] [doi: 10.2196/25442] [Medline: 33301414]
22. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. J Med Internet Res 2020 Aug 25;22(8):e20259 [FREE Full text] [doi: 10.2196/20259] [Medline: 32735549]
23. Kim H, Han D, Kim J, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. J Med Internet Res 2020 Nov 09;22(11):e24225 [FREE Full text] [doi: 10.2196/24225] [Medline: 33108316]
24. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. J Med Internet Res 2021 Feb 10;23(2):e24246 [FREE Full text] [doi: 10.2196/24246] [Medline: 33476281]

## Abbreviations

**AUROC:** area under the receiver operator characteristic
**EHR:** electronic health record
**GRU:** gated recurrent unit
**LIS:** laboratory information system
**ML:** machine learning
**MNAR:** missing-not-at-random
**PCR:** polymerase chain reaction
**RNN:** recurrent neural network

*Edited by C Basch; submitted 03.05.21; peer-reviewed by HJ Kim, S Lalmuanawma; comments to author 28.06.21; revised version received 18.07.21; accepted 11.08.21; published 28.09.21*

*Please cite as:*
*Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, Osborne C, Oliver GR, Lesko J, Bates KL, Khezeli K, Block DR, DiGuardo M, Kreuter J, O'Horo JC, Kalantari J, Klee EW, Salama ME, Kipp B, Morice WG, Jenkinson G*
*COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation*
*J Med Internet Res 2021;23(9):e30157*
*URL: https://www.jmir.org/2021/9/e30157*
*doi: 10.2196/30157*
*PMID: 34449401*

Original Paper

# Using Administrative Data to Explore Potentially Aberrant Provision of Virtual Care During COVID-19: Retrospective Cohort Study of Ontario Provincial Data

Vess Stamenova[1], PhD; Cherry Chu[1], MSc; Andrea Pang[2], MPH; Mina Tadrous[1,2,3], MSc, PhD, PharmD; R Sacha Bhatia[1,2,4], MD, MBA; Peter Cram[2,5], MD, MBA

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

[3]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

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

[5]School of Medicine, University of Texas Medical Branch, Galveston, TX, United States

**Corresponding Author:**
Vess Stamenova, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 268 0985
Email: vess.stamenova@wchospital.ca

## Abstract

**Background:** The COVID-19 pandemic has led to a rapid increase in virtual care use across the globe. Many health care systems have responded by creating virtual care billing codes that allow physicians to see their patients over telephone or video. This rapid liberalization of billing requirements, both in Canada and other countries, has led to concerns about potential abuse, but empirical data are limited.

**Objective:** The objectives of this study were to examine whether there were substantial changes in physicians' ambulatory visit volumes coinciding with the liberalization of virtual care billing rules and to describe the characteristics of physicians who significantly increased their ambulatory visit volumes during this period. We also sought to describe the relationship between visit volume changes in 2020 and the volumes of virtual care use among individual physicians and across specialties.

**Methods:** We conducted a population-based, retrospective cohort study using health administrative data from the Ontario Health Insurance Plan, which was linked to the ICES Physician Database. We identified a unique cohort of providers based on physicians' billings and calculated the ratio of total in-person and virtual ambulatory visits over the period from January to June 2020 (virtual predominating) relative to that over the period from January to June 2019 (in-person predominating) for each physician. Based on these ratios, we then stratified physicians into four groups: low-, same-, high-, and very high–use physicians. We then calculated various demographic and practice characteristics of physicians in each group.

**Results:** Among 28,383 eligible physicians in 2020, the mean ratio of ambulatory visits in January to June 2020:2019 was 0.99 (SD 2.53; median 0.81, IQR 0.59-1.0). Out of 28,383 physicians, only 2672 (9.4%) fell into the high-use group and only 291 (1.0%) fell into the very high–use group. High-use physicians were younger, more recent graduates, more likely female, and less likely to be international graduates. They also had, on average, lower-volume practices. There was a significant positive correlation between percent virtual care and the 2020:2019 ratio only in the group of physicians who maintained their practice ($R$=0.35, $P$<.001). There was also a significant positive correlation between the 2020:2019 ratio and the percent virtual care per specialty ($R$=0.59, $P$<.01).

**Conclusions:** During the early stages of the pandemic, the introduction of virtual care did not lead to significant increases in visit volume. Our results provide reassuring evidence that relaxation of billing requirements early in the COVID-19 pandemic in Ontario were not associated with widespread and aberrant billing behaviors. Furthermore, the strong relationship between the ability to maintain practice volumes and the use of virtual care suggests that the introduction of virtual care allowed for continued access to care for patients.

***(J Med Internet Res 2021;23(9):e29396)*** doi: 10.2196/29396

**KEYWORDS**

telemedicine; virtual care; COVID-19; pandemic; virtual health; telehealth; ambulatory visits; physicians; patients; digital health

## Introduction

The COVID-19 pandemic has led to a rapid increase in virtual care use across the globe [1-5]. In Ontario, Canada's largest province, virtual care increased from 1.6% of all ambulatory visits pre–COVID-19 to 71% during the first wave of the COVID-19 pandemic [3], a much higher rate compared to those reported in other countries such as the United States (30%) and Australia (42%) [4,6,7].

While Ontario had pre-existing virtual care billing codes before the onset of the pandemic, these codes were allowable for a single government-run online platform and only available to specialists and primary care physicians in rostered patient practices or specialized practices. Primary care physicians outside of rostered practices were not included in this model, in order to support continuity of care [8] and respond to growing concerns about fragmentation and poor quality of care received in virtual walk-in clinics as well as funding disruptions in Canada [9] and abroad [10].

In Ontario, the pandemic led to the introduction of temporary billing codes in mid-March 2020 that reimbursed any physician with identical amounts for in-person, video, or telephone visits and eliminated prior restrictions on practice type or allowable technology platforms. This rapid liberalization of billing requirements, both in Canada and other countries [4], has led to concerns about potential abuse, but empirical data are limited [11].

The objectives of this study were to examine whether there were substantial changes in physicians' ambulatory visit volumes coinciding with the liberalization of virtual care billing rules and to describe the characteristics of physicians who significantly increased their ambulatory visit volumes during this period. We also sought to describe the relationship between visit volume changes in 2020 and the volumes of virtual care use in individual physicians and across specialties.

## Methods

We conducted a population-based, retrospective cohort study using health administrative data from the Ontario Health Insurance Plan, which was linked to the ICES Physician Database. Data sets were linked using unique encoded identifiers and analyzed at ICES, an independent, nonprofit research institute. Use of these databases for the purposes of this study was authorized under §45 of Ontario's Personal Health Information Protection Act, which does not require review by a research ethics board. An exemption letter was obtained by the Research Ethics Board at Women's College Hospital, Toronto, Ontario.

We identified a unique cohort of providers based on physicians' billings for in-person and virtual ambulatory visits. We excluded visits for non-Ontario residents and those with an invalid or missing health card number. We also excluded all physicians with clinical volumes that were inconsistent with an active practice during the pre–COVID-19 period (<10 ambulatory visits during the period from January to June 2019).

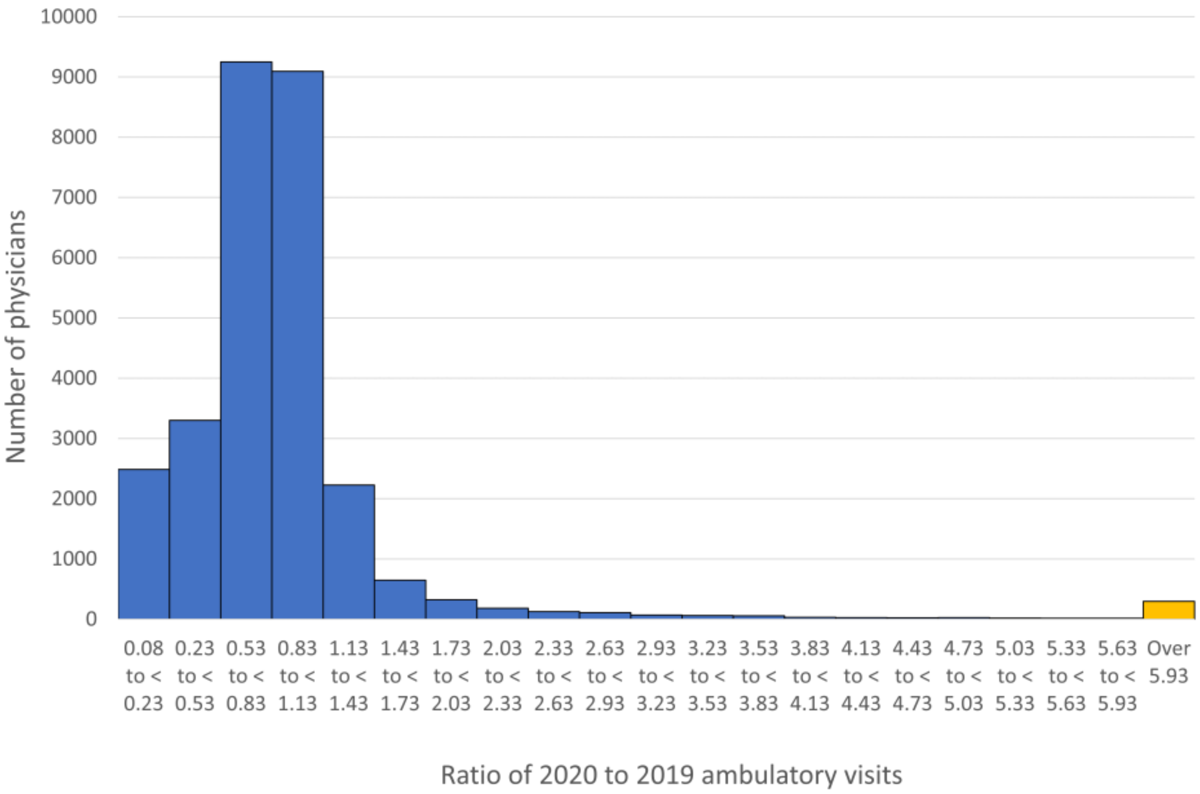
We then calculated the ratio of total in-person and virtual ambulatory visits over the period from January to June 2020 (virtual predominating) relative to that over the period from January to June 2019 (in-person predominating) for each physician. We included first-quarter data in 2020 as they cover the beginning of the pandemic. Data extending past the second quarter of 2020 were unavailable. Based on these ratios, we then stratified physicians into four groups: (1) low-use physicians had ratios from 0 to 0.50 (ie, a 50% or greater reduction in visits in 2020 compared to 2019), (2) same-use physicians were those with ratios over 0.50 but less than 1.25, (3) high-use physicians were those with ratios of at least 1.25 but less than 6.0, and (4) very high–use physicians were those with ratios equal to or greater than 6.0 (ie, an at-least 6-fold increase in visits in 2020 compared to 2019). To explore whether the proportions of physicians falling into each category differed much from previous years, we also calculated the number of physicians falling into each group—defined as the same ratio ranges—for the periods of January to June 2019 relative to January to June 2018.

For all physicians, we also obtained demographic and practice characteristics, including, age, sex, years since graduation, training location, practice type (ie, specialist, family practice, or focused family practice physicians focusing 50% or more of their practice in a specific type of care, such as psychotherapy [12]), and specialty. We also calculated the number of unique patients seen, number of total visits, number of virtual visits, and number of visits per day, virtual or any.

## Results

Among 28,383 eligible physicians in 2020, the mean ratio of ambulatory visits in January to June 2020:2019 was 0.99 (SD 2.53; median 0.81, IQR 0.59-1.0). Only 291 physicians (1.0%) were very high users, 2672 physicians (9.4%) were high users, and 5422 (19.1%) were low users (Figure 1). In comparison, the previous year (2019:2018 visit ratio), among 27,709 eligible physicians, 289 (1.0%) were very high users, 3395 (12.3%) were high users, and 2937 (10.6%) were low users.

**Figure 1.** Histogram showing the total number of physicians by ambulatory visit volume ratio (2020:2019). Proportions of <1 indicate fewer visits in the period of January to June 2020 when compared to January to June 2019, while a proportion of >1 suggested increased visits in the period of January to June 2020. Note, the last bin in orange represents anybody with a ratio over 5.93, representing the top 1%.



High-use physicians were younger, more recent graduates, more likely to be female, and less likely to be international graduates than those who maintained their volumes (ie, same-use group) (*P*<.001) (Table 1). These effects were even more pronounced in the very high–use group. They were also more likely to be specialists than primary care providers (*P*<.001), and physicians in focused primary care practices were more likely to be high or very high users relative to specialists (*P*<.001).

Physicians in the two high-use groups had, on average, lower-volume practices in both 2020 and 2019 (*P*<.001), and volumes were especially smaller in 2019 with fewer patients seen (*P*<.001). The effect was larger for the very high users (*P*<.001). Providers in the two high-use groups had fewer visits per billing day and fewer total patients seen in 2020 (*P*<.001), but they had a higher percentage of virtual care visits (*P*<.001) and a similar number of virtual care visits per day compared to the same-use group (Table 1).

Specialties with a large percentage of their total physician population being in the high-use group included emergency medicine (41/230, 17.8%), psychiatry (378/2061, 18.3%), and internal medicine (147/999, 14.7%). The same specialties were common among the very high–use group (Table 2).

Pearson correlations between the 2020:2019 visit ratio and percent of visits completed virtually showed a significant positive correlation only among physicians from the same-use group (*R*=0.35, *P*<.001) (Figure 2).

Finally, we also calculated the Pearson correlation coefficient between the average 2020:2019 ratio per specialty and the percent virtual care used per specialty. We excluded emergency medicine, diagnostic radiology, and nuclear medicine, as they were outliers and had both the highest 2020:2019 visit ratios (1.6, 1.3, and 2.9, respectively) and the lowest percentages of virtual care (12.4%, 1.5%, and 8.2%, respectively). There was a significant positive correlation between the 2020:2019 ratio and the percent virtual care across specialties (*R*=0.59, *P*<.01) (Figure 3).

**Table 1.** Physician characteristics stratified by ratio of ambulatory visits in January to June 2020:2019.

| Characteristic | Physicians | | | | | *P* value |
|---|---|---|---|---|---|---|
| | All (N=28,383) | Low use: 0 to 0.50 visits[a] (n=5422) | Same use: >0.50 to <1.25 visits (n=19,998) | High use: 1.25 to <6.0 visits[b] (n=2672) | Very high use: ≥6.0 visits (n=291) | |
| Age (years), mean (SD) | 53.9 (11.2) | 56.1 (12.6) | 53.7 (10.8) | 51.9 (11.24) | 49.2 (11.5) | <.001 |
| Years since graduation, mean (SD) | 25.1 (13.3) | 26.7 (15.5) | 25.4 (12.5) | 20.5 (12.60) | 15.8 (11.7) | <.001 |
| Sex (female), n (%) | 11,864 (41.8) | 2287 (42.5) | 8116 (40.6) | 1268 (48.4) | 148 (55.4) | <.001 |
| Canadian or internationally trained (Canadian trained), n (%) | 16,218 (76.4) | 2928 (77.2) | 11,912 (75.7) | 1282 (81.6) | 96 (83.5) | <.001 |
| **Number of unique patients seen in 2019 (January to June)** | | | | | | |
| Mean (SD) | 729.2 (795.0) | 496.53 (675.0) | 858.28 (828.8) | 308.81 (433.4) | 55.46 (60.5) | <.001 |
| Median (IQR) | 539 (163-999) | 257 (75-689) | 691 (295-1129) | 153 (50-402) | 35 (17-72) | <.001 |
| **Number of unique patients seen in 2020 (January to June)** | | | | | | |
| Mean (SD) | 556.7 (618.1) | 155.2 (297.0) | 684.5 (639.5) | 428.0 (577.4) | 439.3 (408.9) | <.001 |
| Median (IQR) | 408 (100-789) | 35 (0-183) | 559 (241-910) | 236 (81-567) | 347 (133-590) | <.001 |
| **Percent virtual care visits in 2020 (January to June)** | | | | | | |
| Mean (SD) | 31.9 (22.7) | 14.95 (24.3) | 34.03 (20.0) | 40.3 (26.2) | 42.3 (28.6) | <.001 |
| Median (IQR) | 34 (11-49) | 0 (0-22) | 37 (19-49) | 44 (18-60) | 42 (21-61) | <.001 |
| **Number of virtual care visits in 2020 (January to June)** | | | | | | |
| Mean (SD) | 400.7 (571.0) | 52.1 (139.5) | 471.3 (570.6) | 411.9 (771.3) | 294.8 (315.1) | <.001 |
| Median (IQR) | 231 (30-551) | 1 (0-37) | 324 (98-639) | 199 (32-494) | 205 (70-432) | <.001 |
| **Number of total visits in 2020 (January to June)** | | | | | | |
| Mean (SD) | 1101.2 (1228.9) | 292.2 (484.9) | 1300.0 (1257.7) | 877.9 (1293.1) | 738.0 (655.7) | <.001 |
| Median (IQR) | 780 (240-1503) | 108 (23-370) | 1004 (454-1715) | 495 (151-1105) | 538 (274-976) | <.001 |
| **Number of total visits in 2019 (January to June)** | | | | | | |
| Mean (SD) | 1304.8 (1420.4) | 800.5 (1103.6) | 1561.5 (1483.3) | 541.4 (855.9) | 66.1 (73.7) | <.001 |
| Median (IQR) | 934 (286-1799) | 392 (101-1128) | 1212 (539-2078) | 267 (74-658) | 43 (20-84) | <.001 |
| **Number of visits per billing day in 2020 (January to June)** | | | | | | |
| Mean (SD) | 12.1 (9.5) | 8.1 (8.3) | 13.2 (9.5) | 9.7 (9.8) | 10.2 (7.3) | <.001 |
| Median (IQR) | 9 (5-15) | 5 (2-10) | 11 (6-16) | 7 (3-12) | 7 (4-12) | <.001 |
| **Number of virtual visits per billing day in 2020 (January to June)** | | | | | | |
| Mean (SD) | 8.6 (6.7) | 5.0 (5.2) | 9.1 (6.5) | 8.1 (7.6) | 7.4 (5.4) | <.001 |
| Median (IQR) | 7 (4-11) | 3 (2-6) | 7 (4-11) | 6 (3-10) | 6 (3-9) | <.001 |
| **Practice type, n (%)** | | | | | | |
| Specialist | 15,201 (53.6) | 3058 (56.4) | 10,419 (52.1) | 1590 (59.5) | 134 (46.0) | <.001 |
| Primary care provider | 9393 (33.1) | 1366 (25.2) | 7560 (37.8) | 436 (16.3) | 31 (10.7) | N/A[c] |
| Focused primary care provider[d] | 426 (1.5) | 120 (2.2) | 190 (1.0) | 97 (3.6) | 19 (6.5) | N/A |
| Miscellaneous | 3363 (50.0) | 878 (13.1) | 1829 (27.2) | 549 (8.2) | 107 (1.6) | N/A |

[a] Visits of 0 to 0.50 correspond to physicians who had a 50% or greater reduction in ambulatory visits between 2019 and 2020.

[b] Visits of 1.25 to <6 correspond to physicians who had a 25% 6-fold increase in visits between 2019 and 2020.

[c] N/A: not applicable; a single test was conducted across all four groups in this section and the *P* value is reported in the row for the first group.

[d] Focused primary care providers are primary care providers who specialize in a specific care (eg, palliative care).

**Table 2.** Physicians per specialty across each user group.

| Specialty | Low use: 0 to 0.5, n (%) | Same use: >0.5 to <1.25, n (%) | High use: 1.25 to <6.0, n (%) | Very high use: ≥6.0, n (%) |
|---|---|---|---|---|
| Family medicine (n=13,244) | 2366 (17.9) | 9715 (73.4) | 1042 (7.9) | 121 (0.9) |
| **Medicine** | | | | |
| Emergency medicine (n=230) | 57 (24.8) | 126 (54.8) | 41 (17.8) | 6 (2.6) |
| Internal medicine (n=999) | 233 (23.3) | 604 (60.5) | 147 (14.7) | 15 (1.5) |
| Infectious diseases (n=148) | 27 (18.2) | 101 (68.2) | 20 (13.5) | 0 (0) |
| Critical care (n=99) | 29 (29.3) | 53 (53.5) | 13 (13.1) | ≤5 (≤5.1) |
| Endocrinology (n=269) | 18 (6.7) | 215 (79.9) | 34 (12.6) | ≤5 (≤1.9) |
| Nuclear medicine (n=48) | 15 (31.3) | 24 (50.0) | 6 (12.5) | ≤5 (≤10.4) |
| Hematology (n=217) | 23 (10.6) | 164 (75.6) | 27 (12.4) | ≤5 (≤2.3) |
| Cardiology (n=684) | 77 (11.3) | 527 (77.0) | 78 (11.4) | ≤5 (≤0.7) |
| Respirology (n=308) | 32 (10.4) | 237 (76.9) | 35 (11.4) | ≤5 (≤1.6) |
| Geriatric medicine (n=152) | 29 (19.1) | 105 (69.1) | 16 (10.5) | ≤5 (≤3.3) |
| Rheumatology (n=213) | 19 (8.9) | 172 (80.8) | 22 (10.3) | 0 (0) |
| Anesthesiology (n=1179) | 480 (40.7) | 597 (50.6) | 101 (8.6) | ≤5 (≤0.4) |
| Nephrology (n=242) | 23 (9.5) | 198 (81.8) | 20 (8.3) | ≤5 (≤2.1) |
| Clinical immunology (n=84) | 22 (26.2) | 56 (66.7) | 6 (7.1) | 0 (0) |
| Gastroenterology (n=345) | 39 (11.3) | 281 (81.4) | 24 (7.0) | ≤5 (≤1.4) |
| Obstetrics and gynecology (n=808) | 98 (12.1) | 659 (81.6) | 45 (5.6) | 6 (0.7) |
| **Other specialties** | | | | |
| Psychiatry (n=2061) | 328 (15.9) | 1326 (64.3) | 378 (18.3) | 29 (1.4) |
| Diagnostic radiology (n=624) | 179 (28.7) | 355 (56.9) | 84 (13.5) | 6 (1.0) |
| Medical oncology (n=269) | 34 (12.6) | 200 (74.3) | 35 (13.0) | 0 (0) |
| Pediatrics (n=1487) | 395 (26.6) | 938 (63.1) | 146 (9.8) | 8 (0.5) |
| Radiation oncology (n=211) | 14 (6.6) | 178 (84.4) | 19 (9.0) | 0 (0) |
| Neurology (n=410) | 53 (12.9) | 321 (78.3) | 33 (8.0) | ≤5 (≤1.2) |
| Physical medicine and rehabilitation (n=211) | 48 (22.7) | 146 (69.2) | 16 (7.6) | ≤5 (≤2.4) |
| Dermatology (n=236) | 52 (22.0) | 173 (73.3) | 11 (4.7) | 0 (0) |
| Surgery (n=2868) | 536 (18.7) | 2212 (77.1) | 104 (3.6) | 16 (0.6) |
| Remaining smaller specialties (n=384) | 91 (23.7) | 246 (64.1) | 43 (11.2) | 4 (1.0) |
| Miscellaneous (n=353) | 105 (29.7) | 69 (19.5) | 126 (35.7) | 53 (15.0) |

**Figure 2.** Correlation between the 2020:2019 visit ratio and percent virtual care in the four groups of providers: those who reduced (0 to 0.50), maintained (>0.5 to <1.25), increased (1.25 to <6), and significantly increased (≥6) their practice in 2020 relative to 2019.

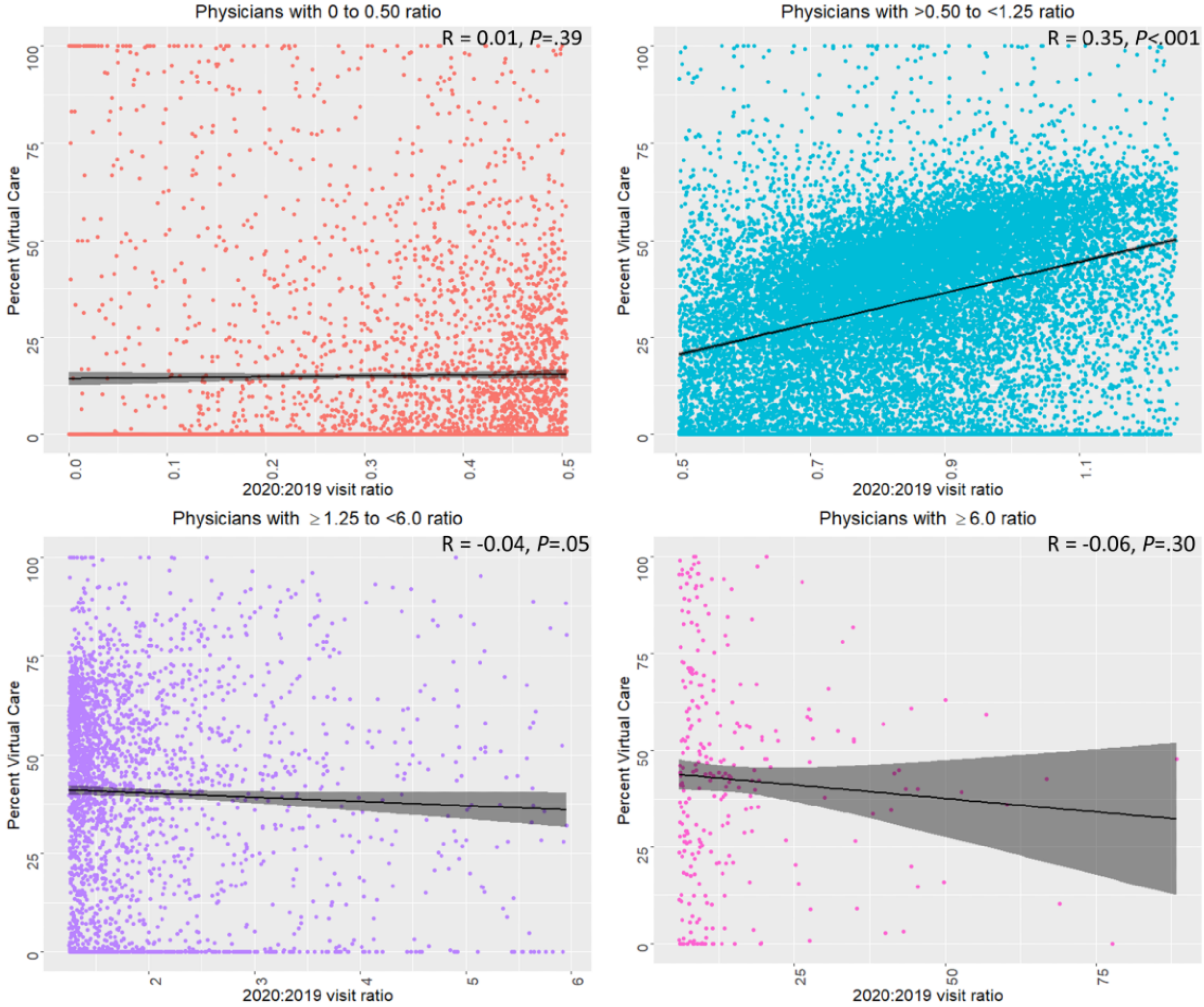
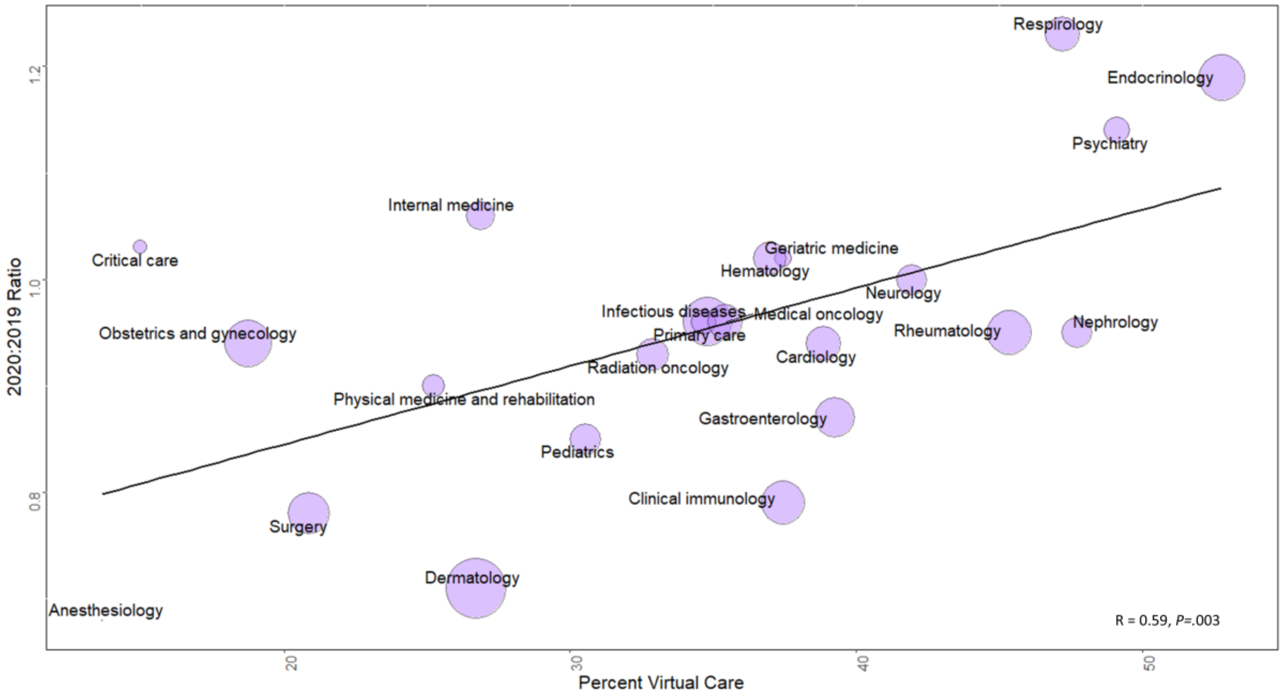


**Figure 3.** Correlation between the 2020:2019 visit ratio and virtual care adoption across specialties. The size of each sphere indicates the number of visits completed.



## Discussion

During the early stages of the pandemic, the introduction of virtual care did not lead to significant increases in visit volume. Only about 10% of physicians increased their visit volumes by 25% or more in 2020 relative to 2019. In total, our results provide reassuring evidence that relaxation of billing requirements early in the COVID-19 pandemic in Ontario were not associated with widespread and aberrant billing behaviors.

Providers who increased their visit volumes tended to be specialists, younger, more recent graduates, and more likely female. Among providers who increased their practice volumes, there was no relationship between the magnitude of increase and virtual care adoption. A significant relationship was observed, however, among providers who maintained their practice. This relationship was also maintained at the specialty level. Endocrinology, respirology, and psychiatry maintained their practices the best and had higher rates of virtual care adoption.

Our results are consistent with data from the United States that showed that despite the introduction of virtual care, overall visit volumes decreased in the early periods of the COVID-19 pandemic [6]. In fact, the introduction of virtual care during the pandemic allowed physicians to maintain their practices. Higher rates of virtual care use among providers who maintained their practice volumes were associated with better maintenance of visit volumes during the pandemic. This trend was also observed in the United States [6]. Here, we confirm these findings with an analysis of the entire physician and patient population in a health care system with a single insurance plan where the introduction of virtual care payment policies occurred at the same time for the entire population.

At least two specialties that showed high virtual care adoption rates and good maintenance of visit volumes during the pandemic were consistent in both Ontario and the United States [6]: psychiatry and endocrinology. Mental health care has the potential to be better suited for virtual care as it often does not require a physical exam and it has been successful in adopting virtual care services both before [13] and after the pandemic [14]. Successful adoption in endocrinology during the pandemic has also been reported [15].

Limitations to our study include a relatively brief time window for evaluating the impact of billing code liberalization, which makes it unclear whether the trends will be maintained in the long term. Our reliance on administrative data also precludes us from robustly evaluating appropriateness of individual visits.

In total, our study suggests that liberalization of virtual care billing requirements coinciding with the COVID-19 pandemic was not associated with an alarming increase in individual physician visit volumes and should serve to assuage concerns over widespread fraud. Furthermore, the strong relationship between the ability to maintain practice volumes and the use of virtual care suggest that the introduction of virtual care allowed continued access to care for patients.

### Conflicts of Interest

None declared.

### References

1. Baum A, Kaboli PJ, Schwartz MD. Reduced in-person and increased telehealth outpatient visits during the COVID-19 pandemic. Ann Intern Med 2021 Jan;174(1):129-131. [doi: 10.7326/m20-3026]
2. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
3. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: A repeated cross-sectional study. CMAJ Open 2021;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
4. Mehrotra A, Bhatia RS, Snoswell CL. Paying for telemedicine after the pandemic. JAMA 2021 Feb 02;325(5):431-432. [doi: 10.1001/jama.2020.25706] [Medline: 33528545]
5. Digital transformation: Shaping the future of European healthcare. Deloitte United Kingdom.: Deloitte Centre for Health Solutions; 2020 Sep. URL: https://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/european-digital-health.html [accessed 2021-03-01]
6. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
7. 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]
8. OMA's Virtual Care FAQs. Toronto, ON: Ontario Medical Association; 2019 Nov 13. URL: https://ontariorheum.ca/wp-content/uploads/Virtual-Care-OTN-FAQs-Nov-2019.pdf [accessed 2021-08-27]
9. Glauser W. Virtual care has potential to fragment primary care and disturb continuity of care, warn doctors. CMAJ News. 2019 Aug 29. URL: http://cmajnews.com/2019/08/29/virtual-care-has-potential-to-fragment-primary-care-and-disturb-continuity-of-care-warn-doctors/ [accessed 2021-03-04]

10. O'Dowd A. Doctors question Hancock's idea of GP video consultations for all. BMJ 2018 Sep 14;362:k3934. [doi: 10.1136/bmj.k3934] [Medline: 30217932]
11. Verma S. Early impact of CMS expansion of Medicare telehealth during COVID-19. Health Affairs Blog. 2020 Jul 15. URL: https://www.healthaffairs.org/do/10.1377/hblog20200715.454789/full/ [accessed 2021-04-01]
12. Schultz SE, Glazier RH. Identification of physicians providing comprehensive primary care in Ontario: A retrospective analysis using linked administrative data. CMAJ Open 2017 Dec 19;5(4):E856-E863 [FREE Full text] [doi: 10.9778/cmajo.20170083] [Medline: 29259018]
13. Mehrotra A, Huskamp HA, Souza J, Uscher-Pines L, Rose S, Landon BE, et al. Rapid growth in mental health telemedicine use among rural Medicare beneficiaries, wide variation across states. Health Aff (Millwood) 2017 May 01;36(5):909-917. [doi: 10.1377/hlthaff.2016.1461] [Medline: 28461359]
14. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid increase in telemental health within the Department of Veterans Affairs during the COVID-19 pandemic. Telemed J E Health 2021 Apr;27(4):454-458. [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
15. Madduri S, Chowdhary R, Sethu Reddy S. Telehealth adoption among endocrinologists during the COVID-19 pandemic. Endocr Pract 2020 Aug;26(8):846-856. [doi: 10.4158/ep-2020-0237]

*Edited by C Basch; submitted 05.04.21; peer-reviewed by E van der Velde, P Yellowlees; comments to author 21.04.21; revised version received 30.04.21; accepted 05.05.21; published 07.09.21*

*Please cite as:*
*Stamenova V, Chu C, Pang A, Tadrous M, Bhatia RS, Cram P*
*Using Administrative Data to Explore Potentially Aberrant Provision of Virtual Care During COVID-19: Retrospective Cohort Study of Ontario Provincial Data*
*J Med Internet Res 2021;23(9):e29396*
*URL: https://www.jmir.org/2021/9/e29396*
*doi: 10.2196/29396*
*PMID: 34313590*

Original Paper

# Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach

Chien-Yu Chi[1*], MD; Shuang Ao[2*], PhD; Adrian Winkler[2], BEng; Kuan-Chun Fu[2], MSc; Jie Xu[2], PhD; Yi-Lwun Ho[3], MD; Chien-Hua Huang[4], MD; Rohollah Soltani[2], MSc

[1]Department of Emergency Medicine, Yunlin Branch, National Taiwan University Hospital, Yunlin, Taiwan
[2]Knowtions Research, Toronto, ON, Canada
[3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[4]Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan
[*]these authors contributed equally

**Corresponding Author:**
Chien-Hua Huang, MD
Department of Emergency Medicine
National Taiwan University Hospital
#7 Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 0972651304
Email: chhuang5940@ntu.edu.tw

## Abstract

**Background:** In-hospital cardiac arrest (IHCA) is associated with high mortality and health care costs in the recovery phase. Predicting adverse outcome events, including readmission, improves the chance for appropriate interventions and reduces health care costs. However, studies related to the early prediction of adverse events of IHCA survivors are rare. Therefore, we used a deep learning model for prediction in this study.

**Objective:** This study aimed to demonstrate that with the proper data set and learning strategies, we can predict the 30-day mortality and readmission of IHCA survivors based on their historical claims.

**Methods:** National Health Insurance Research Database claims data, including 168,693 patients who had experienced IHCA at least once and 1,569,478 clinical records, were obtained to generate a data set for outcome prediction. We predicted the 30-day mortality/readmission after each current record (ALL-mortality/ALL-readmission) and 30-day mortality/readmission after IHCA (cardiac arrest [CA]-mortality/CA-readmission). We developed a hierarchical vectorizer (HVec) deep learning model to extract patients' information and predict mortality and readmission. To embed the textual medical concepts of the clinical records into our deep learning model, we used Text2Node to compute the distributed representations of all medical concept codes as a 128-dimensional vector. Along with the patient's demographic information, our novel HVec model generated embedding vectors to hierarchically describe the health status at the record-level and patient-level. Multitask learning involving two main tasks and auxiliary tasks was proposed. As CA-mortality and CA-readmission were rare, person upsampling of patients with CA and weighting of CA records were used to improve prediction performance.

**Results:** With the multitask learning setting in the model learning process, we achieved an area under the receiver operating characteristic of 0.752 for CA-mortality, 0.711 for ALL-mortality, 0.852 for CA-readmission, and 0.889 for ALL-readmission. The area under the receiver operating characteristic was improved to 0.808 for CA-mortality and 0.862 for CA-readmission after solving the extremely imbalanced issue for CA-mortality/CA-readmission by upsampling and weighting.

**Conclusions:** This study demonstrated the potential of predicting future outcomes for IHCA survivors by machine learning. The results showed that our proposed approach could effectively alleviate data imbalance problems and train a better model for outcome prediction.

***(J Med Internet Res 2021;23(9):e27798)*** doi: 10.2196/27798

**KEYWORDS**

in-hospital cardiac arrest; 30-day mortality; 30-day readmission; machine learning; imbalanced dataset

## Introduction

### Background

In the United States, approximately 209,000 patients experience in-hospital cardiac arrest (IHCA) each year [1]. The rate of survival to hospital discharge is around 14%, and only 7% of IHCA patients could regain an independent life or a partially independent life [2]. In order to reduce the severe effect of IHCA on personal life or society, identifying measures to improve IHCA outcomes is crucial.

### Prior Work

Prognostic factors and prediction tools for survivors of IHCA and their neurologic outcomes have been identified in previous studies [3-8]. However, the evidence of an early warning system for predicting the mortality of IHCA survivors is limited. Current early warning scoring systems using physiologic track-and-trigger systems (TTSs) have been developed for identifying patients at risk for IHCA or other serious outcomes including mortality [9-11]. Most of TTSs rely on the routine observations of vital signs carried out by ward staff. Although many patients could be monitored with this approach, the quality of evidence underpinning the use of TTSs is poor. Specifically, most TTSs have low sensitivity, low positive predictive values, and high specificity [12,13]. In addition to the high mortality after IHCA, readmission after IHCA has a significant cost burden and is associated with comorbidities. Predicting readmission events provides the chance for appropriate interventions and reducing health care costs, including further readmission [14,15].

### Our Study

Here, we first extracted the IHCA cohort from the National Health Insurance Research Database (NHIRD). We assessed their risk based on historical electronic health records (EHRs) in the NHIRD. To provide a long enough window for clinical intervention, we used the 30-day mortality and readmission after IHCA as our prediction targets. In contrast to TTSs, EHRs are prepared by physicians, and they contain several important medical information, including the diagnosis and management of patients. To achieve a better performance, we developed a novel deep learning model, hierarchical vectorizer (HVec), to analyze the patients' historical EHRs and predict mortality and readmission. This study aimed to demonstrate that with the proper data set and learning strategies, we can predict the outcome of IHCA patients based on their historical claims and help clinicians design more effective intervention programs.

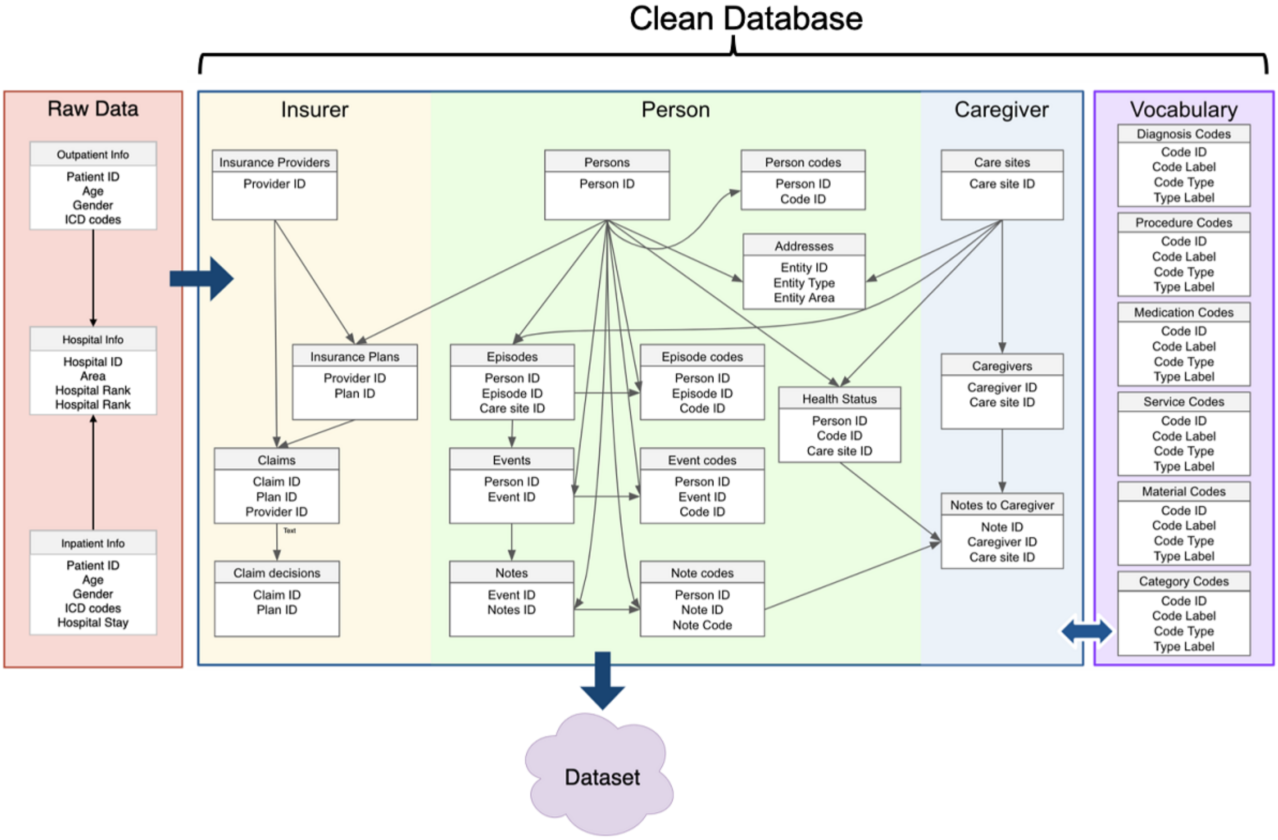
## Methods

### Data Collection

This study was approved by the Institutional Review Board of National Taiwan University Medical College. The IHCA cohort extracted from the NHIRD consisted of 168,693 patients who had at least one IHCA event over 9 years (between January 1, 2002, and December 31, 2010). The Taiwan National Health Insurance program is the only health insurance scheme in Taiwan and covers up to 99.99% of Taiwan's population [16]. The NHIRD contains all health records in inpatient and outpatient settings (clinic or emergency department); however, the records cannot be specifically linked to each patient. International Classification of Disease, 9th Revision (ICD-9) was used during the study period for diagnosis and medical procedures. The NHIRD contains medical information, including gender, age, diagnosis, medical procedure, operation, medication, laboratory test, care site, discharge status, and cost of each hospital visit. Laboratory test results and bedside information, including vital signs, blood pressure, and physical examination, are not included in the NHIRD.

The IHCA population was defined by inpatient records with the ICD-9 procedure codes 99.60 (cardiopulmonary resuscitation, not otherwise specified) and 99.63 (closed-chest cardiac massage) [17]. We used the extract, transform, and load (ETL, see Figure 1) procedure to process raw data into a clean database by eliminating records with missing or invalid information. Raw data in the cleaned database were re-grouped into three major categories (insurer, person, and caregiver) to improve data organization. In addition, vocabulary tables were constructed based on extracted concepts that were used in the raw data.

**Figure 1.** ETL process for converting raw NHIRD data into the data set. The raw data are extracted, transformed, and loaded into the cleaned database after cohort selection and eliminating the invalid data. ETL: extract, transform, load; ICD: International Classification of Disease; NHIRD: National Health Insurance Research Database.



## Experimental Data Set

The 168,693 patients in the data set were split into three data sets: train, validation, and test. The training data set (70% of the data set) was used to train each model. The remaining 30% of patients were split between the validation and test data set evenly to tune the hyperparameters (Table 1) and evaluate model performance, respectively.

For comparison, we trained two single-task models for both mortality and readmission. Person upsampling and event weighting were only performed on the training data set so as not to affect the distribution of the validation and test data set. The F1 score and area under the receiver operating characteristic (AUROC) were our main evaluation metrics.

**Table 1.** Hyperparameter settings.

| Hyperparameter | Value |
|---|---|
| Visit embed size | 200 |
| RNN[a] output size | 128 |
| Learning rate | 1e-3 |
| Dropout rate | 0.5 |
| l2_weights | 0.01 |
| Code embed size | 128 |

[a] RNN: recurrent neural network.

## Input Features

In the retrospective review of the data set, each person may have multiple clinical records (inpatient/outpatient visit) to the hospital within a 9-year period. Each clinical record was set as a unit of analysis. For each clinical record, information was extracted and grouped into input and target features (see Tables 2 and 3). The input features consisted of five major groups described as follows:

1. Medical records consisted of five types of codes, and all codes in the clinical records were mapped to over 400,000 clinical concepts. The health status of the patient in the current record can be determined from this group.

2. Demographics included the age, gender, and information of the targeted patients.
3. Care site information included information of the clinical institution where the patient received treatment.
4. Record statistics provided information on the length (by day) of the record, the number of codes in each International Classification of Disease (ICD) code category, and the total monetary cost involved. This could help the model estimate the severity of the patient’s disease at the record level.
5. Historical information described previous hospital stays and admissions. It was used to estimate the overall health status of the patient in our model.

**Table 2.** Summary of the input features of the model.

| Group and feature name | Feature description | Dimension |
|---|---|---|
| **Medical records** | | |
| Diagnosis | Diagnosis codes from the health record | 128 |
| Procedures | Procedure codes from the health record | 128 |
| Meds | Medication codes from the health record | 128 |
| Tests | Lab test codes from the health record | 128 |
| Other | Other codes from the health record | 128 |
| **Demographics** | | |
| Claim type | Inpatient or outpatient | 2 |
| Age | Age at the event | 1 |
| Gender | Male or female | 2 |
| **Care site information** | | |
| Care site type | Type of site (public, corporate, or private) | 21 |
| Care site specialization | Medical center, community hospital, district hospital, regional hospital, or clinic | 5 |
| Care site rank | Rank of the care site | 17 |
| **Record statistics** | | |
| Hospital stay | Duration of current hospital stay | 1 |
| Total cost | Monetary cost of each of the five ICD[a] codes | 5 |
| Total count | Counts of each of the five ICD codes | 5 |
| **Historical information** | | |
| Past hospitalization duration | The number of days a person spends in the hospital within 3, 6, 12, and 24 months | 4 |
| Past admission count | The number of times a person is admitted to the hospital within 3, 6, 12, and 24 months | 4 |

[a] ICD: International Classification of Disease.

**Table 3.** Summary of the prediction targets.

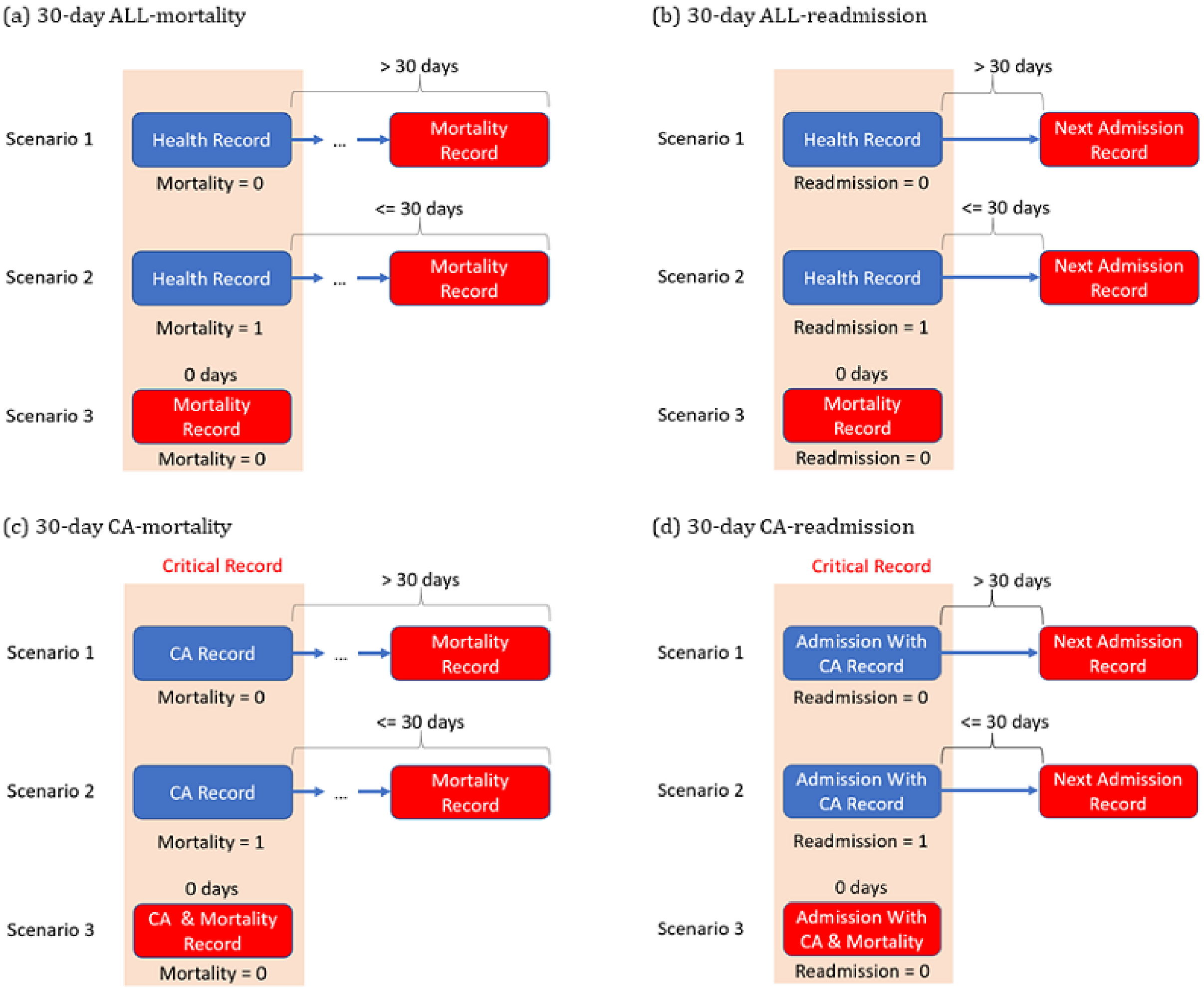
| Group and feature name | Feature description | Dimension |
|---|---|---|
| **Main target** | | |
| Mortality | Whether this event would lead to another mortality event in (within 1 to 30) days | 1 |
| Readmission | Whether this event would lead to another readmission event in (within 1 to 30) days | 1 |

## Targets

A threshold of 30 days was set to predict whether a person would die or readmit within 30 days. Mortality was defined when the patient had an inpatient or outpatient record of mortality or was discharged under critical condition following IHCA. For readmission, whether the patient is readmitted within 1 to 30 days from hospital discharge was predicted. In contrast to other mortality studies, records with mortality (0-day mortality record) were excluded in our study. The main purpose of this strategy was to reduce the “leakage” of features in these records. Our initial results showed that the features of mortality records usually contain information (eg, respiratory failure) explicitly indicating patient mortality. Indeed, these features are significant factors for predicting mortality. However, such cases are not beneficial to our model as the severe condition of these patients makes it hard to treat them with any intervention. Moreover, the high degree of correlation of these features and mortality would cause the model to rely on them and underestimate other potential predictive factors. In order to avoid leakage and let the model focus on other predictive factors, mortality records were set as the negative class, and previous records of mortality (within 30 days) were the positive class.

In clinical practice, the 30-day outcomes of patients after IHCA and discharge from hospitalization are of great interest. The 30-day mortality or readmission after cardiac arrest (CA; CA-mortality/CA-readmission) is a subset of 30-day mortality or readmission. In the rest of the paper, ALL-mortality/ALL-readmission will be used to represent the 30-day mortality or readmission for all records. CA-mortality/CA-readmission refers to the 30-day mortality or readmission following CA (Figure 2).

**Figure 2.** ALL/CA-mortality and ALL/CA-readmission. 3 scenarios of 30-days mortality & readmission after cardiac arrest. Events in red are the outcomes we want to predict. ALL-mortality: 30-day mortality after all records; ALL-readmission: 30-day readmission after all records; CA: cardiac arrest; CA-mortality: 30-day mortality after cardiac arrest records; CA-readmission: 30-day readmission after cardiac arrest records.

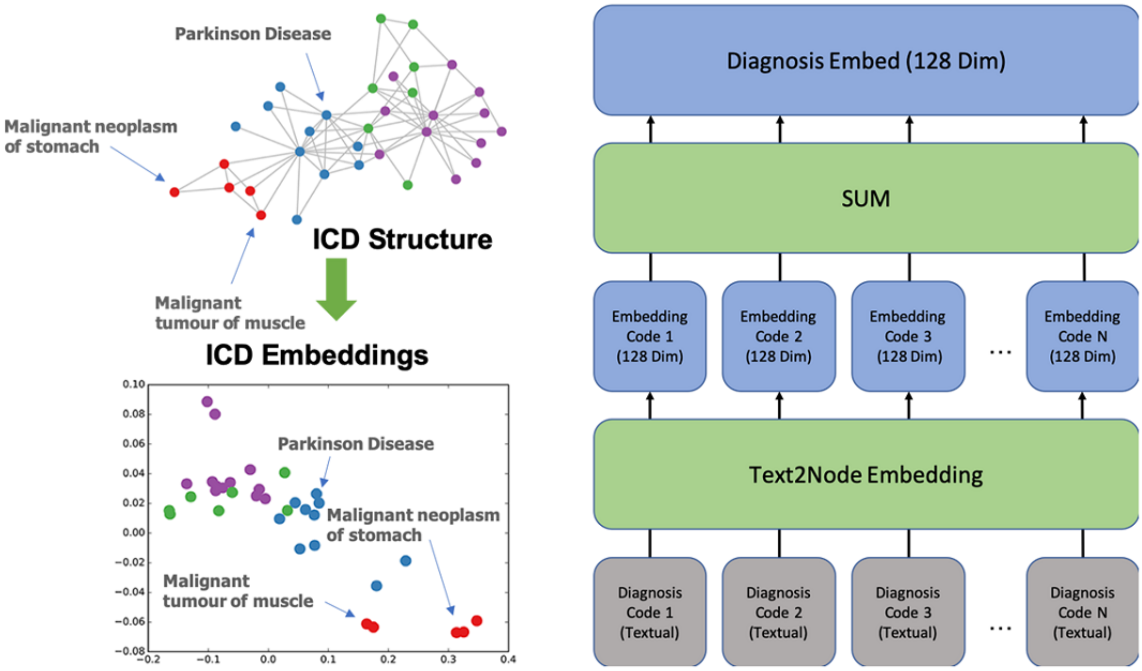


ALL-mortality/ALL-readmission is more common than CA-mortality/CA-readmission as most people would not survive after the first CA event. This means that when we want to predict the future outcome of a recovered CA patient, we do not have enough positive cases for analysis.

## Hierarchical Vectorizer (HVec)

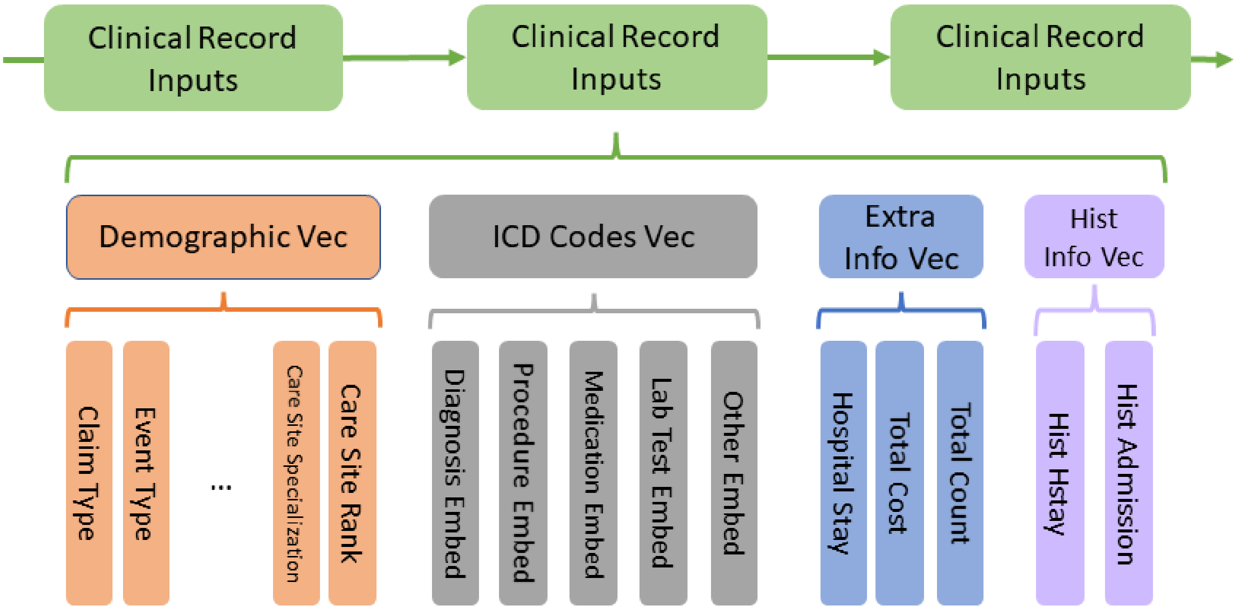
Each record was constructed into a 707-dimension vector for further training. Based on Table 1, all features except for ICD code features (textual features) can be vectorized with one-hot encoding. The features of ICD codes were extracted directly from the health record. A medical knowledge embedding system called Text2Node was used to embed the textual features into vectors [18]. Each of the five categories of ICD codes could contain many ICD codes from a single record, and all codes were added together as a single code for a given category (see Figure 3 as an example). Trained from a substantial medical knowledge database, Text2Node can effectively transform the textual medical concepts into a latent space while preserving the relationship of similar concepts.

**Figure 3.** Example of Text2Node embedding [18] and code embedding for diagnosis codes in a clinical record. Dim: dimension; ICD: International Classification of Disease.



For each clinical record, by concatenating all feature vectors into group vectors hierarchically (see Figure 4), the clinical record vector was obtained. After sorting each clinical record vector according to the date, time series techniques were used to train a model to predict the outcomes of each record.

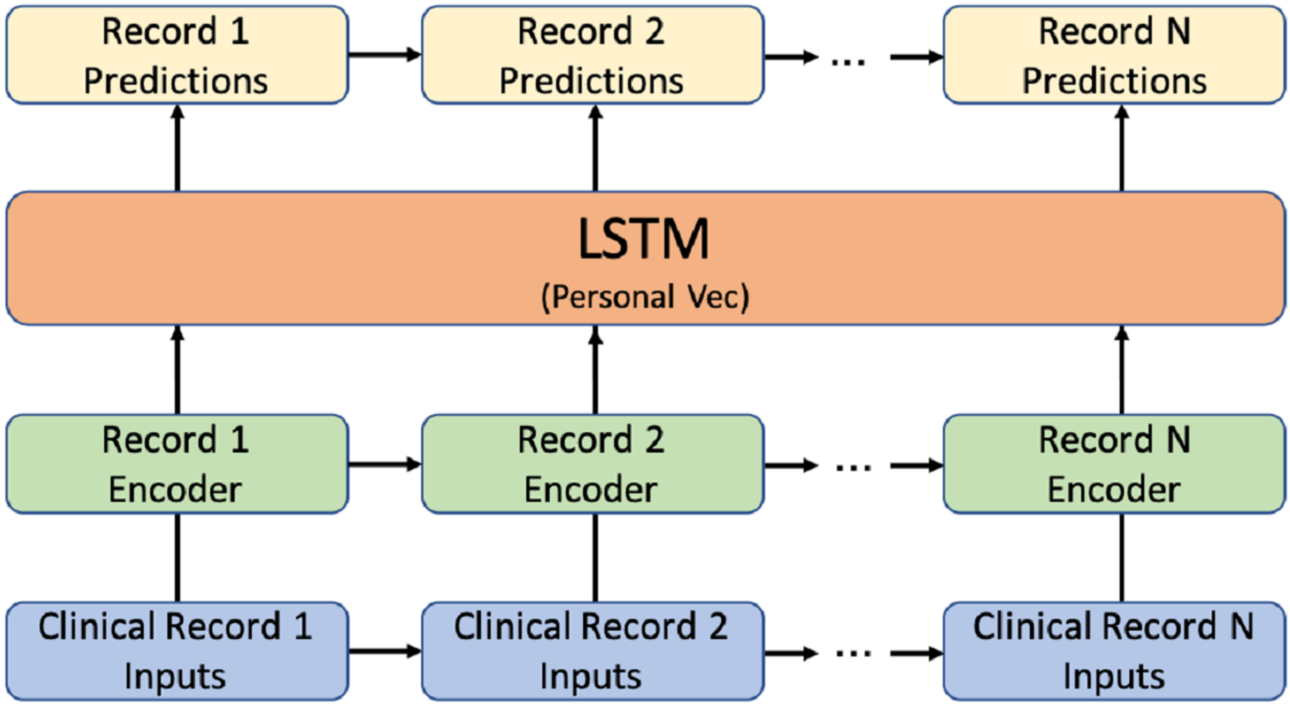
**Figure 4.** Feature concatenation to generate clinical record vectors for time series analysis. ICD: International Classification of Disease; Vec: vector.



### Model Architecture

Deep recurrent neural networks (RNNs) have been proven to be a powerful tool for predicting time series data. In clinical research, the use of RNNs, especially long short-term memory (LSTM), for clinical prediction has been widely investigated [19-21]. Inspired by Choi's work [19], we proposed an HVec model using LSTM networks (Figure 5). In this framework, the record encoder was a fully-connected layer that generated the record embedding for each clinical record independently. The record embedding was a latent vector that contained all the information representing the current clinical record. This latent vector was used as the input of the LSTM to update the person vector (ie, patient status). This person vector was then used to predict our targets.

**Figure 5.** HVec model using LSTM networks. HVec: hierarchical vectorizer; LSTM: long short-term memory; Vec: vector.



However, most previous studies were based on a relatively balanced data set. As we pointed out earlier, the distribution of CA-mortality/CA-readmission is extremely imbalanced. Training LSTM networks with an extremely imbalanced data set is always challenging because, without a carefully designed training strategy, the model could be biased (ie, predicting the negative class for all records). In order to address this problem, two different training strategies have been proposed to alleviate model bias during the training step:

1. The multitask learning framework was introduced to combine several related learning tasks to regularize the gradient and alleviate data imbalance problems during training.
2. At the person level, CA-mortality/CA-readmission records were upsampled, and at the record level, higher weights were assigned to CA records.

## Multitask Learning

The multitask learning framework was proposed, and several auxiliary related tasks were added to the HVec outputs (Figure 6). In this framework, the main functions were mortality and readmission. Here, instead of dividing each main task output into ALL-mortality/ALL-readmission and CA-mortality/CA-readmission independently, an output was considered to cover both because the latter is a subset of the former. Although the distribution was different, they still can be achieved simultaneously with the proposed person upsampling and CA record weighting (described in the next section). Inspired by a previous study [22], three auxiliary autoencoder tasks were introduced to help the model learn the embeddings (Figure 6). Two self-supervised regression tasks were also introduced to allow the embedding to “memorize” the current cost and predict the future cost. Furthermore, another classification task was implemented to predict whether a record is an IHCA record considering that we observed the correlation of IHCA to mortality in the previous analysis.

To monitor the gradients of different tasks and regularize the learning process with auxiliary tasks, Theorem 1 was adopted from Du et al [23].

Theorem 1 given any gradient vector field $G(\theta^{(t)}) = \nabla_\theta \ (\theta)$ (1)

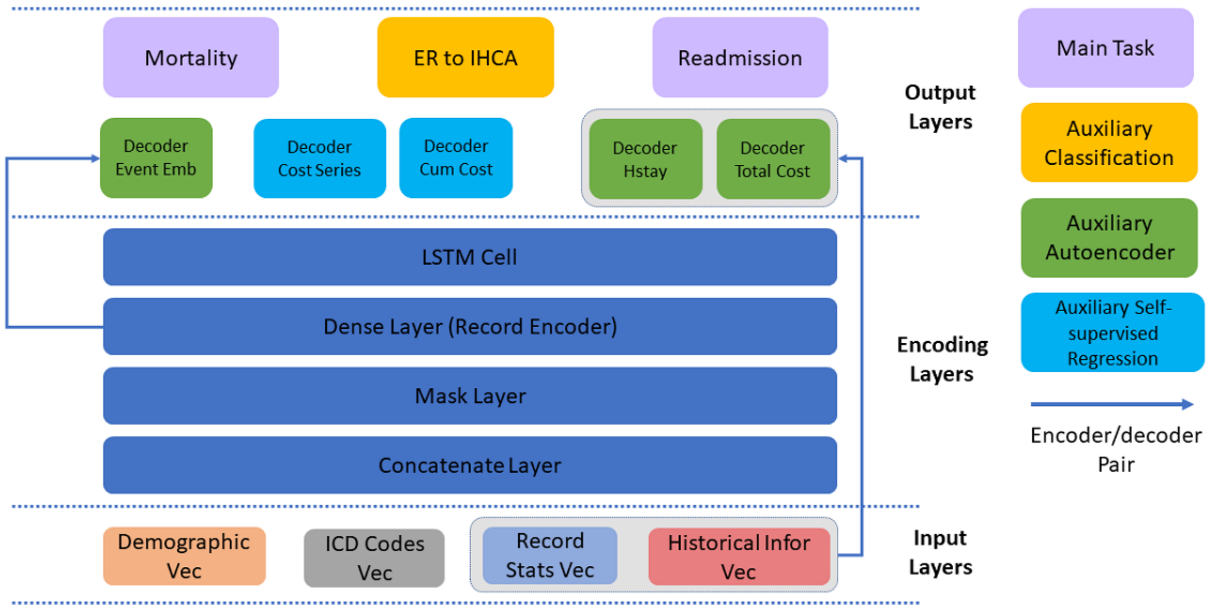
to denote the main task and an arbitrary vector field $V(\theta^{(t)})$ to denote the gradient from another auxiliary task, the update strategy using:

$$\theta^{(t+1)} := \theta^{(t)} - \alpha^{(t)} \left(G(\theta^{(t)}) + V(\theta^{(t)}) + \max\left(0, \cos(G(\theta^{(t)}), V(\theta^{(t)}))\right)\right) \quad \textbf{(2)}$$

with a proper can coverage to a local minimum.

Following this theorem, HVec can learn and converge with a large data set.

**Figure 6.** Multiple outputs for multitask learning. Cum: cumulative; Emb: embedding; ER: emergency room; Hstay: hospital staying; ICD: International Classification of Disease; IHCA: in-hospital cardiac arrest; Infor: information; LSTM: long short-term memory; Stats: statistics; Vec: vector.



## Person Upsampling and CA Record Weighting

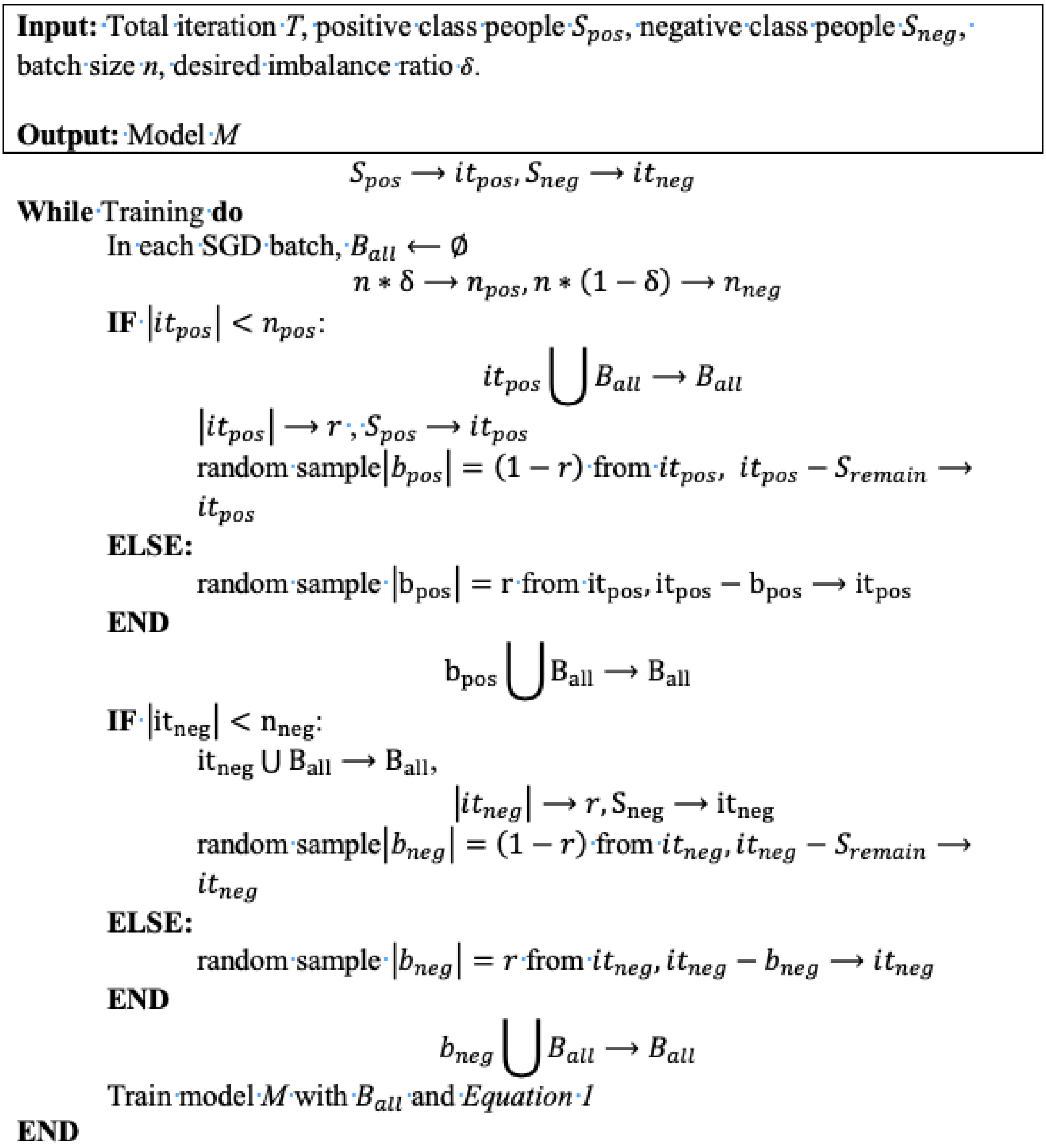
Although the gradients from different tasks can be monitored in multitask learning, with a heavily imbalanced data set (eg, CA-mortality), the auxiliary tasks may fail to regularize the main task. When we trained the HVec model, all records of a person were treated as a single sequential record and fed together into the model. Therefore, in each batch, the batch size was equal to the number of people in the batch. Compared with the number of all records, the number of CA records for each person was relatively rare. CA-mortality and CA-readmission were rare compared with ALL-mortality and ALL-readmission.

The weighting strategy [24] was proposed to solve this problem from two perspectives: at the person level, patients with CA-mortality/CA-readmission records were upsampled per batch (see Figure 7); at the record level, a higher weight was assigned to CA records to make the objective function more sensitive to CA-mortality/CA-readmission records. The upsampling of patients with CA-mortality/CA-readmission records can guarantee that at the person level, there are more CA-mortality/CA-readmission records [25].

Assigning a higher weight to CA records could also emphasize the CA records during training by modifying the loss functions accordingly. Considering *N*-loss functions $\{L_1,\ldots,L_N\}$ corresponding to auxiliary tasks, the loss function can be written as $L(x,y)=\sum_{i}^{N}\mathbb{I}(y_i=1)*w_{pos}*L_i(x,y)+\mathbb{I}(y_i=0)*w_{neg}*L_i(x,y)$ (3).

Where $\mathbb{I}(\cdot)$ (4) is a function that equals 1 if the statement in the bracket is true and otherwise 0; $w_{pos}$ and $w_{neg}$ are the positive and negative class weight, respectively. Combined with the gradient update strategy in equation 1, the HVec can learn from the extremely imbalanced data set effectively.

**Figure 7.** Person upsampling.



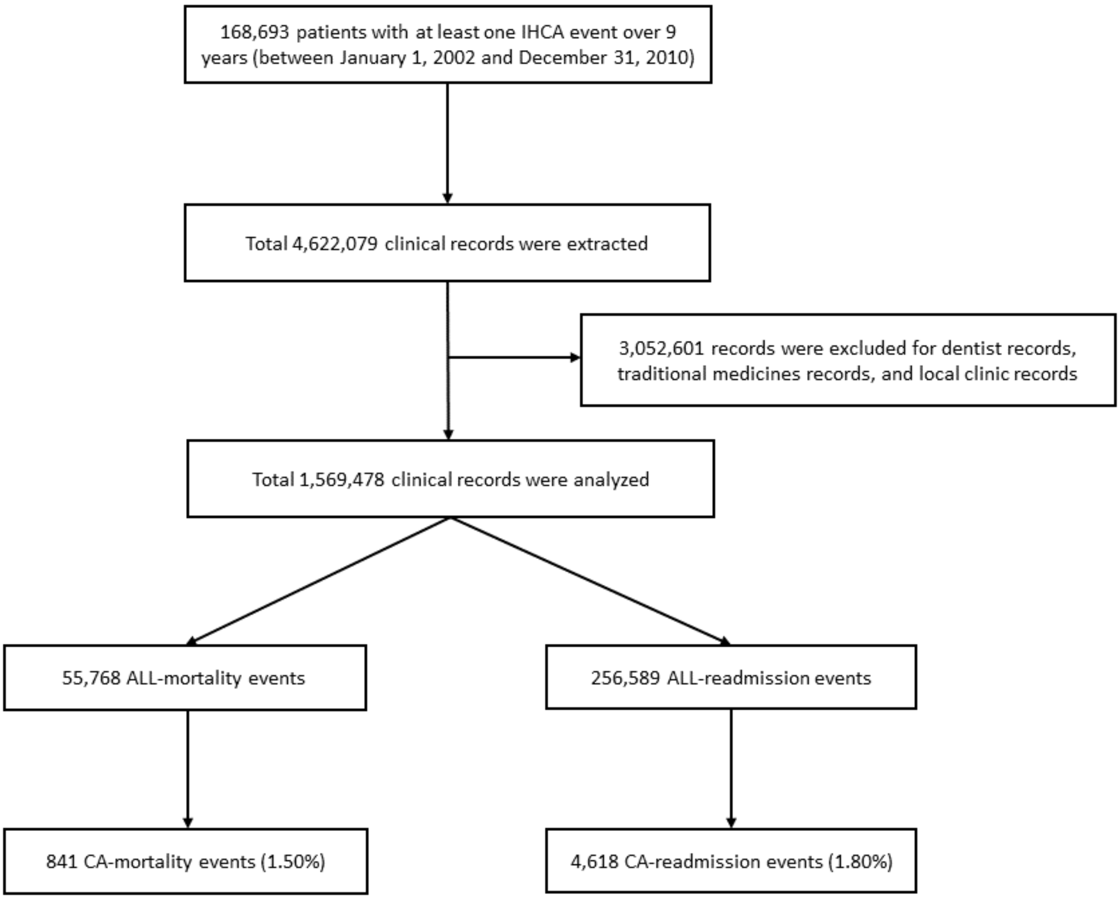
```
Input: Total iteration $T$, positive class people $S_{pos}$, negative class people $S_{neg}$,
batch size $n$, desired imbalance ratio $\delta$.

Output: Model $M$
                    $S_{pos} \longrightarrow it_{pos}, S_{neg} \longrightarrow it_{neg}$
While Training do
    In each SGD batch, $B_{all} \longleftarrow \emptyset$
                    $n * \delta \longrightarrow n_{pos}, n * (1 - \delta) \longrightarrow n_{neg}$
    IF $|it_{pos}| < n_{pos}$:
                    $it_{pos} \bigcup B_{all} \longrightarrow B_{all}$
        $|it_{pos}| \longrightarrow r$ ; $S_{pos} \longrightarrow it_{pos}$
        random sample $|b_{pos}| = (1 - r)$ from $it_{pos}$, $it_{pos} - S_{remain} \longrightarrow$
        $it_{pos}$
    ELSE:
        random sample $|b_{pos}| = r$ from $it_{pos}, it_{pos} - b_{pos} \longrightarrow it_{pos}$
    END
                    $b_{pos} \bigcup B_{all} \longrightarrow B_{all}$
    IF $|it_{neg}| < n_{neg}$:
        $it_{neg} \cup B_{all} \longrightarrow B_{all}$,
                    $|it_{neg}| \longrightarrow r, S_{neg} \longrightarrow it_{neg}$
        random sample $|b_{neg}| = (1 - r)$ from $it_{neg}, it_{neg} - S_{remain} \longrightarrow$
        $it_{neg}$
    ELSE:
        random sample $|b_{neg}| = r$ from $it_{neg}, it_{neg} - b_{neg} \longrightarrow it_{neg}$
    END
                    $b_{neg} \bigcup B_{all} \longrightarrow B_{all}$
    Train model $M$ with $B_{all}$ and Equation 1
END
```

## Results

### Overview

A total of 168,693 patients and 4,622,079 clinical records were extracted from the NHIRD over 9 years, and 3,052,601 records (dentist records, traditional medicine records, or local clinic records) were excluded because these records were concentrated with repetitive conditions and mainly added noise to the machine learning models (Figure 8).

**Figure 8.** CONSORT diagram of the study cohort. ALL-mortality: 30-day mortality after all records; ALL-readmission: 30-day readmission after all records; CA: cardiac arrest; CA-mortality: 30-day mortality after CA records; CA-readmission: 30-day readmission after CA records; IHCA: in-hospital cardiac arrest.



There were 1,569,478 clinical records in the cleaned database, including both inpatient and outpatient records, from 168,693 patients (mean number of records per person: 9.30, SD 10.90) who have experienced at least one IHCA event. The results indicate an imbalanced data set, where the most imbalanced task was CA-mortality with a ratio of 0.53%. The characteristics of the study population are summarized in Table 4. There were 173,345 IHCA records (11.04% of the total clinical records), and on average, there were 1.02 IHCA records for each person.

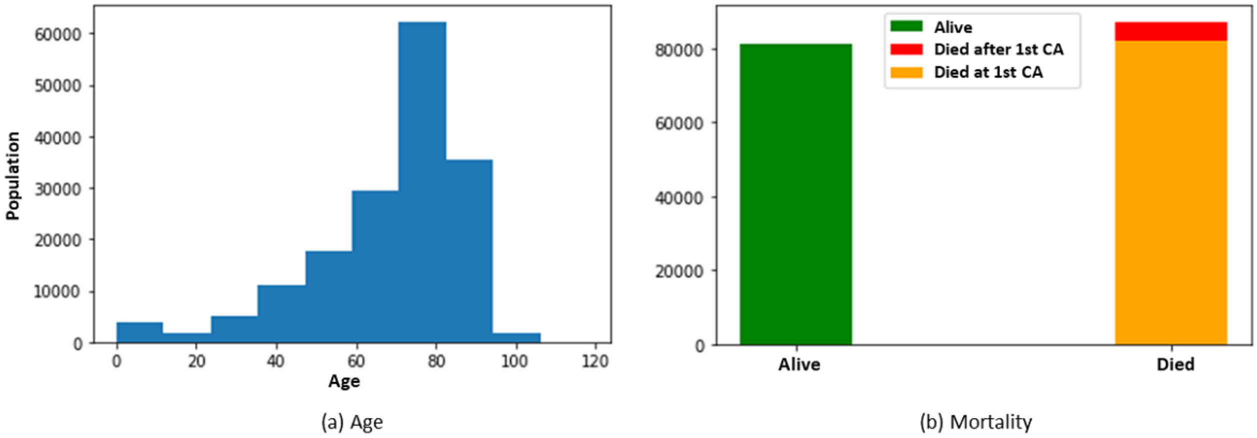
The age of the patients in the data set ranged from 0 (newborn) to 118 years (mean age 68.66, SD 18.96 years), including 104,691 females and 64,002 males. Overall, 164,322 patients (97.4%) had CA only once, 4,174 patients (2.4%) had CA twice, and only 197 patients (0.2%) had CA more than twice. Death was recorded for 87,311 patients (51.75% mortality rate). Of these 87,311 patients, 82,225 patients died during their first hospitalization for CA (94.17%; Figure 9).

**Table 4.** Characteristics of the study population.[a]

| Characteristics | Study population (N=168,693) |
|---|---|
| Age (years), mean (SD) | 68.66 (18.96) |
| Gender (male), n (%) | 64,002 (37.9) |
| Record number per person, mean (SD) | 9.30 (10.90) |
| **Cardiac arrest frequency, n (%)** | |
| 1 | 164,322 (97.4) |
| 2 | 4174 (2.4) |
| ≥3 | 197 (0.2) |
| Mortality, n (%) | 87,311 (51.75) |

[a] Continuous variables are presented as the mean (SD), and categorical variables are presented as the number (percentage of the study population).

**Figure 9.** Age and mortality statistics of the data set. CA: cardiac arrest.



(a) Age

(b) Mortality

## Experiment 1: Single-Task Learning Versus Multitask Learning

In this experiment, person upsampling and event weighting were not applied. The model performance is summarized in Tables 5 and 6.

As shown in Table 5, multitask learning could improve the model performance for ALL-mortality and CA-mortality in terms of the AUROC and F1 scores. Based on single-task and multitask results, there was a relatively high improvement in performance for the extremely imbalanced CA-mortality task compared with the ALL-mortality task. However, the precision was relatively low due to the imbalance ratio, which also affected the F1 score. Moreover, the F1 score of CA-mortality was too low for real-life applications.

**Table 5.** Single-task and multitask learning performance for CA-mortality and ALL-mortality.[a]

| Mortality | CA-mortality | | | | ALL-mortality | | | |
|---|---|---|---|---|---|---|---|---|
| | AUROC[b] | F1 | Precision | Recall | AUROC | F1 | Precision | Recall |
| Single-task learning | 0.658 | 0.014 | 0.010 | 0.024 | 0.663 | 0.130 | 0.101 | 0.180 |
| Multitask learning | 0.752 | 0.049 | 0.041 | 0.060 | 0.711 | 0.147 | 0.093 | 0.349 |

[a] CA-mortality: cardiac arrest mortality (30-day mortality after CA records); ALL-mortality: 30-day mortality after all records.

[b] AUROC: area under the receiver operating characteristic.

As shown in Table 6, the improvement in multitask learning for readmission prediction was not as significant as that for mortality prediction. Furthermore, in CA-readmission prediction, the F1 score and precision were decreased.

**Table 6.** Single-task and multitask learning performance for CA-readmission and ALL-readmission.[a]

| Readmission | CA-readmission | | | | ALL-readmission | | | |
|---|---|---|---|---|---|---|---|---|
| | AUROC[b] | F1 | Precision | Recall | AUROC | F1 | Precision | Recall |
| Single-task learning | 0.847 | 0.214 | 0.162 | 0.315 | 0.872 | 0.554 | 0.424 | 0.801 |
| Multitask learning | 0.852 | 0.209 | 0.152 | 0.335 | 0.889 | 0.562 | 0.430 | 0.811 |

[a] CA-readmission: cardiac arrest readmission (30-day readmission after CA records); ALL-readmission: 30-day readmission after all records.

[b] AUROC: area under the receiver operating characteristic.

In this experiment, compared with single-task learning, multitask learning could achieve a better performance for ALL-mortality and ALL-readmission. However, multitask learning could not solve the extremely imbalanced data set.

## Experiment 2: Improving CA Prediction Performance

Experiment 1 showed that models had difficulties making good predictions based on the CA-mortality and CA-readmission data due to the extremely imbalanced data set. In this experiment, we demonstrated that by applying person upsampling and event weighting, we could further improve the performance for CA-mortality and CA-readmission without considerably affecting ALL-mortality and ALL-readmission. In our experiment, the upsampling rate indicates how many times upsampling was performed for the positive class (patients with CA-mortality or CA-readmission) in a batch, and an upsampling rate of 1 means we did not perform upsampling. The CA event weight indicates the loss weight $w_{pos}$ in equation

3, and we always set $w_{neg}$=1. We used different upsampling rates and event weights in our experiments. Our results are summarized in Tables 7 and 8, including the previous results on single-task model performance (first row of each table).

Overall, compared with the 30-day mortality task, the imbalanced 30-day mortality task showed a larger increase in performance when applying balancing techniques. As shown in Table 7, after upsampling and event weighting, the models demonstrated improved performance for both ALL-mortality and CA-mortality. For CA-mortality, the F1 score was increased by 36.7% (from 0.049 to 0.067).

**Table 7.** Mortality models with their respective hyperparameter configuration and their performance in predicting both CA-mortality and ALL-mortality.[a]

| Mortality | | | CA-mortality | | ALL-mortality | |
|---|---|---|---|---|---|---|
| Upsampling rate | Event weight | Multitask | AUROC[b] | F1 | AUROC | F1 |
| 1 | 1 | No | 0.658 | 0.014 | 0.663 | 0.130 |
| 1 | 1 | Yes | 0.752 | 0.049 | 0.711 | 0.147 |
| 10 | 1 | Yes | 0.808 | 0.064 | 0.728 | 0.155 |
| 10 | 5 | Yes | 0.802 | 0.067 | 0.726 | 0.158 |

[a] CA-mortality: cardiac arrest mortality (30-day mortality after CA records); ALL-mortality: 30-day mortality after all records.

[b] AUROC: area under the receiver operating characteristic.

Similar to ALL-mortality and CA-mortality, the AUROC and F1 score were increased for both ALL-readmission and CA-readmission by applying the two techniques. Based on the results in Table 8, a minor but consistent increase was achieved in CA-readmission prediction. However, there was no significant improvement in ALL-readmission prediction. This is because upsampling and event weighting would not greatly affect the learning process for a more balanced data set.

**Table 8.** Readmission models with their respective hyperparameter configuration and their performance in predicting CA-readmission and ALL-readmission.[a]

| Readmission | | | CA-readmission | | ALL-readmission | |
|---|---|---|---|---|---|---|
| Upsampling rate | Event weight | Multitask | AUROC[b] | F1 | AUROC | F1 |
| 1 | 1 | No | 0.847 | 0.214 | 0.872 | 0.554 |
| 1 | 1 | Yes | 0.852 | 0.209 | 0.889 | 0.562 |
| 5 | 1 | Yes | 0.861 | 0.230 | 0.884 | 0.555 |
| 5 | 5 | Yes | 0.862 | 0.237 | 0.884 | 0.555 |

[a] CA-readmission: cardiac arrest readmission (30-day readmission after CA records); ALL-readmission: 30-day readmission after all records.

[b] AUROC: area under the receiver operating characteristic.

In summary, the results indicated that multitask learning, upsampling, and event weighting could effectively improve the model prediction performance for an imbalanced data set. We also showed that these techniques could be collectively used to achieve better results for an extremely imbalanced data set.

## Discussion

In this study, we constructed a large patient database that includes 9 years of EHRs for over 168,000 IHCA patients, which can be used for future IHCA-related research. In addition, we developed an HVec model (LSTM model) that uses a multitask learning strategy to predict the 30-day mortality and readmission.

The results showed that our model could successfully predict future mortality and readmission using EHR data for IHCA patients. We proposed the person upsampling and record weighting strategies to handle the extremely imbalanced data problem in this study. After applying these techniques, some improvements were achieved in CA-mortality and CA-admission prediction.

In contrast to other studies using deep learning models to predict another IHCA event after CA [26], our study focused on predicting the future outcomes of IHCA patients after discharge. To the best of our knowledge, this is the first study to predict mortality and readmission after IHCA events by machine learning. The model may serve as a surveillance system for those who experienced IHCA. Patients with a high risk of mortality or readmission in the near future could be identified and re-evaluated before discharge. This study also demonstrated the potential of another model for predicting future mortality and readmission after each record using previous EHRs (ALL-mortality/ALL-readmission). The model might help identify those with a high risk in inpatient and outpatient situations. However, mortality and readmission rates are different in the general population compared with the patients selected in this study. Using hospital EHRs with patients' information in the NHIRD, we can construct a real-time alert system based on machine learning methods to predict the adverse events of IHCA survivors and improve their outcomes. Further prospective studies are needed to verify the utility of this system in the general population.

Several studies have reported models for predicting the outcomes of CA patients [8]. In a systematic review of current prediction models, the median AUROC value was 0.84 with an IQR of 0.80 to 0.89 [8]. For IHCA survivors, Chan et al. reported that the cardiac arrest survival post-resuscitation in-hospital (CASPRI) score could be used to predict favorable neurologic outcomes after discharge [3]. The AUROC of the CASPRI score was 0.80. Nanayakkara et al used deep learning models to predict the IHCA events of CA survivors, and the AUROC was 0.87 [26]. In our study, we encountered the difficulty of imbalanced data. Using proper learning strategies, we achieved comparable AUROC values (0.808 for CA-mortality and 0.862 for CA-readmission). However, the recall and precision rates were low in our study due to imbalanced data. This is a challenge we aim to resolve. When using historical medical records to predict outcomes, in many cases, the records contain information that may indicate the outcomes. For example, critical diagnosis and rescue medication are often associated with mortality. Including these types of information can facilitate training and give a high AUROC and F1; however, the model itself is of limited use. Mortality records were set as the negative class in our study to avoid overfitting. In further studies, the model may be improved by adjusting the threshold to optimize the trade-off between specificity and sensitivity.

In our HVec framework, we encoded each person's EHRs in two levels of latent vectors (record-level and person-level) and ensured that the model learns both simultaneously using the unsupervised autoencoder strategy. The predictive results were promising with these latent vectors. The latent vectors in these two levels may be further explored to facilitate clinicians' decision-making and provide better clinical interventions. In addition, the person vector may be used as a biomarker to evaluate the overall health status of a person beyond the health care setting. Along with some recently developed models such as Deep Patient [22] and MixEHR [27], we showed that the use of deep neural networks to extract information from EHRs might solve complex clinical research problems.

An imbalanced data distribution is common in clinical research, especially for disease-related predictions. In comparison with common diseases, many important diseases lack positive cases, making it difficult to train a good model. In a previous study [19], the authors attempted to address this problem by using a balanced, distributed data set and train a deep learning model with the balanced data set. Similarly, in this study involving CA-mortality and CA-readmission, we demonstrated that by carefully designing model learning strategies (eg, multitask learning and upsampling), some common problems in clinical research could be solved effectively with machine learning models.

This study has some limitations. First, the IHCA cohort was retrospectively extracted from the NHIRD. Further studies are needed to evaluate the efficacy of this model as an early warning system and determine how this system affects patients' outcomes. Second, as our model was developed based on the NHIRD, the generalization of this model to other health insurance data sets is not proven. Third, each patient's vital signs and laboratory data were not included in the analysis due to the study design. A combination of EHRs and patients' clinical data may further improve model performance. Fourth, traditional machine learning methods have the limitation of interpretability. Specific risk factors for 30-day mortality/readmission were unknown in this study. In future studies, we plan to develop an explainable model and investigate specific predictive factors in the model.

In summary, our model showed good performance in predicting 30-day mortality and readmission after IHCA, which can help clinicians monitor CA patients' status better. We aim to provide more insights to clinicians with proactive intervention recommendations. Nevertheless, a challenge remains in the interpretative ability of the deep learning model. Our future work will mainly focus on the interpretative power of the model trained using EHRs.

## Acknowledgments

This study was supported by the Ministry of Science and Technology, Taiwan (project: 109-2634-F-002-031). CYC, YLH, and CHH were involved in data collection. SA, AW, KCF, and JX were involved in constructing the deep learning model. CYC and SA contributed equally to this paper and drafted the article. Finally, YLH, CHH, and RS made the final revision. YLH, CHH, and RS contributed equally as corresponding authors. All authors approved the final version of the article.

## Conflicts of Interest

None declared.

## References

1. Writing GM, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation 2016 Jan 26;133(4):e38-e360. [doi: 10.1161/CIR.0000000000000350] [Medline: 26673558]
2. Wang C, Huang C, Chang W, Tsai M, Yu P, Wu Y, et al. Outcomes of Adult In-Hospital Cardiac Arrest Treated with Targeted Temperature Management: A Retrospective Cohort Study. PLoS One 2016;11(11):e0166148 [FREE Full text] [doi: 10.1371/journal.pone.0166148] [Medline: 27820847]

3. Chan PS, Spertus JA, Krumholz HM, Berg RA, Li Y, Sasson C, Get With the Guidelines-Resuscitation Registry Investigators. A validated prediction tool for initial survivors of in-hospital cardiac arrest. Arch Intern Med 2012 Jun 25;172(12):947-953 [FREE Full text] [doi: 10.1001/archinternmed.2012.2050] [Medline: 22641228]
4. Holmberg MJ, Moskowitz A, Raymond TT, Berg RA, Nadkarni VM, Topjian AA, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Derivation and Internal Validation of a Mortality Prediction Tool for Initial Survivors of Pediatric In-Hospital Cardiac Arrest. Pediatr Crit Care Med 2018 Mar;19(3):186-195 [FREE Full text] [doi: 10.1097/PCC.0000000000001416] [Medline: 29239980]
5. Chan PS, Nallamothu BK, Krumholz HM, Spertus JA, Li Y, Hammill BG, American Heart Association Get with the Guidelines–Resuscitation Investigators. Long-term outcomes in elderly survivors of in-hospital cardiac arrest. N Engl J Med 2013 Mar 14;368(11):1019-1026 [FREE Full text] [doi: 10.1056/NEJMoa1200657] [Medline: 23484828]
6. Chelly J, Mpela A, Jochmans S, Brunet J, Legriel S, Guerin L, et al. OHCA (Out-of-Hospital Cardiac Arrest) and CAHP (Cardiac Arrest Hospital Prognosis) scores to predict outcome after in-hospital cardiac arrest: Insight from a multicentric registry. Resuscitation 2020 Nov;156:167-173. [doi: 10.1016/j.resuscitation.2020.09.021] [Medline: 32976962]
7. Chung C, Chiu W, Huang Y, Chan L, Hong C, Chiu H. Identifying prognostic factors and developing accurate outcome predictions for in-hospital cardiac arrest by using artificial neural networks. J Neurol Sci 2021 Jun 15;425:117445. [doi: 10.1016/j.jns.2021.117445] [Medline: 33878655]
8. Carrick RT, Park JG, McGinnes HL, Lundquist C, Brown KD, Janes WA, et al. Clinical Predictive Models of Sudden Cardiac Arrest: A Survey of the Current Science and Analysis of Model Performances. JAHA 2020 Aug 18;9(16):e017625. [doi: 10.1161/jaha.119.017625]
9. Jansen JO, Cuthbertson BH. Detecting critical illness outside the ICU: the role of track and trigger systems. Curr Opin Crit Care 2010 Jun;16(3):184-190. [doi: 10.1097/MCC.0b013e328338844e] [Medline: 20305556]
10. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. Resuscitation 2008 May;77(2):170-179. [doi: 10.1016/j.resuscitation.2007.12.004] [Medline: 18249483]
11. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
12. Sharpley JT, Holden JC. Introducing an early warning scoring system in a district general hospital. Nurs Crit Care 2004;9(3):98-103. [doi: 10.1111/j.1362-1017.2004.0063.x] [Medline: 15152751]
13. Priestley G, Watson W, Rashidian A, Mozley C, Russell D, Wilson J, et al. Introducing Critical Care Outreach: a ward-randomised trial of phased introduction in a general hospital. Intensive Care Med 2004 Jul;30(7):1398-1404. [doi: 10.1007/s00134-004-2268-7] [Medline: 15112033]
14. Chan PS, Nallamothu BK, Krumholz HM, Curtis LH, Li Y, Hammill BG, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Readmission rates and long-term hospital costs among survivors of an in-hospital cardiac arrest. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):889-895 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.114.000925] [Medline: 25351479]
15. Morita Y, Haruna T, Haruna Y, Nakane E, Yamaji Y, Hayashi H, et al. Incidence, predictors, causes, and costs of 30-day readmission after in-hospital cardiopulmonary resuscitation in the United States. Resuscitation 2019 Jan;134:19-25. [doi: 10.1016/j.resuscitation.2018.12.001] [Medline: 30566891]
16. Hsieh C, Su C, Shao S, Sung S, Lin S, Kao Yang Y, et al. Taiwan's National Health Insurance Research Database: past and future. Clin Epidemiol 2019;11:349-358 [FREE Full text] [doi: 10.2147/CLEP.S196293] [Medline: 31118821]
17. DeZorzi C, Boyle B, Qazi A, Luthra K, Khera R, Chan PS, et al. Administrative Billing Codes for Identifying Patients With Cardiac Arrest. J Am Coll Cardiol 2019 Apr 02;73(12):1598-1600 [FREE Full text] [doi: 10.1016/j.jacc.2019.01.030] [Medline: 30922482]
18. Soltani R, Tomberg A. Text2node: a cross-domain system for mapping arbitrary phrases to a taxonomy. arXiv. 2019 Apr 11. URL: https://arxiv.org/abs/1905.01958v1 [accessed 2019-04-11]
19. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
20. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
21. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018 Dec;22(5):1589-1604. [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
22. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
23. Du Y, Czarnecki W, Jayakumar S, Pascanu R, Lakshminarayanan B. Adapting auxiliary losses using gradientsimilarity. arXiv 2018 Dec 5:1812 [FREE Full text]
24. Haibo He, Garcia E. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng 2009 Sep;21(9):1263-1284. [doi: 10.1109/TKDE.2008.239]

25. Ertekin Ş. Adaptive Oversampling for Imbalanced Data Classification. In: Gelenbe E, Lent R, editors. Information Sciences and Systems 2013. Lecture Notes in Electrical Engineering, vol 264. Cham: Springer; Sep 24, 2013:261-269.
26. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. PLoS Med 2018 Nov;15(11):e1002709 [FREE Full text] [doi: 10.1371/journal.pmed.1002709] [Medline: 30500816]
27. Li Y, Nair P, Lu XH, Wen Z, Wang Y, Dehaghi AAK, et al. Inferring multimodal latent topics from electronic health records. Nat Commun 2020 May 21;11(1):2536 [FREE Full text] [doi: 10.1038/s41467-020-16378-3] [Medline: 32439869]

## Abbreviations

**AUROC:** area under the receiver operating characteristic
**CA:** cardiac arrest
**CASPRI:** cardiac arrest survival post-resuscitation in-hospital
**CPR:** cardiopulmonary resuscitation
**EHR:** electronic health record
**ETL:** extract, transform, and load
**HVec:** hierarchical vectorizer
**ICD:** International Classification of Disease
**IHCA:** in-hospital cardiac arrest
**LSTM:** long short-term memory
**NHIRD:** National Health Insurance Research Database
**OHCA:** out-of-hospital cardiac arrest
**RNN:** recurrent neural network
**TTS:** track-and-trigger system

*Edited by R Kukafka; submitted 08.02.21; peer-reviewed by A Ćirković, W Yang, SF Sung, B Richards; comments to author 15.03.21; revised version received 09.05.21; accepted 21.07.21; published 13.09.21*

*Please cite as:*
*Chi CY, Ao S, Winkler A, Fu KC, Xu J, Ho YL, Huang CH, Soltani R*
*Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach*
*J Med Internet Res 2021;23(9):e27798*
*URL: https://www.jmir.org/2021/9/e27798*
*doi: 10.2196/27798*
*PMID:*

Original Paper

# Privacy Practices of Health Information Technologies: Privacy Policy Risk Assessment Study and Proposed Guidelines

Haley M LaMonica, BSc, MA, PhD, ABPP-CN; Anna E Roberts, BAppSc, MExSC, MBMSc; Grace Yeeun Lee, MPH, MHM; Tracey A Davenport, BA(Hons), eMBA; Ian B Hickie, AM, MD, FRANZCP, FASSA

Brain and Mind Centre, The University of Sydney, Camperdown, Australia

**Corresponding Author:**
Haley M LaMonica, BSc, MA, PhD, ABPP-CN
Brain and Mind Centre
The University of Sydney
88 Mallett Street
Camperdown, 2050
Australia
Phone: 61 0426955658
Email: haley.lamonica@sydney.edu.au

## Abstract

**Background:** Along with the proliferation of health information technologies (HITs), there is a growing need to understand the potential privacy risks associated with using such tools. Although privacy policies are designed to inform consumers, such policies have consistently been found to be confusing and lack transparency.

**Objective:** This study aims to present consumer preferences for accessing privacy information; develop and apply a privacy policy risk assessment tool to assess whether existing HITs meet the recommended privacy policy standards; and propose guidelines to assist health professionals and service providers with understanding the privacy risks associated with HITs, so that they can confidently promote their safe use as a part of care.

**Methods:** In phase 1, participatory design workshops were conducted with young people who were attending a participating *headspace* center, their supportive others, and health professionals and service providers from the centers. The findings were knowledge translated to determine participant preferences for the presentation and availability of privacy information and the functionality required to support its delivery. Phase 2 included the development of the 23-item privacy policy risk assessment tool, which incorporated material from international privacy literature and standards. This tool was then used to assess the privacy policies of 34 apps and e-tools. In phase 3, privacy guidelines, which were derived from learnings from a collaborative consultation process with key stakeholders, were developed to assist health professionals and service providers with understanding the privacy risks associated with incorporating HITs as a part of clinical care.

**Results:** When considering the use of HITs, the participatory design workshop participants indicated that they wanted privacy information to be easily accessible, transparent, and user-friendly to enable them to clearly understand what personal and health information will be collected and how these data will be shared and stored. The privacy policy review revealed consistently poor readability and transparency, which limited the utility of these documents as a source of information. Therefore, to enable informed consent, the privacy guidelines provided ensure that health professionals and consumers are fully aware of the potential for privacy risks in using HITs to support health and well-being.

**Conclusions:** A lack of transparency in privacy policies has the potential to undermine consumers' ability to trust that the necessary measures are in place to secure and protect the privacy of their personal and health information, thus precluding their willingness to engage with HITs. The application of the privacy guidelines will improve the confidence of health professionals and service providers in the privacy of consumer data, thus enabling them to recommend HITs to provide or support care.

***(J Med Internet Res 2021;23(9):e26317)*** doi: 10.2196/26317

**KEYWORDS**

privacy; mental health; technology; digital tools; smartphone; apps

# Introduction

## Health Information Technologies

Digital health has quickly become an integral component of best practice health care, transforming the way care is delivered. By capitalizing on digital infrastructure, it is widely recognized that digital health solutions improve access to care, particularly for individuals with mobility or transport restrictions, or for those who live remotely where health care resources may be limited [1,2]. The availability of health information technologies (HITs) is proving invaluable during the COVID-19 pandemic, where face-to-face mental health care is often delivered digitally (eg, videoconferencing) [3]. Beyond access issues, however, digital health also has the potential to optimize or eliminate waitlists and facilitate routine outcome monitoring to strengthen and maintain patient-health professional relationships [4], allowing for shared, data-driven decision-making on appropriate treatment plans [5]. With a greater need for and reliance on digital health solutions for screening, treatment, and ongoing maintenance of health, there is now an increased focus on the privacy and security of personal and health information collected via HITs, such as health-related apps and e-tools (eg, websites and web-based courses).

## Legal and Ethics Rights of Individuals

It is crucial to consider the legal and ethical rights of individuals who choose to both explicitly and passively share their web-based health information. This is essential, particularly in the area of mental health care, where data often contain highly personal information that could cause significant harm and distress if not handled appropriately. There is increasing documentation and guidance in this area, such as the recent release of the National Safety and Quality Digital Mental Health Standards (consultative draft) by the Australian Commission on Safety and Quality in Health Care [6], which includes educational brochures that provide tips to consumers, caregivers [7], and clinicians [8] on choosing a digital mental health service. Specific to privacy, the Australian Privacy Principles require that all organizations have a clearly expressed and current privacy policy detailing how personal information is managed [9]. As personal and health information is deemed particularly sensitive, extra protection concerning its handling was established under the Privacy Act 1988 [10]. The World Economic Forum also has highlighted *trust* as one of the primary issues that needs to be addressed on a global scale to ensure consumers' and health professionals' trust in the privacy and security standards of new digital tools and technology-based therapies [11].

## Privacy Practices of HITs

The use of health-related apps has rapidly increased in recent years, with 47% of Australian consumers using apps in 2018 [12]. Health professionals are also increasingly recommending HITs as part of clinical practice. For example, approximately half of Australian general practitioners responding to an annual technology survey indicated that they recommend HITs for at least monthly use by patients, with mindfulness and mental health apps recommended most often [13]. Although HITs, including apps and e-tools, have gained considerable favor with consumers and health professionals for promoting self-management of health and well-being, the privacy of personal and health information remains a notable area of concern. A systematic review of 79 health and wellness apps certified as clinically safe and trustworthy by the United Kingdom National Health Service Health Apps Library found that 66% of apps transmitting personal information on the internet did so without encryption, and 20% did not have any form of privacy policy [14]. Furthermore, although no app collected or shared information in a manner that was not explicitly stated in the privacy policy, the nature of the personal information included in such transmission was not described in 78% of policies [14]. A recent cross-sectional assessment of 36 top-ranked apps found that 92% of them transmitted data to a third party; however, only 64% of privacy policies made this explicit. In addition, only 43% and 50% of privacy policies disclosed that apps were transmitting data to Google and Facebook, respectively [15]. This begs the question, can users trust that personal and health information collected via HITs will be kept private and secure?

## Objectives

This study aims to use co-design methodologies to better understand young people's preferences for learning about how their personal and health information will be handled by HITs and create prototypes for the InnoWell Platform. The InnoWell Platform was developed by InnoWell, a joint venture between the University of Sydney and PricewaterhouseCoopers (PwC; Australia) through Project Synergy, an Aus $30 million (US $22.1 million) Australian government–funded initiative [16]. As described previously [5,17], the InnoWell Platform is a co-designed digital tool embedded within traditional in-clinic and web-based mental health services to support person-centered, measurement-based care. This study also seeks to develop and apply a privacy policy risk assessment tool to assess whether existing HITs meet the recommended privacy policy standards and present guidelines to assist health professionals and service providers to ask the appropriate questions for themselves and HIT manufacturers to ensure that they can confidently promote the safe use of HITs as part of care.

# Methods

## Phase 1

### *Participatory Design Workshops*

Participatory design (ie, co-design) methodologies are routinely used to ensure that digital tools are designed to meet the needs of the intended user base, thus increasing uptake and engagement [18,19]. Our research team has extensive experience in the use of participatory design, including workshop design and facilitation as well as knowledge translation [19-23].

Our research team conducted a series of 10 participatory design workshops from July to September 2018 in 9 urban and rural *headspace* centers across Australia (Ashfield, Bathurst, Broken Hill, Dubbo, Orange, Wagga Wagga, and Wollongong, New South Wales [NSW]; Townsville, Queensland; Edinburgh North, South Australia). *headspace* centers are primary mental health services providing support to young Australians and their families to promote mental health and engagement with the community. The methods and results of these workshops were previously reported in detail by Cheng et al [24]. In summary, the workshops brought together key stakeholders from the participating *headspace* communities, including help-seeking young people, supportive others, health professionals, and service providers, to collaboratively discuss technology designs, ideas, and principles to support mental health and well-being. In these workshops, technology designs, ideas, and principles, including the concepts of data privacy and security, were evaluated by the participants. Due to varying numbers of participants and other contextual factors, workshops ranged from 2.5 to 4 hours in duration and consisted of the following stages: discovery, evaluation, and prototype. The discovery stage focused on current ways in which technology is used by participants, including for the purposes of supporting mental health, internet access issues in regional communities, and concerns about sharing personal and health information via web-based programs or tools. During the evaluation stage, participants were presented with images of components of the InnoWell Platform and asked to document their feedback. Finally, in the prototype stage, participants were given the opportunity to brainstorm new items, functionalities, and wireframes using sketchbooks for the components or functionalities of the InnoWell Platform. In this paper, we present findings related to the development and inclusion of privacy information in HITs as co-designed by the participants. Importantly, these results were not included in the original publication [24].

### Participants

Participants included individuals from the participating *headspace* communities, including young people attending a participating *headspace* center, a supportive other of a young person attending a participating *headspace* center (eg, family member, caregiver, or friend), or a health professional or service provider working at a participating *headspace* center. The inclusion criteria for participation in the study required participants to be aged ≥12 years, proficient in reading and speaking English, and having completed the participant consent process. Details of recruitment, screening, and informed consent processes have been previously documented by Cheng et al [24].

### Knowledge Translation

The InnoWell Platform consists of a multidimensional assessment evaluating a range of biopsychosocial domains (eg, psychological distress, sleep, alcohol use, and physical health) to provide a holistic view of the consumer. The assessment results are available in real time and designed to be reviewed collaboratively by the consumer and their health professional to promote shared decision-making about care options, accounting for consumer preferences. A consumer's progress can then be routinely tracked and monitored over time using assessment tools to inform treatment planning, clinical review, and coordinated care within and between services.

As previously described [24], workshop notes and descriptive artifacts were reviewed by an independent knowledge translation team with 2 young researchers without previous knowledge of the InnoWell Platform or the fundamental principles underpinning its design. Each team member, taking note of their general observations, reviewed all data independently and subsequently identified the key concepts noted by workshop participants to then produce prototype designs of the components of the InnoWell Platform.

### Ethics

This study was approved by the University of Sydney's Human Research Ethics Committee (protocol number: 2018/130).

## Phase 2

### Development of the Privacy Policy Risk Assessment Tool

Drawing from digital health privacy and security criteria published by existing research and professional associations, our research team developed a privacy policy risk assessment tool (Multimedia Appendix 1) to evaluate privacy policies for HITs, including apps and e-tools. Developed in 2018 by a multidisciplinary expert panel of health professionals, informaticists, medical students, and consumers with lived experience of mental illness, the American Psychiatric Association app evaluation model includes risk, privacy and safety questions as part of their simple four-stage hierarchal app review process [25]. Although the American Psychiatric Association model served as the primary reference, to ensure an all-encompassing assessment tool, we also incorporated details from the National and Safety Quality Digital Mental Health Standards developed by the Australian Commission on Safety and Quality in Health Care to improve the quality of digital mental health care and protect the privacy of service users [6]. The final privacy policy risk assessment tool consists of 23 items covering (1) privacy policy (6 items), (2) personal health information (PHI; 4 items), (3) data security and storage (9 items), and (4) other aspects of privacy (4 items). As the questions and concepts included in the privacy policy risk assessment tool were drawn from previously published privacy and security standards, we were confident that the measure has face and construct validity; however, a specific validity analysis was not conducted.

### Review of Digital Health Tool Privacy Policies

Importantly, the privacy policy risk assessment tool is broadly applicable to HITs. To demonstrate its utility, in this study, we evaluated the privacy policies of the apps and e-tools in the *youth* configuration of the InnoWell Platform [16]. Within the InnoWell Platform, there are two types of care options: clinical and nonclinical. Clinical care options require health professionals' involvement, such as individual therapy and group therapy, whereas a consumer can immediately access and begin using nonclinical care options, such as apps and e-tools, without the support of a health professional [26]. During the co-design process, care options are tailored to the consumer population; in this case, young people receiving care through primary youth mental health services (eg, *headspace* centers).

Aligned with established evaluation processes [15], privacy policies and any related material that may contain privacy-related content, such as terms and conditions, were sourced from associated websites and app store links. A nonrestrictive process was used, allowing all hyperlinks from the app store or within the privacy policy to be considered. All available information was collated and reviewed in accordance with the privacy policy risk assessment tool developed primarily by a member of the research team (AER), with support from a research intern (Toby Wong).

### *Assessing Readability*

The readability levels of each privacy policy were assessed as part of the evaluation process. There are multiple readability formulas available; however, for the purpose of this study, we used the Flesh-Kincaid readability tests because of their acceptance in the health care literature [27]. The readability tests were designed to indicate how difficult an English passage of writing is to understand using a formula calculated from the average number of syllables per word and the average number of words per sentence [28]. Both the Flesch reading ease and the Flesch-Kincaid reading grade level were calculated using the web-based readability tools [29]. The scores indicate the readability of a passage of text on a scale from “very easy to read" to "very difficult to read” and “fifth grade reading level” to “professional reading level,” respectively. Scores were calculated using the first two paragraphs of the privacy policies. The paragraphs (at least 100 words) were copied into a readability formula calculator, and a score was given.

### *Data Analysis*

Descriptive statistics were used to analyze all aspects of the assessment data. SPSS version 25 (IBM Corp) was used for all analyses.

## Phase 3

### *Development of Privacy Guidelines: Consultations*

As a result of Project Synergy and the development of the InnoWell Platform, a set of core principles and privacy guidelines were used as the starting point to formalize a more encompassing set of privacy guidelines. A series of consultations were held between 2014 and 2016 (phase 1) and then again between 2017 and 2018 (phase 2) to develop a set of privacy guidelines for Project Synergy. Initial consultations were conducted by Orygen (the National Centre of Excellence in Youth Mental Health) [30]. The subsequent consultations were conducted by the Project Synergy research and development team (led by the authors TAD and IBH). These consultations were held at the University of Sydney’s Brain and Mind Centre, either in-person or via teleconference, and brought together key stakeholders across relevant organizations.

### *Participants*

Participants included key stakeholder groups, including Orygen (the National Centre of Excellence in Youth Mental Health), the Young and Well Cooperative Research Centre, Mental Health Commission of NSW (Pacific Privacy Consulting Pty Ltd), the Project Synergy research and development team (the University of Sydney’s Brain and Mind Centre), InnoWell, and PwC (Australia). Select individuals were nominated by each organization, with participants contributing diverse expertise and experience, such as the involvement of 2 ex-serving privacy commissioners for NSW and Victoria (Australia).

### *Preliminary Development of Privacy Guidelines for Phase 1 Project Synergy (2014-2016)*

From the outset of the development cycle of the prototype and as part of phase 1 of Project Synergy (2014-2016) [16], this ongoing series of consultations were envisaged to inform the development of the privacy guidelines. Therefore, an initial set of guidelines was developed; these guidelines were produced by Orygen and supported by the Young and Well Cooperative Research Centre and the Mental Health Commission of NSW [30]. They were also reviewed by the Project Synergy research and development team at the University of Sydney’s Brain and Mind Centre (the authors) and Pacific Privacy Consulting.

### *Ongoing Development of Privacy Guidelines for Phase 2 of Project Synergy (2017-2020)*

As a result of the review of the initial guidelines developed in phase 1 of Project Synergy, a narrower focus was decided upon and used as the starting point for the development of more encompassing privacy guidelines for phase 2 of Project Synergy (2017-20). Specifically, upon review of the initial guidelines whereby the University of Sydney and Pacific Privacy Consulting determined that privacy concerns were the most important priority, a narrower focus was given to 8 core foundation principles to be followed by organizations using the prototype in phase 1 of Project Synergy (2014-2016) [16]. The core foundation principles included responsibility for legal compliance, anonymous or pseudonymous services wherever practicable, individual control, transparency, interaction with individuals, encryption, deidentification, and cross-border processing risk. Development of the broad structure and content of the guidelines was guided by the Project Synergy research and development team, with review and input by Pacific Privacy Consulting and InnoWell.

This paper presents privacy guidelines to assist health service providers in considering the privacy of their consumers when using HITs as part of care. The guidelines were first drafted by the Project Synergy research and development team based on the information gathered through the initial collaborative consultation process. The checklist was then reviewed, discussed, and evaluated by the research team, ultimately resulting in agreement by consensus.

# Results

## Knowledge Translation Findings

The results of the knowledge translation process highlighted that participants wanted privacy information to be presented before being required to create an account. Specifically, they emphasized the need for privacy information to be readily available, allowing a user to be completely comfortable when entering more sensitive information into a HIT, such as the InnoWell Platform (eg, “Always ask could this site be more secure with my information” [Wollongong workshop]). This included the ability to change permissions concerning data sharing at their discretion (eg, “[I] would want privacy settings in place so that not everyone that shares the system can see" [Broken Hill workshop]). Participants noted that privacy information is frequently confusing and difficult to understand, leaving them unsure whether they should trust the HIT to protect their personal and health information. Thus, multiple participants suggested a pin code or password (eg, "Consider password security like in bank apps." [Townsville workshop]) to access certain data so the consumer controls who has access to their information in the InnoWell Platform. Importantly, the idea of consumer control extended beyond HIT manufacturers such as InnoWell and included health professionals and supportive others (ie, family members and carers) accessing personal and health information (eg, "Need privacy setting like Facebook...can filter who can see the information" [Bathurst workshop]).

## App and e-Tool Privacy Policy Assessment

We evaluated 34 privacy policies using the privacy policy risk assessment tool. Most of these apps and e-tools were designed for both youth and adult users (28/34, 82%), whereas the remaining 18% (6/34) were specifically designed for youth audiences (aged ≤25 years). Most apps and e-tools (20/34, 59%) were self-help or self-management tools supporting mental health and well-being, including three specifically using cognitive behavioral therapy techniques. There were also 12% (4/34) symptom trackers, 6% (2/34) web-based counseling services, 6% (2/34) planning and time management tools, 6% (2/34) psychoeducational websites, and 2% (1/34) mindfulness and meditation app. The remaining apps and e-tools supported fitness (2/34, 6%) and relationships (1/34, 2%).

## Summary of Privacy Policy Information

### *Overview*

The summary results from the review of privacy policies are presented in Table 1. All apps and e-tools had privacy policies. Some of the policies were not readily accessible directly from the app or e-tool but rather were hosted on an external website that the app or e-tool privacy policy fell under (ie, a privacy policy for a hospital or government department). Importantly, most of the policies (26/34, 76%) explicitly stated that they met the standards of the Privacy Act 1988 (Australia) or international equivalent (ie, Health Insurance Portability and Accountability Act). In addition, most manufacturers (31/34, 91%) introduced the purpose of the privacy policy, stating that the policy explained the manufacturers' approach to privacy, protection, and management of personal information. Similarly, 97% (33/34) of privacy policies were noted to provide adequate information for all potential users. In contrast, manufacturers (26/34, 76%) frequently did not provide adequate information about their organization and how or why they operated.

**Table 1.** Summary results of the privacy policy assessment results (N=34).

| Privacy policy questions and responses | Value, n (%) |
|---|---|
| **Is there a privacy policy?** | |
| No | 0 (0) |
| Yes | 34 (100) |
| **Does the app or e-tool claim to meet the standards of the Privacy Act 1988 (Australia), HIPAA[a] (United States), or another international equivalent?** | |
| No | 8 (24) |
| Yes | 26 (76) |
| **Did the manufacturer introduce the purpose of the privacy policy?** | |
| No | 3 (9) |
| Yes | 31 (91) |
| **Does the privacy policy provide an introduction to the organization, including its vision and purpose?** | |
| No | 26 (76) |
| Some information | 3 (9) |
| Yes | 5 (15) |
| **Does the privacy policy provide adequate information (both targeted and general) relevant for all users, including consumers seeking care and health professionals?** | |
| No | 0 (0) |
| Some information | 1 (3) |
| Yes | 33 (97) |

[a] HIPAA: Health Insurance Portability and Accountability Act.

### Readability of Privacy Policies

In relation to the Flesch reading ease, the privacy policies were all found to fall into the top three of the seven available score categories, with 9% (3/34) rated as *fairly difficult to read*, an additional 53% (18/34) rated as *difficult to read*, and the remaining 38% (13/34) rated as *very difficult to read.* The Flesh-Kincaid grade-level test illustrated similar results with only the top two levels represented out of eight possible levels. All but three policies fell into the *college graduate* grade level (31/34, 91%), with the remaining considered *professional-level* reading difficulty (3/34, 9%).

### Collection of PHI

Table 2 presents results from the PHI assessment. Most apps and e-tools (32/34, 94%) collected some form of PHI, ranging from simple demographic information to more sensitive information such as information related to mental health. The remaining 6% (2/34) of apps and e-tools were considered informational tools and collected data such as email addresses. Of the 32 apps and e-tools that collected PHI, most (28/32, 87%) shared this information in some manner (data sharing is reviewed in greater detail below). This use was disclosed in most privacy policies (27/32, 85%), explicitly stating that reasonable steps were taken to ensure the security of the PHI; however, for 6% (2/32) and 9% (3/32) of privacy policies how PHI was shared was *somewhat clear* or *not clear*, respectively. Although more than half of the privacy policies (17/32, 53%) clearly stated how and when PHI was deleted, this information was either *somewhat clear* or *not clear* in the remaining 3% (1/32) or 44% (14/32) privacy policies, respectively.

**Table 2.** PHI[a] assessment results (N=34).

| PHI questions and responses | Value, n (%) |
|---|---|
| **Does the app or e-tool collect PHI (ie, demographic information, medical histories, test and laboratory results, mental health conditions, or insurance information)?** | |
| No | 2 (6) |
| Yes | 32 (94) |
| **Is PHI (ie, demographic information, medical histories, test and laboratory results, mental health conditions, or insurance information) shared?** | |
| No | 4 (13) |
| Yes | 28 (87) |
| **Is it clear if the organization has taken reasonable steps to ensure the security of PHI?** | |
| Not clear | 3 (9) |
| Somewhat clear | 2 (6) |
| Yes—clear | 27 (85) |
| **Is it clear how and when their organization will delete PHI?** | |
| No | 14 (44) |
| Somewhat clear | 1 (3) |
| Yes | 17 (53) |

[a] PHI: personal health information.

### Data Sharing and Use of Information

Results of the assessment questions related to data sharing, data preferences, and data storage are presented in Table 3. Most privacy policies (32/34, 94%) declared how data were used and for what purposes, with approximately two-thirds (22/34, 65%) stating that users were allowed to delete data. Notably, only 3% (1/34) of apps allowed data sharing preferences to be changed, although this required the user to contact the manufacturer via email; 9% (3/34) of apps and e-tools enabled users with some permissions specific to data sharing, such as receiving push notifications or allowing the user to choose whether to share data such as location with the mobile app.

**Table 3.** Data sharing and data use assessment results (N=34).

| Data sharing and data use questions and responses | Value, n (%) |
|---|---|
| **Does the privacy policy declare data use and purpose?** | |
| No | 2 (6) |
| Yes | 32 (94) |
| **Is shared data deidentified? (ie, is data anonymous—is personal information masked or severed from the identity of the contributor)** | |
| No | 11 (33) |
| Yes | 23 (67) |
| **Can the user change their preferences regarding data sharing (ie, switch it on or off)?** | |
| No | 30 (88) |
| Yes | 1 (3) |
| Some permissions | 3 (9) |
| **Can the user delete their data from the app or e-tool?** | |
| No | 10 (29) |
| Yes | 22 (65) |
| Informational webpage only | 2 (6) |
| **Is user data stored on the device?** | |
| No | 27 (79) |
| Yes | 4 (12) |
| Unspecified | 3 (9) |
| **Is user data stored on a server?** | |
| Yes | 32 (94) |
| Unspecified | 2 (6) |
| **For how long is user data stored?** | |
| Unspecified or unclear | 12 (35) |
| Until no longer needed | 12 (35) |
| 1 year or less | 2 (6) |
| 1-3 years | 3 (9) |
| More than 3 years | 3 (9) |
| At user discretion | 2 (6) |
| **What type of server is used to store user data? (eg, Amazon Web Services, within Australian borders)** | |
| Secure Australian server | 6 (19) |
| Secure overseas server | 11 (34) |
| Unspecified server | 7 (22) |
| University server | 2 (6) |
| Hospital or PHN[a] | 2 (6) |
| Unclear | 4 (13) |
| **In what country is the server located ?** | |
| Australia | 13 (38) |
| United States | 9 (26) |
| Canada | 3 (9) |
| Europe | 1 (3) |
| Multiple countries | 2 (6) |
| Unclear or unspecified | 6 (18) |

[a] PHN: Primary Health Network.

Most apps and e-tools (32/34, 94%) stored data on a server, with a small number (4/34, 12%) storing data on both the device (ie, PHI) and a server (ie, email address and website activity). In addition, 6% (2/34) of apps and e-tools did not specify where the data were stored. Data storage duration ranged from up to 1 year (2/34, 6%), 1 to 3 years (3/34, 9%), and more than 3 years (3/34, 9%). Although more than one-third of the apps and e-tools were unclear or did not specify for how long data were stored (12/34, 35%), approximately one-third (12/34, 35%) stored the data until no longer needed by the organization. The remaining 6% (2/34) of privacy policies stated that the data would be deleted at the user's discretion.

Of the 32 apps and e-tools that stored data on a server, 11 (34%) stored data on a secure overseas server, 6 (19%) stored data on a secure Australian server, 7 (22%) stored data on an unnamed or unspecified server. In addition, 6% (2/32) of apps and e-tools stored data on a university server, 6% (2/32) stored data on a hospital or primary health network server system, and 13% (4/32) were unclear on the type of server used. The location of data storage was mixed between Australia (13/34, 38%), the United States (9/34, 26%), Canada (3/34, 9%), Europe (1/34, 3%), and multiple locations (2/34, 6%), with the remainder unclear as to where data were stored (6/34, 18%).

### Review of How Data Are Shared

Most apps and e-tools (27/34, 79%) shared data with relevant third parties, including but not limited to partners, suppliers, collaborators, advisers, and business associates. The types of data shared varied from PHI to aggregated user data, such as location. A small number of apps and e-tools (4/34, 12%) shared information with irrelevant third parties, including social media platforms such as Facebook. In addition, 38% (13/34) of apps and e-tools shared data with a research partner or university, 15% (5/34) shared information with government departments, and 38% (13/34) shared data with a health-related group or person (eg, a support person or health professional).

### Additional Information

All privacy policies were assessed for their inclusion of various other details, which are summarized in Table 4. More than half of the apps and e-tools (20/34, 59%) used third-party vendors, such as Google Analytics, to evaluate and track consumer use of websites, collect demographic data, and evaluate other information related to the apps or e-tool website and the user's device. Less than one-quarter of privacy policies (8/34, 24%) adequately explained how the manufacturer would respond to a data breach, although most policies (31/34, 91%) provided some details as to how to provide feedback or lodge a complaint either with the manufacturer or through an expert third party (ie, Office of the Australian Information Commissioner). Finally, only 24% (8/34) of privacy policies explicitly warned individuals about privacy risks involved in accessing services that are outside the control of the service provider, such as third-party advertisements, with an additional 9% (3/34) of policies providing some detail in this regard.

**Table 4.** Overview of additional details provided by privacy policies (N=34).

| Additional questions and responses | Value, n (%) |
|---|---|
| **Does the app or e-tool use third-party vendors (eg, Google Analytics)?** | |
| No | 11 (32) |
| Yes | 20 (59) |
| Unspecified | 3 (9) |
| **Is the manner in which the organization will respond to a data breach adequately explained?** | |
| No | 26 (76) |
| Yes | 8 (24) |
| **Does the privacy policy inform users how they can make inquiries and provide feedback and lodge complaints, including both contact details for the relevant party within the organization and a third-party expert (eg, Office of the Australian Information Commissioner)?** | |
| No—does not provide either | 3 (9) |
| Some—provides information for internal or third party only | 16 (47) |
| Yes—provides both internal and third-party expert | 15 (44) |
| **Does the privacy policy explicitly warn users about privacy risks involved in accessing services offered that are outside the control of the organization ?** | |
| No | 23 (67) |
| Somewhat | 3 (9) |
| Yes | 8 (24) |

## Privacy Considerations for Health Services

As generated through the collaborative consultation process described previously, Textbox 1 presents privacy guidelines to assist health professionals and service providers to ask the appropriate questions—of themselves and to HIT manufacturers—before confidently promoting the safe use of HITs as part of mental health care.

**Textbox 1.** Privacy guidelines—health professional and service provider considerations regarding the use of health information technologies (HITs) for care.

**Privacy guidelines**

- The HIT manufacturer has clearly introduced the purpose of its privacy policy.
- The privacy policy includes an introduction to the HIT manufacturer and how and why their organization operates.
- The privacy policy provides adequate information and addresses my concerns.
    - If no, I am aware who I need to contact to seek clarification...
- The privacy policy is written clearly.
    - If no, I am aware who I need to contact to seek clarification...
- The privacy policy adequately explains how the HIT manufacturer will collect and use personal data.
- The privacy policy adequately explains how and when the HIT manufacturer will disclose personal data to third parties.
    - If the HIT manufacturer shares data with third parties, I am confident that the third-party partners are reputable and will comply with all legislative requirements when collecting, storing, and sharing data.
    - If no, I am aware who I need to contact to seek clarification...
- I am confident the HIT manufacturer has taken the appropriate steps to protect everyone’s data, adopting the strongest security measures.
- I have been made aware of how end users can access and correct their personal information on the HIT.
    - If no, I am aware who I need to contact to seek clarification...
- The privacy policy outlines how and when the HIT manufacturer will delete personal data.
- The privacy policy outlines how the HIT manufacturer will respond to any data breaches.
- The privacy policy includes information on how I can inquire, provide feedback, or make complaints.
- There is the opportunity for me to contact a third-party expert to inquire about the privacy policy (such as the Office of the Australian Information Commissioner).
    - If yes, they are...
- From what I read, I feel comfortable using the proposed HIT as part of my clinical care and/or practice.

# Discussion

## Listening to Consumers

With their increased experience and exposure, consumers are becoming more sophisticated users of HITs. They can offer valuable insights into how privacy information should be presented to ensure that it is clear, informative, and transparent. Participants of our co-design workshops highlighted the need for all HITs to have a privacy policy that provides relevant data security information before collecting information from an individual, preceding the account creation process. In addition, participants stated that privacy policies should be accessible, transparent, and user-friendly, ensuring that consumers understand what personal and health data will be collected, stored, and shared and, in turn, enabling them to trust the HIT to protect their data. These findings align with those reported by Schueller et al [31], which indicated that 70% of mental health app users rate the inclusion of a privacy policy and data encryption as important. Of note, when a mental health app was deemed to be from a *trusted source* (no shared definition of what constituted a trusted source), users assumed that the app adequately protected their data [31], potentially leaving their data vulnerable to unrecognized data sharing pathways.

## Privacy Policy Risk Assessment

As evidenced by our co-design results, consumers are calling for greater clarity and transparency in the privacy policies of HITs so that they can be confident that they understand how their personal and health information may be used. Importantly, all apps and e-tools included in this study had a privacy policy. All but 1 of those policies provided explicit details explaining the manufacturer’s approach to privacy and how personal information is managed and protected. However, approximately one-quarter of the privacy policies did not meet the standards of the Australian Privacy Act 1988 or other international equivalents, raising concerns regarding undisclosed data sharing and poorly secured data storage. Even when the use of data adheres to privacy standards, issues of transparency often arise. For example, a recent review of the data sharing practices of 24 health-related apps found that data were shared with 55 unique entities, including app developers, their parent companies, and third parties (ie, service providers). Subsequently, third parties shared user data with an additional 216 integrated fourth parties (eg, Facebook sharing data with data brokers to enable targeted advertising) [32].

In addition to poor overall transparency, our results also confirm that privacy policies, when present, are fairly difficult to read and require a college or professional reading level, essentially rendering them useless for a large portion of potential users (eg, children and young people or individuals with cognitive impairments or intellectual disabilities). This aligns with previous research by Robillard et al [33], which highlighted that the readability of privacy policies of mental health apps is typically written at too high a level for the general population. Strikingly, a longitudinal review of privacy policies found both a decline in readability and a marked increase in length [34]. To improve the readability of privacy policies, it is suggested that HIT manufacturers either compare different versions of their policy to determine which one has the best readability score or that the policy be rewritten until it meets a predetermined grade level. Researchers have suggested that a grade 9-10 reading level is likely to be appropriate for the general population [35].

Most apps and e-tools included in this study collected some form of PHI (32/34, 94%), including, in some cases, information related to mental health, with 87% (28/34) of those apps and e-tools then sharing these data in some manner. Although data sharing was disclosed in most privacy policies (27/32, 84%), how PHI was shared was not transparent in 15% (5/32) of the policies. Most apps and e-tools shared data with relevant third parties (27/34, 79%); however, 12% (4/34) also shared information with third parties deemed to be irrelevant, such as social media platforms (ie, Facebook). Of note, few apps and e-tools (4/31, 13%) allowed users to update their permissions concerning data sharing.

Although it is unlawful in Australia, for example, to share PHI for purposes other than those stated in a privacy policy, the complexities of the web-based environment frequently preclude a full understanding of how data are shared and for what purpose [36]. Users must recognize that shared data enters into a *supply chain*, being passed from apps and e-tools to parent companies and then on to third parties such as data trackers, aggregators, and brokers [37]. Subsequently, data may be sold to researchers and government agencies for advertising purposes. In addition to driving targeted advertising campaigns, this aggregated data may also be used to influence employment and financial and insurance decisions, with potentially marked consequences at the individual level [37], potentially leading to incarceration or human rights abuses in some countries [11]. Given the risks described above, explicit and transparent documentation of how data are shared is critical to ensure that users are able to provide informed consent. Furthermore, it is critical that governance structures and regulatory standards are established globally to ensure ethical practice in digital mental health care, including the handling of PHI. As advised by the World Economic Forum, regulations should not be designed exclusively by governments but rather in collaboration with consumers with lived experience, mental health professionals, technology manufacturers, and policy makers, with the aim of facilitating efficient access to effective and safe digital tools to address growing mental health needs [11].

Most apps and e-tools in our sample stored data on a server (32/34, 94%), with more than half (21/32, 66%) storing data on an unnamed or unspecified server. Although other server types included university servers (2/32, 6%) and a hospital or primary health network server system (2/32, 6%), the type of server used to store data for an additional 22% (7/32) of apps and e-tools was unclear. Once data are transmitted to a third-party server, it is often difficult to determine the robustness of the privacy and confidentiality standards in place to protect the PHI. For example, Cultura Colectiva, a digital media company with access to user information from Facebook, stored data on a publicly accessible server, resulting in the exposure of 540 million individual records, including user IDs and names [38]. In addition, although it is ideal that all information being transmitted is encrypted, personal and health information may still be visible in server logs with few restrictions in terms of access [39]. Notably, when different apps and e-tools use the same server, it may be possible to link different PHIs together to create digital profiles of users [40] with potentially negative repercussions such as impacts on employment and insurance. Despite the frequency of data breaches [41], 76% (26/34) of the privacy policies reviewed in this study failed to document how the organization would respond to data breaches, leaving the user to wonder what steps might be taken to protect their data from exposure and misuse (ie, identify theft).

### Privacy Guidelines for Health Professionals and Service Providers

Given the concerns regarding accessibility, transparency, and readability outlined above, through a consultative process with key stakeholders, our team developed privacy guidelines (Textbox 1) to prompt health professionals and service providers to ask informed questions when reviewing an HIT privacy policy to ensure adequate data privacy and security measures are in place. The guidelines aim to support health professionals and service providers to confidently promote the safe use of HITs as part of care and within the broader service. The privacy guidelines have considered the privacy policy of HIT manufacturers to emphasize the importance of building trust between users and HIT manufacturers through transparency [42]. Thus, it is the responsibility of the HIT manufacturer to be aware of all (both current and emerging) regulatory requirements and best practice principles [42] to ensure that the privacy policy is communicated to all users. This not only minimizes potential harm but also allows users to be well informed and to have more control when consenting to use HITs.

As few consumers will review academic literature before accessing HITs, they are more likely to learn about available apps and e-tools via the internet, app stores, social media, word of mouth, or health professionals. In relation to the latter, it is recommended that health professionals and service providers conduct their own risk assessment before implementing HITs into their service to ensure appropriate risk strategies are in place [3]. Not only should providers have an understanding of privacy risks but it is also important that they work with consumers to ensure they are aware of the potential for privacy breaches to ensure they are providing informed consent before engaging with an HIT [37].

### Limitations

This study has some limitations that are important to note. Although we engaged in a thorough collaborative consultation process to develop the broad structure and content of the privacy guidelines for Project Synergy, the development of the privacy policy risk assessment tool and the guidelines for health professionals and service providers was conducted by the research team, independent of this broader stakeholder group. Therefore, we acknowledge that both the assessment tool and the guidelines may benefit from further input or revision by individuals with expertise in data privacy and security, both from a legal perspective and regarding manufacturers of digital tools. As highlighted by the co-design work presented in phase 1, our group recognizes the importance of including the voice of those with lived experience in our work to reform mental health services and systems of care, including the ethical development and application of digital tools. With that being said, we acknowledge that the privacy policy risk assessment tool and guidelines were developed without contributions from consumers with lived experience of mental ill health.

## Conclusions

Given the increasing uptake of HITs, both by individuals for the purposes of self-management and by health professionals as a means to complement clinical services, it is essential that all users have a clear understanding of what personal and health information will be collected, how these data will be shared and stored, and what privacy and security measures are in place to ensure it is protected. Our findings highlight the ubiquitous poor readability and lack of transparency in existing privacy policies, a stark contrast to what consumers emphasized as essential factors in the presentation of privacy information. Although consumers, health professionals, and services are becoming increasingly reliant on HITs to deliver, support, or enhance care, concerns regarding the privacy of health and personal information are likely to undermine user confidence and willingness to engage with HITs. Therefore, we provide suggested guidelines that can be easily adopted by health professionals and service providers when considering the implementation of HITs, including apps and e-tools, into their service. We recommend that these guidelines be adopted to ensure that HITs are used to their full potential to maximize patient health outcomes while minimizing risk and that users are informed of privacy and security considerations to make educated decisions as to whether they would like to share their personal and health information.

## Acknowledgments

The authors IBH and TAD were integral in securing funding to support this study. This study was designed by HML, AER, TAD, and GL. All data analyses were conducted by HML and AER. All authors have contributed to and approved the final manuscript. This research was conducted on behalf of the Australian Government Department of Health as part of Project Synergy (2017-20). InnoWell was formed by the University of Sydney and PricewaterhouseCoopers (Australia) to deliver the Aus $30 million (US $22.1 million) Australian government-funded Project Synergy. The authors would also like to acknowledge Future Generation Global for funding the Youth Mental Health & Technology Program, which aims to improve young people's access to quality mental health care, including through the use of co-design methodologies. The authors would like to acknowledge and thank Hannah Yee for her contributions to the knowledge translation and Toby Wong for his assistance with data collection. The authors would also like to thank the participants in the co-design workshops for their insights on how to best present privacy information to users.

## Conflicts of Interest

IBH is the codirector of health and policy at the Brain and Mind Centre, University of Sydney. The Brain and Mind Centre operates an early-intervention youth service 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 PricewaterhouseCoopers (Australia; 45% equity) to deliver the Aus $30 million (US $22.1 million) 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. TAD is now the director of research and evaluation at the Design and Strategy Division of the Australian Digital Health Agency. The funding source does not entail any potential conflicts of interest for the other members of the Project Synergy research and development team.

## Multimedia Appendix 1

Privacy risk assessment tool.
[DOCX File , 16 KB-Multimedia Appendix 1]

## References

1. Patient-centric technology improves access, efficiency, and quality of care. Wolters Kluwer. 2018 Apr 24. URL: https://www.wolterskluwer.com/en/expert-insights/patient-centric-technology-improves-access-efficiency-and-quality-of-care [accessed 2021-08-06]
2. Health technology in Australia. Australian Government Department of Health. URL: https://www.health.gov.au/health-topics/health-technology [accessed 2021-08-06]

3. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
4. Davenport TA, Cheng VW, Iorfino F, Hamilton B, Castaldi E, Burton A, et al. Flip the clinic: a digital health approach to youth mental health service delivery during the covid-19 pandemic and beyond. JMIR Ment Health 2020 Dec 15;7(12):e24578 [FREE Full text] [doi: 10.2196/24578] [Medline: 33206051]
5. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. Front Psychiatry 2019 Aug 23;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]
6. National Safety and Quality Digital Mental Health (NSQDMH) standards. Australian Commission on Safety and Quality in Health Care. 2020 May 29. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-04/UPDATED%20national_safety_and_quality_digital_mental_health_standards_-_consultation_draft_-_feb_2020_0.pdf [accessed 2020-10-16]
7. Tips for choosing a digital mental health service - information for consumers and carers. Australian Commission on Safety and Quality in Health Care. 2020. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-05/tips_for_choosing_a_digital_mental_health_service_-_consumers_and_carers_fact_sheet_-_may_2020.pdf [accessed 2020-10-16]
8. Tips for choosing a digital mental health service - information for clinicians. Australian Commission on Safety and Quality in Health Care. 2020. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-05/tips_for_choosing_a_digital_mental_health_service_-_clinicians_fact_sheet_-_may_2020.pdf [accessed 2020-10-16]
9. Read the Australian privacy principles. Australian Government Office of the Australian Information Commissioner. 2014. URL: https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles/ [accessed 2021-08-06]
10. Privacy Act 1988. Australian Government Federal Register of Legislation. 1988. URL: https://www.legislation.gov.au/Details/C2018C00456 [accessed 2020-08-28]
11. Empowering 8 billion minds enabling better mental health for all via the ethical adoption of technologies. World Economic Forum. 2019 Jul. URL: http://www3.weforum.org/docs/WEF_Future%20Council_Mental_Health_and_Tech_Report.pdf [accessed 2021-08-06]
12. Granwal L. Penetration rate of health applications on mobiles and tablets in Australia 2014-2018. Statista. 2020 Apr 09. URL: https://www.statista.com/statistics/985116/use-health-apps-mobile-tablet-australia/#statisticContainer [accessed 2020-10-16]
13. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
14. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
15. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
16. 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 7:3-39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
17. LaMonica HM, Davenport TA, Braunstein K, Ottavio A, Piper S, Martin C, et al. Technology-enabled person-centered mental health services reform: strategy for implementation science. JMIR Ment Health 2019 Sep 19;6(9):e14719 [FREE Full text] [doi: 10.2196/14719] [Medline: 31538938]
18. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]
19. Davenport TA, Milton AC, Ospina-Pinillos L, Whittle L, Ricci CS, Burns JM. Project Synergy research and development cycle: iterative processes of participatory design, user testing, implementation and feasibility testing. Med J Aust 2019 Oct 07;211(7 Suppl):8-11 [FREE Full text]
20. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-enabled mental health service reform for open arms - veterans and families counselling: participatory design study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
21. Davenport T, Ospina-Pinillos L, Ricci C, Milton A, Burns J, Hickie I. Fit Uni Life to thrive: an online health and wellbeing prototype for young people. Med J Aust 2019 Oct 07;211(7 Suppl):12-16 [FREE Full text]

22. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
23. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021 Jan 06;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 33404509]
24. Cheng VW, Piper SE, Ottavio A, Davenport TA, Hickie IB. Recommendations for designing health information technologies for mental health drawn from self-determination theory and co-design with culturally diverse populations: template analysis. J Med Internet Res 2021 Feb 10;23(2):e23502 [FREE Full text] [doi: 10.2196/23502] [Medline: 33565985]
25. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
26. Davenport TA, LaMonica HM, Whittle L, English A, Iorfino F, Cross S, et al. Validation of the Innowell Platform: Protocol for a clinical trial. JMIR Res Protoc 2019 May 31;8(5):e13955 [FREE Full text] [doi: 10.2196/13955] [Medline: 31152524]
27. Wang L, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
28. Kincaid J, Fishburne JR, Rogers R, Chissom B. Derivation of New Readability Formulas: (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel. Institute for Simulation and Training, University of Central Florida, Millington. 1975. URL: https://stars.library.ucf.edu/istlibrary/56/?utm_source=stars.library.ucf.edu%2Fistlibrary%2F56&utm_medium=PDF&utm_campaign=PDFCoverPages [accessed 2021-08-24]
29. Automatic readability checker. Readability Formulas. 2021. URL: https://readabilityformulas.com/free-readability-formula-tests.php [accessed 2021-08-12]
30. Robinson J, Donaldson A, Hetrick S, Rice S, Bailey E. Guidelines for Project Synergy: Development of a set of medico-legal ethical guidelines. Young and Well CRC, Melbourne. 2016. URL: https://www.youngandwellcrc.org.au/ [accessed 2021-08-25]
31. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
32. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. Br Med J 2019 Mar 20;364:l920 [FREE Full text] [doi: 10.1136/bmj.l920] [Medline: 30894349]
33. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Mar 06;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
34. Milne GR, Culnan MJ, Greene H. A longitudinal assessment of online privacy notice readability. J Public Policy Mark 2006 Sep 1;25(2):238-249. [doi: 10.1509/jppm.25.2.238]
35. Krumay B, Klar J. Readability of privacy policies. In: Singhal A, Vaidya J, editors. Data and Applications Security and Privacy XXXIV. Cham: Springer; 2020:388-399.
36. Mobile privacy: a better practice guide for mobile app developers. Office of the Australian Information Commission. 2014 Sep 05. URL: https://www.oaic.gov.au/agencies-and-organisations/guides/guide-for-mobile-app-developers [accessed 2020-10-22]
37. Martin K. Ethical issues in the big data industry. In: Strategic Information Management. Oxfordshire, England: Routledge; Apr 2020:450-471.
38. Lecher C. Facebook app developers leaked millions of user records on cloud servers, researchers say. The Verge. 2019 Apr 03. URL: https://www.theverge.com/2019/4/3/18293978/facebook-app-developers-leak-user-records-data-cloud-servers [accessed 2020-10-23]
39. He D, Naveed M, Gunter CA, Nahrstedt K. Security concerns in android mHealth apps. AMIA Annu Symp Proc 2014 Nov 14;2014:645-654 [FREE Full text] [Medline: 25954370]
40. Vallina-Rodriguez N, Sundaresan S. 7 in 10 smartphone apps share your data with third-party services. The Conversation. 2017 May 30. URL: https://theconversation.com/7-in-10-smartphone-apps-share-your-data-with-third-party-services-72404 [accessed 2021-08-06]
41. 2019 healthcare data breach report. HIPAA Journal. URL: https://www.hipaajournal.com/2019-healthcare-data-breach-report/ [accessed 2020-10-23]
42. Parker L, Karliychuk T, Gillies D, Mintzes B, Raven M, Grundy Q. A health app developer's guide to law and policy: a multi-sector policy analysis. BMC Med Inform Decis Mak 2017 Oct 02;17(1):141 [FREE Full text] [doi: 10.1186/s12911-017-0535-0] [Medline: 28969704]

## Abbreviations

**HIT:** health information technology
**NSW:** New South Wales
**PHI:** personal health information
**PwC:** PricewaterhouseCoopers

---

*Edited by R Kukafka; submitted 07.12.20; peer-reviewed by N Shen, L Ospina-Pinillos; comments to author 05.02.21; revised version received 15.02.21; accepted 04.05.21; published 16.09.21*

*Please cite as:*
*LaMonica HM, Roberts AE, Lee GY, Davenport TA, Hickie IB*
*Privacy Practices of Health Information Technologies: Privacy Policy Risk Assessment Study and Proposed Guidelines*
*J Med Internet Res 2021;23(9):e26317*
*URL: https://www.jmir.org/2021/9/e26317*
*doi: 10.2196/26317*
*PMID:*

Original Paper

# Patterns for Patient Engagement with the Hypertension Management and Effects of Electronic Health Care Provider Follow-up on These Patterns: Cluster Analysis

Dan Wu[1*], BSc; Jiye An[1*], PhD; Ping Yu[2], PhD; Hui Lin[1], BSc; Li Ma[3], MSc; Huilong Duan[1], PhD; Ning Deng[1], PhD

[1]College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China

[2]School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[3]General Hospital of Ningxia Medical University, Yinchuan, China

[*]these authors contributed equally

**Corresponding Author:**
Ning Deng, PhD
College of Biomedical Engineering and Instrument Science
Ministry of Education Key Laboratory of Biomedical Engineering
Zhejiang University
Zhouyiqing Bldg 512, Yuquan Campus
38 Zheda Rd
Hangzhou, 310027
China
Phone: 86 571 2295 269
Fax: 86 571 2295 269
Email: zju.dengning@gmail.com

## Abstract

**Background:** Hypertension is a long-term medical condition. Electronic and mobile health care services can help patients to self-manage this condition. However, not all management is effective, possibly due to different levels of patient engagement (PE) with health care services. Health care provider follow-up is an intervention to promote PE and blood pressure (BP) control.

**Objective:** This study aimed to discover and characterize patterns of PE with a hypertension self-management app, investigate the effects of health care provider follow-up on PE, and identify the follow-up effects on BP in each PE pattern.

**Methods:** PE was represented as the number of days that a patient recorded self-measured BP per week. The study period was the first 4 weeks for a patient to engage in the hypertension management service. K-means algorithm was used to group patients by PE. There was compliance follow-up, regular follow-up, and abnormal follow-up in management. The follow-up effect was calculated by the change in PE (CPE) and the change in systolic blood pressure (CSBP, SBP) before and after each follow-up. Chi-square tests and *z* scores were used to ascertain the distribution of gender, age, education level, SBP, and the number of follow-ups in each cluster. The follow-up effect was identified by analysis of variances. Once a significant effect was detected, Bonferroni multiple comparisons were further conducted to identify the difference between 2 clusters.

**Results:** Patients were grouped into 4 clusters according to PE: (1) PE started low and dropped even lower (PELL), (2) PE started high and remained high (PEHH), (3) PE started high and dropped to low (PEHL), and (4) PE started low and rose to high (PELH). Significantly more patients over 60 years old were found in the PEHH cluster ($P \leq .05$). Abnormal follow-up was significantly less frequent ($P \leq .05$) in the PELL cluster. Compliance follow-up and regular follow-up can improve PE. In the clusters of PEHH and PELH, the improvement in PE in the first 3 weeks and the decrease in SBP in all 4 weeks were significant after follow-up. The SBP of the clusters of PELL and PELH decreased more (–6.1 mmHg and –8.4 mmHg) after follow-up in the first week.

**Conclusions:** Four distinct PE patterns were identified for patients engaging in the hypertension self-management app. Patients aged over 60 years had higher PE in terms of recording self-measured BP using the app. Once SBP reduced, patients with low PE tended to stop using the app, and a continued decline in PE occurred simultaneously with the increase in SBP. The duration and depth of the effect of health care provider follow-up were more significant in patients with high or increased engagement after follow-up.

***(J Med Internet Res 2021;23(9):e25630)*** doi: 10.2196/25630

**KEYWORDS**

hypertension; health care services; mHealth; patient engagement; electronic follow-up; cluster analysis

# *Introduction*

## Background

Hypertension is a lifestyle-induced chronic disease that affects health-related quality of life. Once the disease deteriorates, it may cause complications such as stroke, myocardial infarction, cardiac failure, and renal failure [1]. Controlling hypertension requires patients to follow long-term self-management plans, including measuring and recording blood pressure (BP), taking medicine, conducting physical activities, and healthy eating. These plans should preferably be established with guidance from the health care providers and be followed continuously for hypertension control [2].

The rapid pace of electronic and mobile technology development has promoted the implementation of out-of-hospital health care services [3-5]. Mobile health (mHealth) service is defined as the use of mobile phone and wireless technologies to support medical and public health care services [6]. Abundant evidence supports the introduction of mHealth services into chronic disease control and promoting positive patient behavior [4,7-10]. Although recent studies have demonstrated efficacy for some mHealth services, some have performed poorly [11-14]. Thus, there are mixed results for using mHealth services to support patient self-management of hypertension in the community [15,16].

Patient engagement (PE) is essential for bringing an improvement of health outcomes using the mHealth app [11,17,18]. PE refers to the activation of a patient to engage with the interventions that are designed to promote positive health behavior [19]. In a study of a hypertension self-management app, Toto-Ramos et al [20] found that the patients with sustained PE, as measured by the number of weeks that the patients engaged with the app, experienced significant reduction in systolic blood pressure (SBP). Goyal et al [21] evaluated the behavior of daily blood glucose reading of patients and found a significant relationship between increased number of readings and improved glycated hemoglobin. In comparison with the traditional methods, mHealth service is advantageous in improving PE for self-management of hypertension [22,23]. For example, Kaplan et al [24] designed an mHealth app to facilitate PE in managing hypertension. The app provides a mobile platform for recording and tracking self-measured BP, periodic reminders to measure BP, and wireless BP measurement devices. Their research resulted in significant improvement in BP for patients and a high level of PE.

In the mHealth field, complex PE patterns emerge when mobile apps are used for self-management. Some patients register but never use the app, some use it intermittently, and others frequently use it for a long period of time. PE can be objectively measured as amount, duration, breadth, and depth of using the mHealth app [11], and it can also be subjectively measured as interest, intrigue, focus, inattention, enjoyment, pleasure, etc [25]. In a study of pain management with a mobile app, Rahman et al [26] used a clustering technique to identify PE patterns with the app. They measured PE by 3 key usage features: duration of app use, frequency of app use, and the number of usage records. They then used a K-means clustering algorithm to find groups of patients as indicated by PE because the PE pattern reflects the behavioral characteristics of patients using the app. Sanatkar et al [27] also used a clustering technique to analyze the mHealth app usage data so as to distinguish PE with the app in an e-mental health community. They measured 5 usage features of PE: number of user logins, number of daily trackers used, number of learning activities started and completed, and number of reminders received. Distinct usage patterns were observed in the frequency of using the app.

To date, the studies on patterns of PE and the associations between these PE patterns and health outcomes have usually used various types of static count data (eg, number of logins and number of records). However, PE is a dynamic, ongoing process [28]. The change of this dynamic process cannot be captured simply by analyzing count data captured at a single date point in the cross-sectional data analysis. Longitudinal change is useful for identifying trends through analyzing the time series data. Furthermore, as hypertension management requires long-term efforts, understanding the trends of PE is important for the long-term successful management of hypertension.

Health care provider follow-up can offer continuous and personalized attention to guide patient's self-management behavior in response to the patient's current BP level [29]. Without health care provider support, it is difficult for patients to maintain BP control and high levels of PE with an mHealth app [30], which has often resulted in a high level of dropout [31,32]. Therefore, follow-up by health care providers is recommended by the hypertension management guideline in order to know, track, and intervene in patient's hypertension self-management in the long-term [33-35]. In the traditional outpatient management model, follow-up often takes the form of home visits by health care providers to promote PE and then achieve BP control. Despite its proven benefits for hypertension management, follow-up is not fully implemented by health care providers due to the high human resources required for patient follow-up.

In comparison with the traditional format of follow-up, mHealth health care provider follow-up is a low-cost, convenient means of follow-up, and patient's BP, medication, exercise, diet, etc., can be checked in the office. In a series of our previous studies into hypertension management service [36,37], after receiving the BP data that a patient entered through the mobile app, the server could automatically conduct BP data analysis, with the web-based platform reporting the patient's BP condition to the health care provider and reminding the health care provider which patients needed to follow-up. In response to different patient health statuses, health care providers can conduct follow-up through calling and sending short SMS text messages. Our previous study designed 3 types of follow-ups: compliance follow-up, regular follow-up, and abnormal follow-up [36,37]. These were conducted to improve compliance, track BP status, redesign a new plan, and respond to the abnormal BP of patients.

To date, few studies have explored the effect of health care provider follow-up on PE and BP control in mHealth services for hypertension management [23]. It is not clear if follow-up is effective, which type of follow-up is effective, when the follow-up effect emerges, if there is a differential effect of follow-up in different patient groups, and how long the follow-up effect lasts [28,38]. Answers to these questions are essential for informing the implementation of value-based health care and providing optimal outcomes in improving the quality and reducing cost of hypertension management.

### Objective

The aim of this study was to explore the patterns of PE with a hypertension management service and the effect of health care provider follow-up on PE and BP within the first 4 weeks of mHealth app usage. This included 3 objectives: (1) to discover the patterns of PE with the mHealth app and the association between a series of related variables (age, gender, education level, the mean SBP, and the number of follow-ups) and the PE pattern, (2) to explore the health care provider follow-up effect on PE, and (3) to examine the effect of follow-up on SBP in each PE pattern.

## *Methods*

### Description of the mHealth App

Blood Pressure Assistant (BPA) is a mHealth hypertension self-management app available for patients in the General Hospital of Ningxia Medical University [36,37,39]. BPA was launched in 2015, and since then, 2129 patients have registered to use the app. The mHealth app was designed in accordance with a customized care pathway in compliance with the Chinese guideline for hypertension management. The care pathway defines tasks for hypertension management for patients and health care providers.

In this care pathway, each patient is required to register and provide basic demographic information (ie, name, gender, date of birth, and education level). The patient is then assigned to a health care provider, who is responsible for formulating a tailored management plan, conducting follow-up, and supervising patients' uploaded data. The management plan includes the frequency of self-measured and recorded BP, and recommendations for medication, physical activities, and diet. The patient's uploaded data includes self-measured BP, medications taken, physical activities, and diet records. The most important task for health care providers is to track patient's current BP level through the web-based platform and use mobile phones for patient follow-ups to assist in BP control. There are 3 types of follow-ups. (1) Compliance follow-up is performed as a response to the identified low patient compliance with the hypertension management plan [36]. In this case, health care providers need to remind the patient to measure and record BP on a regular basis. (2) Regular follow-up is conducted on regular basis in order to check the BP level and decide whether or not to maintain or update the management plan [37]. (3) Abnormal follow-up is required when patient's self-measured BP data are abnormal [36]. Health care providers need to check and understand the causes of abnormal conditions and intervene in a timely manner. With consent, the health care provider enrolls the patient they manage into the mHealth hypertension management program in an online community. Patients can use the app to check their self-management plans. Their main task is to measure and record BP data on a daily basis.

### Data Collection

#### *Statement of Ethics*

Ethical approval was granted by the Ethic Committee for the Conduct of Human Research at General Hospital of Ningxia Medical University (#NXMU-GH-2017-273). Patients in this study signed the informed consent forms.

#### *Sample*

All data were stored and extracted from the BPA server, which contains the demographic information, self-management plans, patient-uploaded data, and follow-up records of health care providers. The primary data set included 2129 patients. We selected the patient records based on the following criteria: patients were at least 18 years of age; patients registered to the app between March 27, 2016, and July 10, 2019 (as the main functions of the app were consistent during this period, this could ensure that the patient's behavior was not affected by the changes in app functions); and patients continuously measured and recorded BP more than 4 weeks after registration (as the control of BP requires at least a 4-week observation period according to the relevant guidelines) [33-35].

#### *Data Extraction*

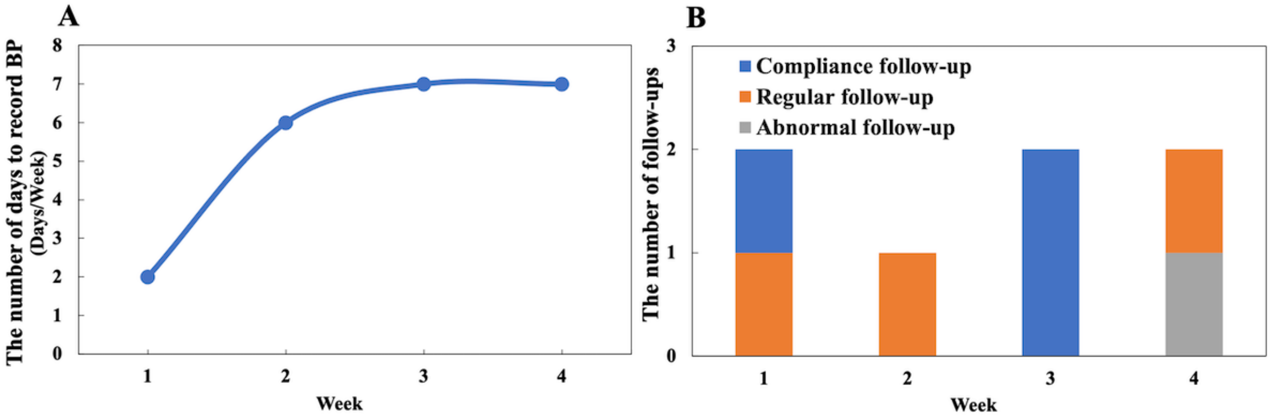
We extracted 3 types of data from the database: demographics, BP records, and follow-up records. The demographics information included the patient identification, data of birth, gender (male and female), and education level (primary or secondary school, high school, or university and above). The BP record data included SBP, diastolic blood pressure, and the uploaded date (containing year, month, day, minute, and second). Due to the high correlation between SBP and diastolic blood pressure, only SBP was used for analysis. The information specific to the follow-up records included patient ID, follow-up type (compliance follow-up, regular follow-up, and abnormal follow-up), and follow-up date.

### Data Analysis

#### *Measuring PE*

In accordance with the hypertension management guidelines [33-35], we defined the time unit of observation as 1 week. The behavior of measuring BP is a basic behavior in hypertension management because other behaviors (eg, taking medicine, doing physical activities, and eating healthy) need to be based on patient's current BP level. Therefore, in this study, patient engagement was indicated by the number of days that a patient recorded self-measured BP per week (see Figure 1). The analysis period was the first 4 consecutive weeks after the initial patient registration.

**Figure 1.** Example of patient engagement and follow-up for a patient within 4 weeks. (A) Example of patient engagement. (B) Example of the number of different types of follow-ups. BP: blood pressure.



### Cluster Analysis

Cluster analysis was conducted to group the patients into different PE patterns. K-means algorithm (Python 3; Python Software Foundation) was used to cluster patients by PE [40]. Dynamic time warping was used to calculate the similarity of PE between patients because this feature was represented as time series data [41]. We then used silhouette score to determine the optimal number of clusters [42]. The silhouette score measured the distance between clusters based on the distance between the PE of patients as determined by dynamic time warping. A higher silhouette score indicates tighter clusters, where each cluster is completely separate from the others.

### Characterizing the Clusters

An optimal clustering result was reached based on the silhouette score of a different number of clusters. For each cluster, we analyzed the demographic features (the distribution of age, gender, and education level); BP features, including the distribution of mean SBP in the first week (representing the initial BP conditions in hypertension management) and the trend of weekly mean SBP; and follow-up features (the distribution of the number of the 3 types of follow-ups in each of the 4 weeks; see Figure 1).

### The Effect of 3 Types of Follow-ups on PE and SBP

The follow-up effect was calculated by the change in PE (CPE) and the change in SBP (CSBP) before and after each follow-up, which were defined as follows:

$$CPE = \frac{\sum_1^N (PE_{i+1} - PE_i)}{N}$$

$$CSBP = \frac{\sum_1^N (SBP_{i+1} - SBP_i)}{N}$$

where $PE_{i+1}$ is $PE$ in the week after a follow-up event, $PE_i$ is PE in the week of the follow-up even, $SPB_{i+1}$ is the mean SBP in the week after a follow-up event, $SPB_i$ is the mean SBP in the week of the follow-up event, and N is the number of follow-ups performed by health care providers.

Statistical analysis was conducted in SPSS version 24 (IBM Corp). Chi-square test was performed to evaluate the statistical significance of associations between the clusters and the discrete variables (demographic, BP, and the number of follow-ups). Pairwise comparisons were conducted using the *z* scores to compare difference in the proportion of discrete variables between the clusters. Analysis of variance (ANOVA) was employed to analyze the changes in outcome parameters (CPE and CSBP) after follow-up in each of the 4 weeks. Once a significant change was detected, Bonferroni multiple comparisons were further conducted to examine the differences between the clusters. A *P* value <.05 was used to determine whether the difference was statistically significant.

## Results

### PE Patterns

A total of 562 patients met the selection criteria and were included in the study. We found that the silhouette score was the highest with 4 clusters of patients (see Figure 2). Hence, we accepted the 4-cluster output of K-means (see Table 1) for further analysis. The PE of each cluster within 4 weeks was significantly different (*P*<.001).

**Figure 2.** Comparison of the silhouette score for the different numberd of clusters (range from 3 to 7).

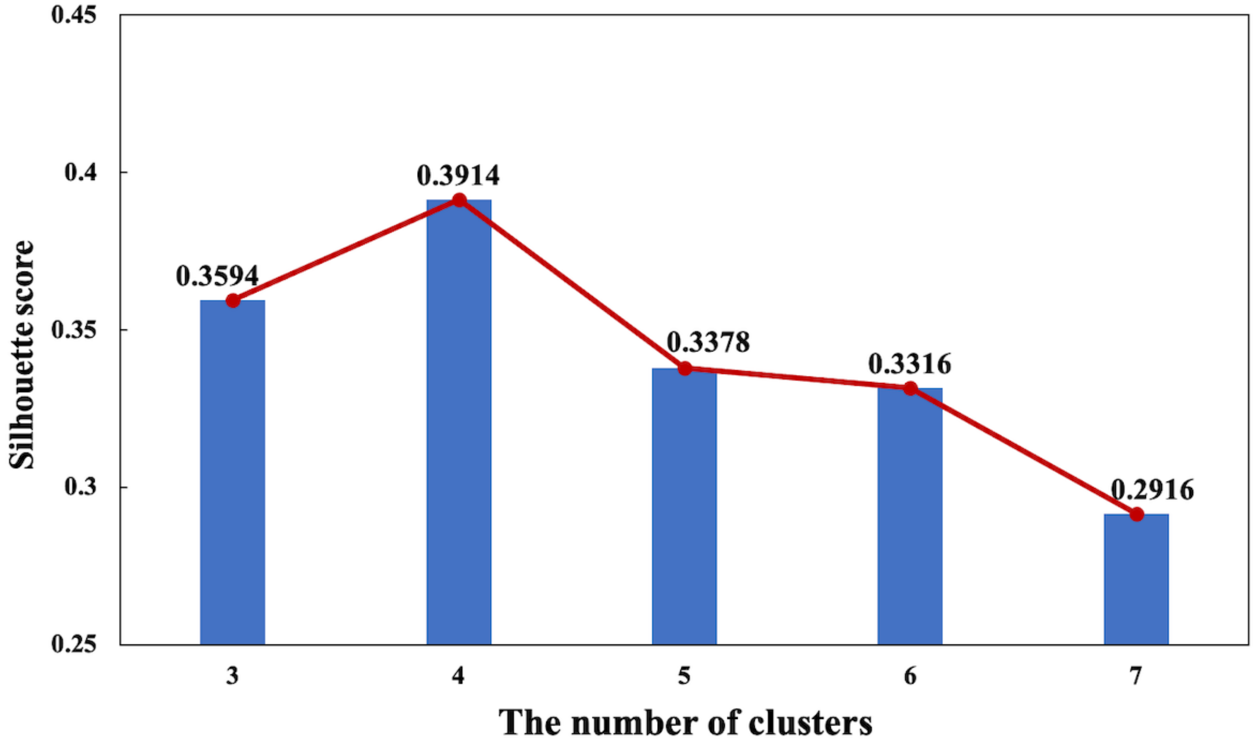


**Table 1.** Comparison of PE in the 4 clusters.

| PE[a] (Days/week) | PELL[b], mean (SD) (n=183) | PEHH[c], mean (SD) (n=142) | PEHL[d], mean (SD) (n=148) | PELH[e], mean (SD) (n=89) | *P* value |
|---|---|---|---|---|---|
| Week 1 | 1.9 (1.1) | 6.4 (1.1) | 5.8 (1.2) | 2.4 (1.3) | <.001 |
| Week 2 | 0.8 (1.3) | 6.6 (0.8) | 4.8 (1.9) | 3.5 (2.0) | <.001 |
| Week 3 | 0.8 (1.3) | 6.4 (0.9) | 3.2 (2.1) | 4.3 (2.0) | <.001 |
| Week 4 | 0.6 (0.8) | 6.3 (0.9) | 1.8 (1.3) | 4.6 (1.5) | <.001 |

[a] PE: patient engagement.

[b] PELL: patient engagement started low and dropped even lower.

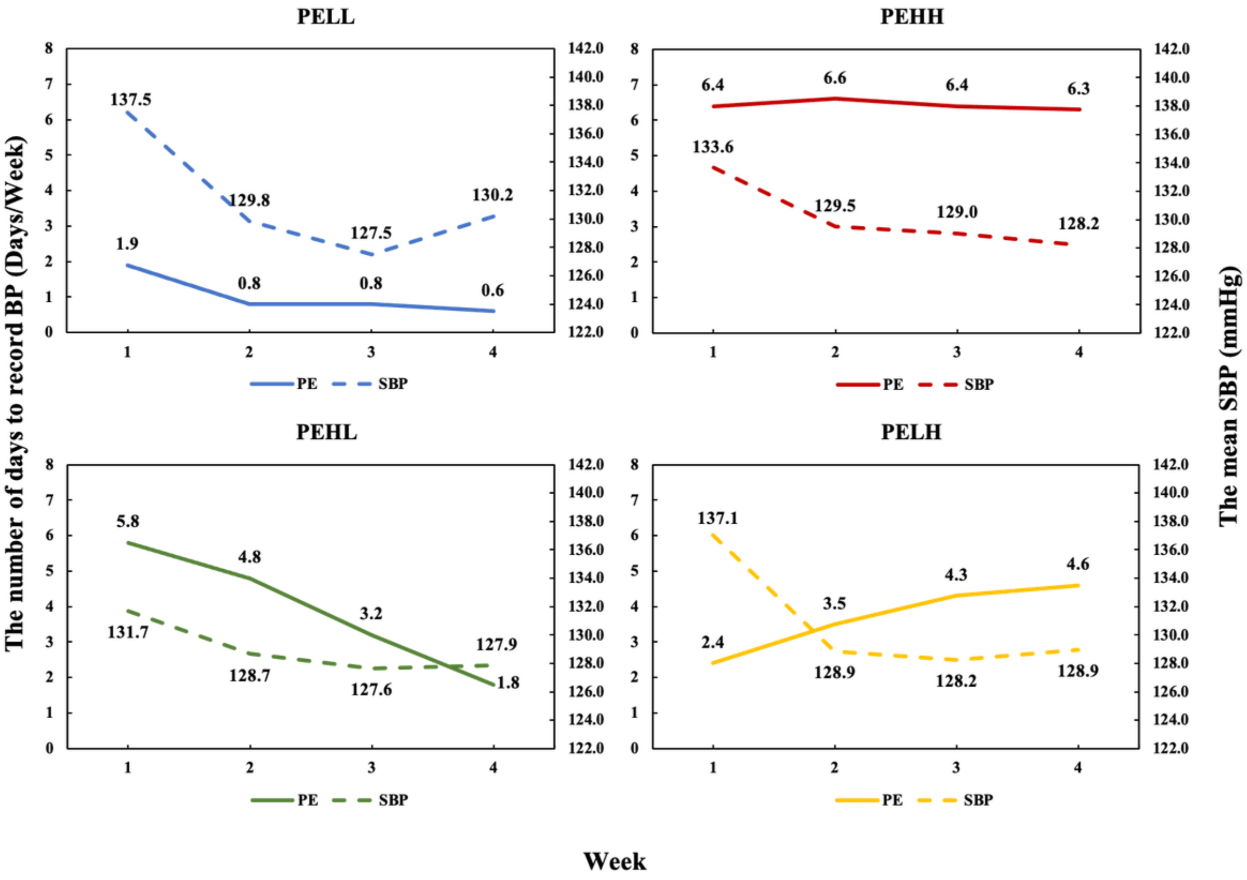
[c] PEHH: patient engagement started high and remained high.

[d] PEHL: patient engagement started high and dropped to low.

[e] PELH: patient engagement started low and rose to high.

There were 4 distinctive change patterns of PE (see Figure 3). The first cluster contained 183 patients. They started recording their BP 1.9 days/week (SD 1.1 days/week) and then decreased every week to around 0.6 days/week (SD 0.8 d days/week) in the fourth week. We referred this cluster as “PE started low and dropped even lower” (PELL) cluster. The second cluster contained 142 patients who were consistently active engaging in BP recording throughout the whole period. On average they recorded the BP for more than 6 days per week. Therefore, we referred them as the “started high and remained high” (PEHH) cluster. The third cluster contained 148 patients who started with a high level of recording (5.8 days/week, SD 1.2 days/week) that then decreased every week to around 1.8 days/week (SD 1.3 days/week) in the fourth week. We referred them as the “PE started high and dropped to low” (PEHL) cluster. The fourth cluster contained 89 patients who began recording self-measured BP 2.4 days/week (SD 1.3 days/week) and gradually increased every week to around 4.6 days/week (SD 1.5 days/week) in the fourth week. We referred them as the “PE started low and then rose to high” (PELH) cluster.

**Figure 3.** The patient engagement patterns of the 4 clusters and the 4-week mean SBP curve of the 4 clusters. BP: blood pressure; SBP: systolic blood pressure. PELL: patient engagement started low and dropped even lower; PEHH: patient engagement started high and remained high; PEHL: patient engagement started high and dropped to low; PELH: patient engagement started low and rose to high.



## The Demographics, SBP, and Follow-up Characteristics of the PE Patterns

### *The Association Between Demographic Factors and the PE Pattern*

The distribution of age, gender, and education level in each cluster is presented in Table 2. The majority of patients (387/562, 68.9%) were between 40 and 60 years old. Chi-square test found a statistically significant association between age and the PE pattern (*P*=.001). There was no significant difference in the distribution of patients under 60 years old in each PE pattern, but there were many more patients aged over 60 years old in the PEHH cluster than in the other clusters (*P*<.05).

Chi-square test did not find a statistically significant association between the PE pattern and gender or education level. There were many more male patients (338/562, 60.1%) using the BPA than female patients. Moreover, 50.5% (284/562) of patients had a university degree.

**Table 2.** The distribution of demographic characteristics in each cluster.

| Characteristic | PELL[a] (n=183) | PEHH[b] (n=142) | PEHL[c] (n=148) | PELH[d] (n=89) | *P* value |
|---|---|---|---|---|---|
| **Age, n (%)** | | | | | .001 |
| 18<age≤40 | 24 (13.1) | 14 (9.8) | 27 (18.2) | 11 (12.4) | |
| 40<age≤60 | 131 (71.6) | 86 (60.6) | 104 (70.3) | 66 (74.1) | |
| Age>60 | 28 (15.3) | 42 (29.6) | 17 (11.5) | 12 (13.5) | |
| **Gender, n (%)** | | | | | .37 |
| Male | 102 (55.7) | 84 (59.2) | 95 (64.2) | 57 (64.0) | |
| Female | 81 (44.3) | 58 (40.8) | 53 (35.8) | 32 (36.0) | |
| **Education, n (%)** | | | | | .45 |
| Primary or secondary school | 14 (7.6) | 6 (4.2) | 5 (3.4) | 2 (2.2) | |
| High school | 79 (43.2) | 61 (43.0) | 68 (45.9) | 43 (48.3) | |
| University and above | 90 (49.2) | 75 (52.8) | 75 (50.7) | 44 (49.4) | |

[a] PELL: patient engagement started low and dropped even lower.

[b] PEHH: patient engagement started high and remained high.

[c] PEHL: patient engagement started high and dropped to low.

[d] PELH: patient engagement started low and rose to high.

## *The Association Between the Initial BP Conditions and the PE Pattern*

The mean SBP in the first week represented the initial BP conditions in hypertension management. The chi-square test showed a statistically significant association between the mean SBP in the first week and the PE pattern (*P*<.001; see Table 3). *z* scores identified a significantly higher proportion of patients with high SBP (between 160 mmHg and 179 mmHg) in the PELL cluster in the first week than in the PEHL cluster (*P*<.05). The majority of patients (316/562, 56.2%) had a mean SBP between 120 mmHg and 139 mmHg in the first week.

**Table 3.** The distribution of the first week mean SBP in each cluster (*P*<.001).

| SBP[a] | PELL[b], n (%) (n=183) | PEHH[c], n (%) (n=142) | PEHL[d], n (%) (n=148) | PELH[e], n (%) (n=89) |
|---|---|---|---|---|
| SBP<120 | 19 (10.4) | 13 (9.2) | 23 (15.5) | 7 (7.9) |
| 120≤SBP≤139 | 90 (49.2) | 91 (64.1) | 91 (61.5) | 44 (49.4) |
| 140≤SBP≤159 | 53 (29.0) | 35 (24.6) | 32 (21.6) | 34 (38.2) |
| 160≤SBP≤179 | 16 (8.7) | 3 (2.1) | 2 (1.4) | 4 (4.5) |
| SBP≥180 | 5 (2.7) | 0 (0) | 0 (0) | 0 (0) |

[a] SBP: systolic blood pressure.

[b] PELL: patient engagement started low and dropped even lower.

[c] PEHH: patient engagement started high and remained high.

[d] PEHL: patient engagement started high and dropped to low.

[e] PELH: patient engagement started low and rose to high.

## *The Trend of Weekly Mean SBP in Each PE Pattern*

Across 4 weeks of management, the mean SBP of all 4 clusters declined at different rates (see Figure 3). Pairwise comparisons of the mean SBP between 4 weeks in each cluster revealed that in the clusters of PELL (*P*=.002), PEHH (*P*=.006), and PELH (*P*<.001), the mean SBP had significantly reduced from the second week. In the PEHL cluster, the significant reduction of SBP appeared from the third week (*P*=.02). The PELL cluster experienced a greater reduction in the mean SBP in the third week compared to the other clusters, but the SBP of users in this cluster increased more than did the others in the fourth week.

## *The Number of Follow-ups in Each PE Pattern*

The amount of the 3 types of follow-ups were significantly different (*P*=.03) in the 4 clusters (see Table 4). Abnormal follow-up was performed significantly less frequently (*P*<.05) in the PELL cluster than in the clusters of PEHH and PEHL. The distribution of follow-up in each of the 4 weeks was not significantly different in the 4 clusters. Most (338/680, 49.7%) follow-ups were performed in the first week.

**Table 4.** The distribution of the number of 3 types of follow-ups and the number of follow-ups in each of the 4 weeks in each cluster.

| Value | PELL[a] (n=183) | PEHH[b] (n=142) | PEHL[c] (n=148) | PELH[d] (n=89) | *P* value |
|---|---|---|---|---|---|
| **Type, n (%)** | | | | | .02 |
| Compliance | 77 (46.1) | 53 (31.4) | 74 (39.2) | 57 (41.9) | |
| Regular | 80 (47.9) | 87 (51.5) | 85 (45.0) | 61 (44.9) | |
| Abnormal | 10 (6.0) | 29 (17.2) | 30 (15.9) | 18 (13.2) | |
| **Time, n (%)** | | | | | .11 |
| Week 1 | 90 (51.7) | 86 (50.3) | 96 (49.2) | 66 (47.1) | |
| Week 2 | 50 (28.7) | 44 (25.7) | 40 (20.5) | 36 (25.7) | |
| Week 3 | 13 (7.5) | 14 (8.2) | 23 (11.8) | 23 (16.4) | |
| Week 4 | 21 (12.1) | 27 (15.8) | 36 (18.5) | 15 (10.7) | |

[a] PELL: patient engagement started low and dropped even lower.

[b] PEHH: patient engagement started high and remained high.

[c] PEHL: patient engagement started high and dropped to low.

[d] PELH: patient engagement started low and rose to high.

## The Effect of Follow-up on PE and SBP in Each PE Pattern

### *The Effect of Each Type of Follow-up on PE*

The results of ANOVA revealed the significant main effects of cluster (*F*=15.09; *P*<.001) and type (*F*=5.442, *P*=.005), and the cluster–type interaction effect (*F*=2.60; *P*=0.02) on PE (see Table 5). Bonferroni multiple comparisons found that after the 3 types of follow-up, the CPE between the 4 clusters had the following statistically significant differences: (1) for the compliance follow-up, PE in the PELL cluster had a significantly larger drop than did the PEHH (*P*=.02) and PELH (*P*<.001) clusters, and improvement in PE in the PELH cluster was significantly higher than that of the PELL (*P*<.001) and PEHL (*P*<.001) clusters; (2) for the regular follow-up, PE in the PEHL cluster had a significantly larger drop than did the clusters of PELL (*P*=.01), PEHH (*P*=.01), and PELH (*P*<.001); and (3) for the abnormal follow-up, the decrease in PE in the PEHL cluster was significantly lower than that of the PEHH cluster (*P*=.01). Compliance follow-up improved PE in the PEHH and PELH clusters. Regular follow-up improved PE in the PELH cluster.

**Table 5.** The effect of 3 types of follow-ups on patient engagement and the effect of follow-up in each of the 4 weeks on PE. Multiple comparisons of the 4 clusters (at the .05 level).

| CPE[a] (Days/week) | PELL[b], mean (SD) (n=183) | PEHH[c], mean (SD) (n=142) | PEHL[d], mean (SD) (n=148) | PELH[e], mean (SD) (n=89) | *P* value |
|---|---|---|---|---|---|
| **Type** | | | | | |
| Compliance | –0.6 (1.7) | 0.5 (1.5)[f] | –0.3 (2.5) | 1.31 (2.3)[f,h] | <.001 |
| Regular | –0.1 (1.2) | –0.2 (1.4) | –1.1 (2.3)[f,g] | 0.67 (2.7)[h] | <.001 |
| Abnormal | –0.1 (1.0) | –0.1 (1.4) | –1.8 (2.3)[g] | –0.39 (2.3) | .01 |
| **Time** | | | | | |
| Week 1 | –0.7 (1.6) | 0.3 (1.7)[f] | –0.8 (2.3)[g] | 1.2 (2.5)[f,g,h] | <.001 |
| Week 2 | 0.1 (1.2) | 0 (0.9) | –1.5 (2.7)[f,g] | 1.4 (2.1)[f,g,h] | <.001 |
| Week 3 | –0.2 (1.3) | 0.4 (1.0) | –1.6 (2.1)[g] | 0.3 (2.5)[h] | .01 |
| Week 4 | –0.2 (1.4) | –0.7 (1.7) | –0.1 (2.7) | –1.5 (2.5) | .10 |

[a] CPE: change in patient engagement.

[b] PELL: patient engagement started low level and dropped even lower.

[c] PEHH: patient engagement started high and remained high.

[d] PEHL: patient engagement started high and dropped to low.

[e] PELH: patient engagement started low and rose to high.

[f] The given cluster is significantly different from the PELL cluster.

[g] The given cluster is significantly different from the PEHH cluster.

[h] The given cluster is significantly different from the PEHL cluster.

### *The Effect of Follow-up on PE in Each of the 4 Weeks*

The results of ANOVA revealed the significant main effects of the cluster (*F*=10.335; *P*<.001) and time (*F*=2.870; *P*=.04), and the cluster–time interaction effect (*F*=5.168; *P*<.001) on PE (see Table 5). Bonferroni multiple comparisons found that after the follow-up in each of the 4 weeks, the CPE between the 4 clusters had the following statistically significant differences: in week 1, the improvement in PE in the PELH cluster was significantly higher than that of the clusters of PELL (*P*<.001), PEHH (*P*=.03), and PEHL (*P*<.001); in week 2, the PE in the PEHL cluster had a significantly larger drop than that of the clusters of PELL (*P*=.001), PEHH (*P*=.01), and PELH (*P*<.001), and the improvement in PE in the PELH was significantly higher than that of the clusters of PELL (*P*=.01), PEHH (*P*=.01), and PEHL (*P*<.001); and in week 3, the PE in the PEHL cluster had a significantly larger drop than that of the PEHH (*P*=.02) and PELH (*P*=.01) clusters. Follow-up improved PE in the PEHH and PELH clusters in first 3 weeks and only in the second week in the PELL cluster, but had no effect on the PEHL cluster in any of the 4 weeks.

### *The Effect of Each Type of Follow-up on SBP*

The results of ANOVA revealed significant main effects of cluster (*F*=2.789; *P*=.04) and type (*F*=1.137; *P*=.32), but no cluster–type interaction (*F*=0.956; *P*=.46) effect on SBP (see Table 6). The mean SBP of all 4 clusters of patients reduced after 3 types of follow-ups except for the abnormal follow-up in the PELL cluster. Bonferroni multiple comparisons showed that after the compliance follow-up, the PELH cluster had a significantly higher level of SBP decline than did the PEHL cluster (*P*=.003).

**Table 6.** The effect of 3 types of follow-ups on systolic blood pressure and the effect of follow-up in each of the 4 weeks on systolic blood pressure. Multiple comparisons of the 4 clusters (at the .05 level).

| CSBP[a] (mmHg) | PELL[b], mean (SD) (n=183) | PEHH[c], mean (SD) (n=142) | PEHL[d], mean (SD) (n=148) | PELH[e], mean (SD) (n=89) | *P* value |
|---|---|---|---|---|---|
| **Type** | | | | | |
| Compliance | –3.4 (11.4) | –2.5 (5.8) | –0.9 (6.2) | –6.7 (10.5)[f] | .01 |
| Regular | –3.7 (11.3) | –2.1 (6.4) | –2.4 (9.1) | –4.5 (8.7) | .33 |
| Abnormal | 1.8 (8.5) | –2.6 (6.3) | –1.7 (7.2) | –3.1 (11.0) | .66 |
| **Time** | | | | | |
| Week 1 | –6.1 (12.7) | –3.0 (6.5) | –2.5 (7.1) | –8.4 (9.8)[f,g] | <.001 |
| Week 2 | –2.1 (7.7) | –1.3 (5.2) | –0.1 (5.5) | –4.2 (9.0) | .09 |
| Week 3 | 1.4 (13.5) | –3.2 (6.3) | 2.2 (8.9) | –0.3 (10.3) | .04 |
| Week 4 | 2.6 (6.2) | –1.7 (6.3) | –3.7 (10.7) | –3.4 (4.8) | .28 |

[a] CSBP: change in systolic blood pressure

[b] PELL: patient engagement started low and dropped even lower.

[c] PEHH: patient engagement started high and remained high.

[d] PEHL: patient engagement started high and dropped to low.

[e] PELH: patient engagement started low and rose to high.

[f] The given cluster is significantly different from the PEHL cluster.

[g] The given cluster is significantly different from the PEHH cluster.

### *The Effect of Follow-up on SBP in Each of the 4 Weeks*

The results of ANOVA revealed significant main effects of cluster ($F$=2.697; $P$=.045) and time ($F$=7.561; $P$<.001), but no cluster–time interaction effect ($F$=1.600; $P$=.11) on SBP (see Table 6). Bonferroni multiple comparisons showed that the PELH cluster in the first week had a significantly higher level of SBP decline than did the PEHH ($P$=.001) and PEHL ($P$<.001) clusters. The SBP decreased within the first 2 weeks in all 4 clusters and continued to fall in the PEHH and PELH clusters over 4 weeks. Overall, the SBP was more reduced in the PELL (–6.1 mmHg) and PELH (–8.4 mmHg) clusters than in the others in the first week.

# Discussion

## Principal Findings

### *PE Patterns*

This study explored patterns of PE with the hypertension self-management app and identified the effect of health care provider follow-up on PE and SBP in each PE pattern in the first 4 weeks after registration. For the first time, we found 4 dynamic trends of PE in a sample of 562 patients who used the mHealth app to record self-measured BP. Two clusters started with a high level of engagement, with one remaining at a high level throughout and the other dropping. Two clusters started with low engagement, with one increasing the level of engagement and the other dropping.

The majority of patients (387/562, 68.9%) were mainly between 40 and 60 years old. There was no difference in the distribution of patients under 60 years of age in each PE cluster; however, there were more patients aged over 60 years in the continuously high engagement cluster than in the other clusters. This may suggest that the patients over 60 years of age were more likely to engage in recording self-measured BP using the mHealth app for hypertension self-management. This may be attributed to the high level of awareness of the risk of hypertension because age is an important contribution factor to the development of hypertension [16,43]. Our finding is consistent with that of Goyal et al [15] in which older participants completed more planning challenges for chronic disease management than did younger participants. However, the finding of Kruse et al [10] was not consistent with our observations. They found that patients over the age of 65 years were less likely to use the mHealth service due to problems understanding the information, difficulty using technology, and inability to access the internet [10].

In this study, there were more male patients than female patients, which may be attributed to males being more prone to using mHealth services for hypertension management [44,45]. However, we did not find a significant difference in the proportion of males and females in each PE pattern. It appears that, although more males used the mHealth services than did females [44], they had the same level of PE with the mHealth service. Abd-alrazaq et al [30] found that gender did not affect intention to use the mHealth service, which was in line with our findings. However, Chung et al [46] found that females with heart failure were more adherent to the sodium-restricted diet than males, and Goyal et al [15] found that PE levels among female users with chronic disease were slightly higher than those among male users. Therefore, the gender differences in PE with mHealth apps does not appear to be conclusive.

The relationship between PE and SBP in the PELL cluster is worth noting. Patients in the PELL cluster reported a higher

SBP in the first week, and after 3 weeks of management, the SBP decreased more compared with the other clusters. Patients then began to stop using the app, and this decline in PE continued simultaneously with the increase in SBP. It may be the case that patients felt they reached the BP goal and ceased use of the app [27,47]. This may give evidence to suggest that controlling BP requires continuous engagement in hypertension self-management [35].

### *The Effect of 3 Types of Follow-ups on PE and SBP*

Health care provider follow-up is essential for the prevention and treatment of hypertension. It is necessary to understand which type of follow-up is effective for which patient group [2]. Compliance follow-up, which was provided when the patients showed signs of reducing compliance in recording self-measured BP, had a positive effect on PE in the PEHH and PELH clusters. Regular follow-up, which was provided at a fixed interval, only had a positive effect on PE in the PELH cluster. These follow-ups might have been taken as the cue for engagement behavior for these patients, who may have awareness of hypertension self-management, but need cues for action. Interestingly, abnormal follow-up had no effect on any cluster of patients despite it being performed when the patient's BP was abnormal. These findings suggest that different types of follow-ups had different effects on patients' behavior of recording self-measured BP.

The content and method of follow-up are important to achieving a follow-up effect [48-50]. One study showed that health care providers have individual inherent preferences for the type of follow-up content to be provided to the patient [18]. Abd-alrazaq et al [30] found that the health care provider factor was related to patients' intension to engage in mHealth services. In our study, follow-up was performed through the mobile phone after health care providers received a reminder from the mHealth app. Other ways to improve PE, such as routine reminders combined with games, have yielded higher user engagement [16], and users seem to prefer simple, short voice messages over text messages because of communication trust and increased accessibility [51]. Cechetti et al [52] developed and implemented an mHealth app with a gamification method for hypertension management, which proved to be effective in promoting PE [52].

The mean SBP of all 4 clusters of patients reduced after 3 types of follow-ups except for abnormal follow-up in the PELL cluster. We also found abnormal follow-up was performed less frequently in the PELL cluster than in the others. This may be attributed to patients in the PELL cluster having a low level of engagement in recording self-measured BP, which resulted in the mHealth app not detecting the patient's abnormal BP condition and reminding the health care provider to follow-up. The PELH cluster developed positive PE and achieved SBP control after 3 types of follow-ups. This is in agreement with the observation that follow-up can motivate patients to engage in self-management. Positive feedback from PE after follow-up has been shown to be beneficial for BP control [53-55].

### *The Effect of Follow-up in Each of the 4 Weeks on PE and SBP*

One study reported that 74.84% of app-only users stopped using an mHealth physical activity management app by day 43 [56]. This suggests that the duration of patient "stickiness" with the fully automatic mHealth services is limited and needs to be complemented by human support to keep momentum. Little is still known concerning the extent and duration of the effect of health care provider follow-up on self-management of hypertensive patients. This study found that the effect of follow-up on the PEHH and PELH clusters lasted until the third week, only had a small impact on the PELL cluster in the second week, and had no effect on the PEHL cluster at any time point. The effect duration of different types of follow-ups varied across each PE pattern. This may suggest that a fixed care pathway based only on patient's BP level would not work for all types of patients. Chronic disease management also needs to consider the patient's behavior and personal preferences. Our results support the design of patient-centered follow-up plans that incorporate social behavioral characteristics and preferences of patients into chronic disease management [14].

After 4 weeks of mHealth services being used, the mean SBP of all 4 clusters decreased. Interestingly, in the first week, the PELL and PELH clusters had a higher level of decline in SBP than did the other clusters. This may be the reason why the PELL cluster was complacent and did not actively engage in recording self-measured BP. Taking the advice of the health care providers may motivate those in the PELH cluster to increase their level of engagement. In the first 2 weeks, all 4 clusters of patients experienced a decline in SBP. Only the PEHH and PELH clusters maintained the SBP reduction through the fourth week, supporting the notion that hypertension management requires ongoing effort in monitoring BP to help patients improve their awareness of their own condition.

## Strengths and Limitations

The strength of this study was that, first, the study used the longitudinal data collected in the first 4 weeks of patient registration with the mHealth service, which is informative for characterizing the changing trend of PE with hypertension management under an mHealth service. The findings are useful for continuous improvement of mHealth services for hypertension management. To the best of our knowledge, no other study has used longitudinal data to describe patterns of PE with a hypertension self-management app. Second, we analyzed the effect of health care provider follow-up on PE and SBP in each PE pattern. Our findings revealed the relationship between PE, BP, and health care provider follow-up. This provides evidence to support the further design of appropriate types of follow-ups for patients. Finally, we observed PE from real-world patients, which can reveal patient behavior in a natural setting, rather than recruited patients who would be more likely to overcome the burden associated with research work [57]. This helped to generate implementable, practical insights for the engagement of actual patients in daily hypertension management.

The study had 4 limitations. First, we excluded patients who used the app for fewer than 4 weeks after registration. This might have produced a bias toward more positive findings of PE with the mHealth app and thus may limit the generalizability of our findings to those patients who drop out of the mHealth service early. Second, we only investigated the PE trend with the mHealth service in the first 4 weeks of hypertension management. Future study can further investigate the PE pattern after this period of time. Third, we defined PE as the behavior of recording self-measured BP. There are many other usage behaviors, such as taking medicine, engaging in physical activities, and keeping a healthy diet, which should be analyzed in future studies. Fourth, the different engagement behaviors could have arisen from various demographic and social psychological characteristics of patients (such as marital status, profession, anxiety, depression), so these factors need to be considered in future studies.

### Conclusions

By analyzing the 4-week log data from a hypertension self-management app, BPA, this study identified the 4 distinct PE patterns in using an mHealth app for hypertension self-management. We also characterized the different effects of 3 types of health care provider follow-up on PE and SBP. Results showed how patients engaged with the mobile app and how health care provider follow-up affects or does not affect their engagement and BP. Our findings may inform the design and help strengthen health care provider follow-up strategies to improve outpatient engagement with mHealth apps for hypertension management. Future work needs to clarify the long-term engagement of patients with hypertension health care services. The indicators of PE should be broadened to include multiple types of usage behavior, and the effect of patient provider follow-up needs to be associated with patients' various demographic and socio-psychological characteristics.

## Acknowledgments

This study was supported by the National Key Research and Development Program of China (#2020YFC2006400 and #2020YFC2003403), and the Major Science and Technology Project in Hainan Province of China (# ZDKJ2019012).

## Conflicts of Interest

None declared.

## References

1. Global status report on noncommunicable diseases 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf;jsessionid=3C0C9A267DA218A8836BDCA9EB2C5C0D?sequence=1 [accessed 2021-09-05]
2. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
3. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
4. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
5. Wong A, Bhyat R, Srivastava S, Boissé Lomax L, Appireddy R. Patient Care During the COVID-19 Pandemic: Use of Virtual Care. J Med Internet Res 2021 Jan 21;23(1):e20621 [FREE Full text] [doi: 10.2196/20621] [Medline: 33326410]
6. Chen X, Koskela M, Hyväkkä J. mHealth new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2021-09-05]
7. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
8. 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]
9. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
10. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
11. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]

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 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
13. Sakakibara BM, Ross E, Arthur G, Brown-Ganzert L, Petrin S, Sedlak T, et al. Using mobile-health to connect women with cardiovascular disease and improve self-management. Telemed J E Health 2017 Mar;23(3):233-239. [doi: 10.1089/tmj.2016.0133] [Medline: 27623231]
14. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Ann Intern Med 2003 Feb 04;138(3):256-261. [Medline: 12558376]
15. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
16. 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]
17. 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]
18. Tyagi H, Sabharwal M, Dixit N, Pal A, Deo S. Leveraging providers' preferences to customize instructional content in information and communications technology-based training interventions: retrospective analysis of a mobile phone-based intervention in India. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e15998 [FREE Full text] [doi: 10.2196/15998] [Medline: 32130191]
19. James J. Patient Engagement. Health Affairs Health Policy Brief 2013;14. [doi: 10.1377/hpb20130214.898775]
20. 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]
21. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
22. Rehman H, Kamal AK, Morris PB, Sayani S, Merchant AT, Virani SS. Mobile health (mHealth) technology for the management of hypertension and hyperlipidemia: slow start but loads of potential. Curr Atheroscler Rep 2017 Mar;19(3):12. [doi: 10.1007/s11883-017-0649-y] [Medline: 28210974]
23. Cheikh-Moussa K, Mira JJ, Orozco-Beltran D. Improving engagement among patients with chronic cardiometabolic conditions using mHealth: critical review of reviews. JMIR Mhealth Uhealth 2020 Apr 08;8(4):e15446 [FREE Full text] [doi: 10.2196/15446] [Medline: 32267239]
24. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. Health Inf Sci Syst 2017 Dec;5(1):4 [FREE Full text] [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]
25. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
26. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, manage my pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
27. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
28. 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]
29. Mosby. Follow-up. (n.d.). Medical Dictionary for the Health Professions and Nursing Internet. URL: https://medical-dictionary.thefreedictionary.com/follow-up [accessed 2021-09-05]
30. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
31. 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]
32. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
33. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison HC, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation,

and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2017 Nov 13:e13-e115. [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]

34. National Center for Cardiovascular Diseases of China. China Hypertension Community Management Guidelines 2014. Chinese Journal of Health Management 2015 Sep 1;9(1):10-30. [doi: 10.3760/cma.j.issn.1674-0815.2015.01.004]
35. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension-A report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. J Geriatr Cardiol 2019 Mar;16(3):182-241 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2019.03.014] [Medline: 31080465]
36. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
37. Wang Z, Li C, Huang W, Chen Y, Li Y, Huang L, et al. Effectiveness of a pathway-driven eHealth-based integrated care model (PEICM) for community-based hypertension management in China: study protocol for a randomized controlled trial. Trials 2021 Jan 22;22(1):81 [FREE Full text] [doi: 10.1186/s13063-021-05020-2] [Medline: 33482896]
38. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
39. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
40. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]
41. Huy VT, Anh DT. An efficient implementation of anytime K-medoids clustering for time series under dynamic time warping. New York, NY, USA: Association for Computing Machinery; 2016 Presented at: Proceedings of the Seventh Symposium on Information and Communication Technology (SoICT '16); 2016; Ho Chi Minh, Vietnam p. 22-29. [doi: 10.1145/3011077.3011128]
42. Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. 2010 Presented at: 2010 International Conference on Data Mining; Dec 13, 2010; Sydney, NSW, Australia p. 911-916.
43. Rao J, Chen X, Zhang J, Yang J, Pang M, Pan B, et al. The risk of hypertension doubles every 10 years in China: Age, period, and birth cohort effects on the prevalence of hypertension from 2004 to 2013. Am J Hypertens 2019 Apr 22;32(5):492-502. [doi: 10.1093/ajh/hpz003] [Medline: 30668634]
44. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
45. Khatun F, Heywood AE, Hanifi SMA, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
46. Chung ML, Moser DK, Lennie TA, Worrall-Carter L, Bentley B, Trupp R, et al. Gender differences in adherence to the sodium-restricted diet in patients with heart failure. J Card Fail 2006 Oct;12(8):628-634 [FREE Full text] [doi: 10.1016/j.cardfail.2006.07.007] [Medline: 17045182]
47. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):1-4 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
48. Billue KL, Safford MM, Salanitro AH, Houston TK, Curry W, Kim Y, et al. Medication intensification in diabetes in rural primary care: A cluster-randomised effectiveness trial. BMJ Open 2012 Sep 17;2(5):e000959. [doi: 10.1136/bmjopen-2012-000959]
49. Estrada CA, Safford MM, Salanitro AH, Houston TK, Curry W, Williams JH, et al. A web-based diabetes intervention for physician: a cluster-randomized effectiveness trial. Int J Qual Health Care 2011 Dec;23(6):682-689 [FREE Full text] [doi: 10.1093/intqhc/mzr053] [Medline: 21831967]
50. Crenshaw K, Curry W, Salanitro AH, Safford MM, Houston TK, Allison JJ, et al. Is physician engagement with web-based CME associated with patients’ baseline hemoglobin A1c levels? The rural diabetes online care stud. Academic Medicine 2010;85(9):1511-1517. [doi: 10.1097/acm.0b013e3181eac036]
51. Steinman L, Heang H, van Pelt M, Ide N, Cui H, Rao M, et al. Facilitators and barriers to chronic disease self-management and mobile health interventions for people living with diabetes and hypertension in Cambodia: qualitative study. JMIR Mhealth Uhealth 2020 Apr 24;8(4):e13536 [FREE Full text] [doi: 10.2196/13536] [Medline: 32329737]
52. Cechetti NP, Bellei EA, Biduski D, Rodriguez JPM, Roman MK, De Marchi ACB. Developing and implementing a gamification method to improve user engagement: A case study with an m-Health application for hypertension monitoring. Telematics and Informatics 2019 Aug;41:126-138. [doi: 10.1016/j.tele.2019.04.007]
53. Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]

54. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
55. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
56. 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;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
57. Sanders I, Short CE, Bogomolova S, Stanford T, Plotnikoff R, Vandelanotte C, et al. Characteristics of adopters of an online social networking physical activity mobile phone app: cluster analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e12484 [FREE Full text] [doi: 10.2196/12484] [Medline: 31162130]

## Abbreviations

**ANOVA:** analysis of variance
**BP:** blood pressure
**BPA:** Blood Pressure Assistant
**CPE:** change in patient engagement
**CSBP:** the change in systolic blood pressure
**mHealth:** mobile health
**PE:** patient engagement
**PELL:** patient engagement started low and dropped even lower
**PEHH:** patient engagement started high and remained high
**PEHL:** patient engagement started high and dropped to low
**PELH:** patient engagement started low and rose to high
**SBP:** systolic blood pressure

*Edited by R Kukafka; submitted 09.11.20; peer-reviewed by L Liu, J Lacroix; comments to author 05.02.21; revised version received 10.03.21; accepted 10.08.21; published 28.09.21*

*Please cite as:*
*Wu D, An J, Yu P, Lin H, Ma L, Duan H, Deng N*
*Patterns for Patient Engagement with the Hypertension Management and Effects of Electronic Health Care Provider Follow-up on These Patterns: Cluster Analysis*
*J Med Internet Res 2021;23(9):e25630*
*URL: https://www.jmir.org/2021/9/e25630*
*doi: 10.2196/25630*
*PMID:*

Original Paper

# Attitudes Toward the Environment and Use of Information and Communication Technologies to Address Environmental Health Risks in Marginalized Communities: Prospective Cohort Study

Jose G Perez-Ramos[1], PhD; Scott McIntosh[2], PhD; Emily S Barrett[3], PhD; Carmen M Velez Vega[4], PhD; Timothy D Dye[1], PhD

[1]Department of Obstetrics and Gynecology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[2]Department of Public Health Sciences, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[3]Environmental and Occupational Health Sciences Institute, Rutgers School of Public Health, Rutgers University, Piscataway Township, NJ, United States

[4]Escuela Graduada de Salud Pública, Recinto de Ciencias Médicas, Universidad de Puerto Rico, San Juan, Puerto Rico

**Corresponding Author:**
Jose G Perez-Ramos, PhD
Department of Obstetrics and Gynecology
School of Medicine and Dentistry
University of Rochester
601 Elmwood Ave
Box 668
Rochester, NY, 14622
United States
Phone: 1 5852768755
Email: j.perezramos@rochester.edu

## Abstract

**Background:** Information and communication technologies, including mobile health (mHealth), can help isolated communities address environmental health challenges. The Puerto Rican island of Culebra has faced multiple sociopolitical and economic factors that have distressed the island's environment and health. *Culebrenses* are technologically engaged and have demonstrated a use of technology that transcends socioeconomic barriers. As a result, technological interventions could potentially help manage environmental risks on the island.

**Objective:** This study aims to test and evaluate the potential benefits of an mHealth tool, termed ¡mZAP! (*Zonas, Acción y Protección*), for engaging communities with environmental risks through technology.

**Methods:** Participants using ¡mZAP! (N=111) were surveyed. Bivariate analyses were used to examine associations of mHealth use with sociodemographics, technology use, an adapted environmental attitudes inventory, and the multidimensional health locus of control. Logistic regression was used to examine associations between attitudes toward environmental health risks and mHealth use.

**Results:** Higher positive attitudes toward the environment were significantly associated with the use of ¡mZAP! (odds ratio 5.3, 95% CI 1.6-17.0). Environmental attitudes were also associated with the multidimensional health locus of control *powerful others* subscale ($P$=.02), indicating that attitudes toward the environment become more negative as feelings controlled by others increase. Participants felt that the authorities would resolve the challenges (63/111, 56.7%).

**Conclusions:** Perceived lack of control could present barriers to collective actions to address salient environmental health challenges in communities. The ongoing dependency on government-based solutions to community problems is worrisome, especially after the hurricane experiences of 2017 (which may potentially continue to be an issue subsequent to the more recent 2020 earthquakes).

***(J Med Internet Res 2021;23(9):e24671)*** doi: 10.2196/24671

**KEYWORDS**

community engagement; environmental health risk; epidemiology; ICT; mHealth; mobile phone; Puerto Rico

## Introduction

### Background

Information and communication technologies (ICTs), including mobile health (mHealth), can help isolated communities address environmental health challenges [1,2]. The World Health Organization considers environmental risks a public health priority [3]. Social and ecological determinants of health, along with environmental health challenges, often exacerbate pre-existing health disparities [4]. Traditional health paradigms have been changed by mHealth tools to address such health disparities. For example, a study used a mobile app termed *FAITH!* (Fostering African American Improvement in Total Health) to improve cardiovascular health among an African American church community (n=86) in the United States following a multiphase community-based participatory research design [5]. Other examples include using a web-based intervention for smoking cessation among low socioeconomic status populations in the United States (n=1440) [6] and an mHealth tool to reduce health care access gaps among marginalized rural populations in Laos (n=983) and Thailand (n=1158) [7].

The benefits of using ICTs to address environmental risks following crowdsourcing approaches have been successfully implemented in the United States using the mobile phones of users [8,9]. Mobile phones are also being used to engage populations in new and innovative learning modalities [10,11]. Previous studies have shown that mobile tools such as cell phones can help users develop positive attitudes toward the environment [12,13]. Mobile tools facilitate engaging with users by using their devices (eg, smartphones and smartwatches) as part of the solution to existing environmental challenges [8,13].

Although the use of technology has transcended socioeconomics barriers, providing digital access to different levels of society, a significant gap perpetuates digital access disparities among marginalized communities worldwide [14-16].

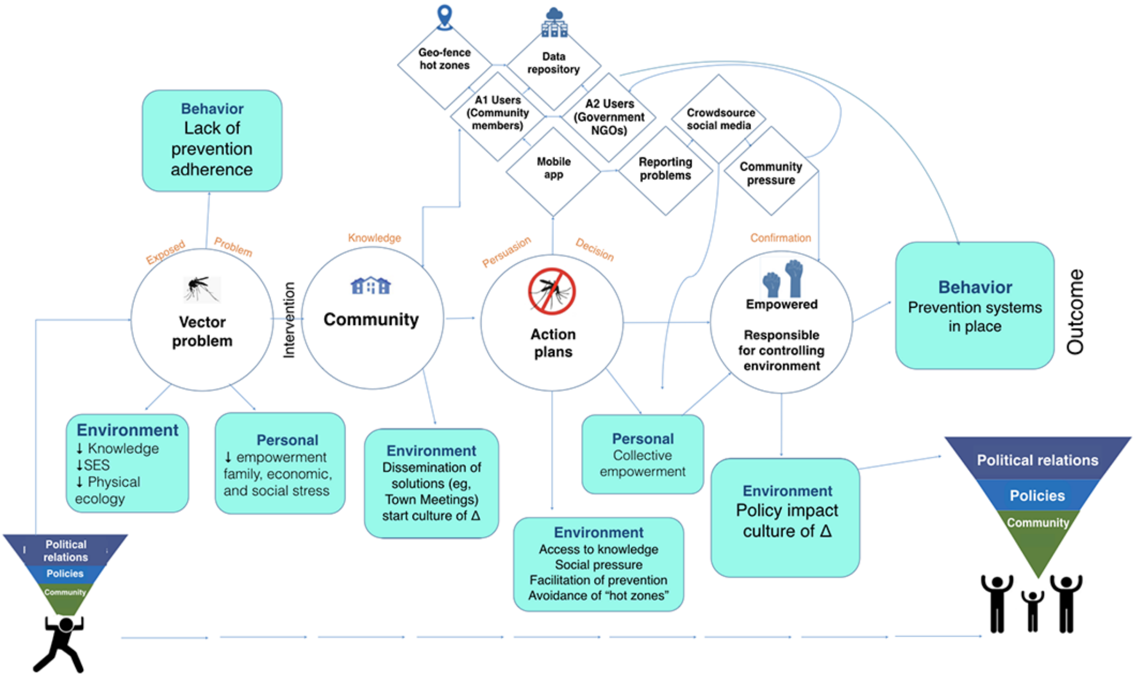
The power of being interconnected via modern technology (eg, social media, cell phones, and the internet) has been shown to (1) catalyze sociopolitical changes [17,18], (2) close the gap in physical distances [19], (3) foster the development of health advances [20], and (4) provide access to information [21]. In contrast, the use of technology has also been linked to the spread of misinformation, resulting in the exacerbation of social injustices among marginalized groups. For example, WhatsApp was used in Maceió, Brazil, to diffuse conspiracy theories about vaccines as the cause of microcephaly in children testing positive for the Zika virus [22,23]. Thus, communities often have untapped and unrecognized resources that may be mobilized to work with the government sector, health programs, and nonprofit organizations to address risks in their environments [24].

Many communities, particularly islands, are dependent on ICT to reduce the impact of their geographical and societal isolation [25]. Island communities face many social and health challenges that are not necessarily faced by other communities, which can intensify risk [26-28]. The remote island of Culebra, off the eastern coast of Puerto Rico’s *Big Island* (*Isla Grande*), is geographically and systemically isolated from the rest of Puerto Rico. Culebra’s isolation also results in limited political power because of the small population, leading to a system of dependency that negatively impacts health access, utilities, waste management, transportation, water safety, and air and soil contamination, which is common in rural and isolated areas [29-33].

Hurricanes Irma and María in 2017 demonstrated this dependency, with Culebra losing connection to the national electrical grid for more than a year [34] and having difficulties in resuming ferry and other island services. Although impacted by these challenges, *Culebrenses* have a strong history of collective engagement against social injustices, such as facilitating the US Navy’s exit from the island in 1975 and protesting against maritime transportation policies [35]. In addition to this community engagement, *Culebrenses* are technologically engaged and have demonstrated a use of technology that transcends socioeconomic barriers [36]. As a result, technological interventions could potentially help manage environmental risks on the island.

This study was guided by a multi-theoretical framework (social cognitive theory [SCT] and the diffusion of innovations theory) which, along with community-centered design methodology, can (1) provide a model for behavioral changes and (2) guide the process to understand the adoption of this mHealth tool better [37-39]. This study also followed a human-centered design framework that provides an inclusion process whereby users are central and active agents necessary for the conception, testing, implementation, and evaluation of solutions to Culebra’s environmental health risk challenges [40]. The SCT addresses behavior change and the effect of the community members’ environment [41]. The SCT also helps identify community problems and priorities, allowing for an in-depth understanding of personal factors, current community behaviors, and environmental influences associated with environmental health risks that can affect community-wide behavioral change. The diffusion of innovations theory facilitates the process of development and sustainability [42]. Owing to the complexity of these two theories, a theoretical framework model titled *Community-Centered Environmental Health Risk Control Model* was developed for this study (Figure 1).

**Figure 1.** Community-centered environmental health risk control model. NGO: nongovernmental organization; SES: socioeconomic status.



## Objective

Following this theoretically informed model facilitated the work as part of a community-driven research project to collaborate with community partners to implement a tailored, community-centered, and crowdsourced ICT termed ¡mZAP! (*Zonas, Acción y Protección*) as part of a community-driven approach to assess the ability of ICTs to engage islanders in collective environmental actions. Specifically, in this study, we evaluate the ability of psychosocial, environmental attitudes, and technological variables to predict the use of the ¡mZAP! app.

# Methods

## Overview

We used a prospective design to assess the ¡mZAP! app use. The study surveyed community participants at baseline to capture their perceptions and attitudes toward the environment, access to and use of technology, community actions toward environmental challenges, knowledge about vector-borne diseases, and health control in Culebra, Puerto Rico. App registrants were followed prospectively for 3 months (November 2018-February 2019) to assess the use of ¡mZAP!

## The Archipelago of Culebra

Culebra is an archipelago municipality located on the east coast of Puerto Rico, 17 miles from *Isla Grande* (*Big Island*—as it is commonly referred to in Puerto Rico by the residents of outer islands; Figure 2). On the basis of the 2018 US Census annual estimates of the resident population, Culebra has an estimated population of 1500 [43]. Culebra's geographical location provides a unique, biodiverse ecosystem of drylands, mangroves, lagoons, white sand beaches, and coral reefs, all in almost 12 square miles [35,44]. Largely owing to this biodiversity and geographical location, tourism is Culebra's primary economy.

Culebra also presents another face; for many decades, this set of islands has experienced marginalization by the United States and from the local government [33,35], including military bombing drills that left unknown contamination and pollution effects [29-31], unreliable maritime transportation [45], environmental hazards [46], high presence of vector agents (mosquitoes) [35], and the lack of a proper health care facility [32,47].

**Figure 2.** Map of Culebra in relation to the rest of Puerto Rico.

## ¡mZAP! App

The ¡mZAP! app was designed and developed as a community-oriented ICT to stimulate collective action to mitigate environmental health risks (eg, trash and mosquitoes). This study design involved direct input from community organizations and community members in Culebra, including women's advocacy organizations, sports organizations, private businesses, health organizations, and schools, primarily through focus group and qualitative engagement processes, as previously described [48].

In this app, users may take actions to identify, report, and address community-based priorities such as abandoned structures, trash, stray animals, stagnant water, mosquito breeding grounds, and pests. Although not directly relevant to environmental issues, a feature to track ferry and ship locations was included to further incentivize app use as the community had identified maritime transportation as a major challenge.

## Survey

A survey was deployed at baseline using the REDCap (Research Electronic Data Capture, Vanderbilt University) platform version 9.0, which also hosted the survey data. As defined in previous studies, the REDCap database is a secured, encrypted, and Health Insurance Portability and Accountability Act–compliant data capture application developed for large-scale research projects [49].

The survey contained adaptations of two previously validated instruments and original items self-administered by participants via the mobile app tool. The survey aimed to better understand users' attitudes toward the environment, community technology use, and community challenges. Questions related to community health, demographic characteristics, the environment, attitudes toward community actions, and technology use were included. Consistent with recommendations for pretesting and pilot-testing of surveys evaluating mHealth interventions [49,50], the survey was pretested with approximately 10 iterations with feedback from 4 bilingual Latinos from the author's institution and volunteers from the target community.

## ¡mZAP! Databases

The ¡mZAP! app was built as a cross-platform mobile and web-based app using the *Ionic* framework by an app developer that worked as part of the ¡mZAP! core team at the University of Rochester. *Ionic* is an open-source toolkit that facilitates the building of mobile apps and web-based applications using a single coding structure (eg, HTML, Cascading Style Sheets, and JavaScript), eliminating the need to create a new coding process for each mobile and web-based platform. An integration of *Angular*, a Google-maintained platform, was used to build the app across mobile and web-based systems (eg, iOS and Android). *Firebase* software was used as a database repository. *Firebase* is a Google-based software with the capacity to secure file uploads and downloads, including images, videos, and other user-generated data under Google encrypted (AES256 and AES128) and cloud storage. The ¡mZAP! mHealth tool back-end process was built to assign each user a unique ID linking all the data generated by users while using the app. This process facilitated a better understanding of ¡mZAP! users' behaviors. An application programming interface was used to access the REDCap survey directly from the app. Unique user identifications and email addresses were used to merge the two databases (¡mZAP! use and the REDCap survey). A weekly data transfer stored participants' app use to secure servers at the lead author's institution.

## Sample Size Calculation

This study was powered to detect an effect size of 3.0 ($\alpha$=.05 and $\beta$=.20) of app use by the environmental attitudes' category. We estimated that 50% of participants with positive attitudes toward the environment would use the app and 25% of participants with negative attitudes would use ¡mZAP! On the basis of these calculations, the sample size required for this study was 116 participants.

## Participants and Study Implementation

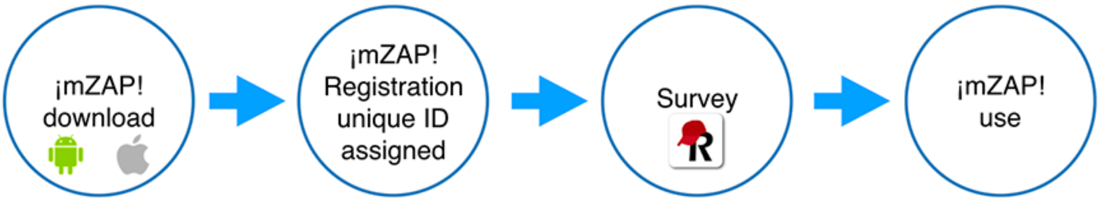
Eligible participants were ≥18 years and self-reported living in Culebra for at least 6 months before the start of the study. Survey questions were administered in Spanish, which is the primary language of the island.

Participants were recruited at baseline following a snowball sampling methodology [51], community meetings, flyer distributions, and social media posts. During in-person visits, eligible participants were invited to download the ¡mZAP! app and complete the survey during the registration process before starting the use of app functionalities (Figure 3). Participants completed the survey using their devices, and via REDCap, the surveys were uploaded to the University of Rochester's firewall-protected, encrypted cloud-based servers. Participants did not receive any compensation for completing the survey or using ¡mZAP! For recruitment purposes, a local area code telephone number and a unique email address were acquired, and a Facebook fan page was created specifically for the project, enabling participants to reach the research team for questions and concerns about the project. This study adhered to the Reporting of Studies Conducted using Observational Routinely Collected Data guidelines and checklist to ensure the inclusion of all components deemed necessary for a scientific research report [52].

¡mZAP! was launched in a series of in-person community meetings in fall of 2018 and was disseminated among community members during a 3-month trial period (November 2018-February 2019). A total of 124 people downloaded the app (Apple iOS 53/124, 42.7% and Android 71/124, 57.3%). Out of the 124 people, 116 (93.5%) participants started registration; 5 (4%) respondents did not meet the inclusion criteria of being a resident of Culebra and were excluded. Out of the 116 participants, 111 (95.7%) participants successfully registered and completed the survey.

This study was reviewed and approved by the institutional review board of the first author's (RSRB00073777) university and followed the 1964 Declaration of Helsinki and its subsequent amendments. HealthproMed, a federally qualified health center in Culebra, reviewed the study protocol and approved the study. An information letter was electronically provided to the participants during their registration for the app.

**Figure 3.** ¡mZAP! user implementation process. ¡mZAP!: Zonas, Acción y Protección.



## Measures

### Overview

The study data included two distinct databases: ongoing baseline survey data and ¡mZAP! use data. The app documented user actions and collected data within a 3-month period. *¡mZAP! use* was defined as any action by users within the mobile app (eg, reports, comments, messages, likes, and resolves) during this follow-up period. Multiple user data points were synchronized by random user unique identification, which traced each user's behavior while using the ¡mZAP! tool.

### The Environmental Attitudes Inventory

The environmental attitudes inventory [53] scale was adapted (adapted environmental attitudes inventory [a-EAI]) to assess participants' attitudes toward the environment. This environmental scale was developed to evaluate the multidimensional nature of environmental attitudes and the perceptions or beliefs of participants regarding the natural environment and associated factors. For this study, 12 items from the environmental attitudes inventory scale were selected and combined for the a-EAI summary scale. We converted the EAI summary scale to a standardized *t* score with mean 50 (SD 10) [54]. In addition to analyzing the a-EAI as a continuous variable, we reduced the a-EAI to discrete categories with tertiles as the cutoff.

### The Wallston Multidimensional Health Locus of Control

The multidimensional health locus of control (MHLC) scale assesses perceptions of control over one's health. The MHLC scale has been used in previous studies to measure the influences and perceptions of community members regarding external factors that affect health outcomes [55]. A Spanish-validated version of the MHLC was used in this study [56]. The three subscales were MHLC internal (internal health locus of control), MHLC external (powerful others health locus of control), and MHLC chance (chance health locus of control) [57]. The rationale for using the MHLC was that perceptions of health control could be associated with overall control and a positive inclination to act against environmental health risks. The subscale scores were standardized using the Streiner and Norman [54] method.

### Community Environmental Priorities, Technology Use, Mosquito Knowledge, and Demographics

For addressing community environmental priorities, two additional sections were incorporated from previous study findings [48,49,58] to assess the use of technology, environmental challenges, community resolutions, and access to technology at home (eg, cell phones and the internet).

### Data Analysis

The survey data were exported from REDCap, cleaned, and analyzed using SPSS Statistics (IBM Corporation) version 25 [59]. Chi-square tests and analysis of variance were used to test bivariate associations among all key variables. Odds ratios (ORs) with 95% CIs were used to examine the magnitude of the association of predictor variables, including the main predictor variable (attitudes toward the environment, a-EAI) and the primary outcome (use of ¡mZAP!). Forced and forward stepwise logistic regression models [60] were generated to control for potential confounders that were marginally associated (ie, $P$<.25) with a-EAI and mZAP! use. Where necessary, $P$<.05 was considered significant.

## Results

### Study Sociodemographic Characteristics

In total, 110 participants completed the download of ¡mZAP! and the intake survey. As Table 1 shows, most (71/110, 64.5%) of the survey participants were aged >35 years. In addition, 65.4% (72/110) had completed a higher education degree and 68.5% (76/111) had lived in Culebra for >11 years. An equal number of females and males participated in the study (n=54 each). A comparison of the study sample with the US Census 2013-2017 American Community Survey in Puerto Rico [43] revealed no statistically significant differences between the study population and the population of the municipality of Culebra regarding gender and ethnicity (data not shown), suggesting that the study sample was similar to the overall population of Culebra.

**Table 1.** Sociodemographics and technology attitudes among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019 (N=110).

| Variables | Participants, n (%) |
|---|---|
| **Age (years)** | |
| 18-25 | 10 (9.1) |
| 26-36 | 29 (26.4) |
| 37-47 | 34 (30.9) |
| 48-54 | 22 (20) |
| ≥55 | 15 (13.6) |
| **Age category (years)** | |
| ≤36 | 39 (35.5) |
| ≥37 | 71 (64.5) |
| **Education level** | |
| Less than or equal to high school | 34 (32.1) |
| Higher education | 72 (67.9) |
| **Gender** | |
| Female | 54 (49.1) |
| Male | 54 (49.1) |
| Prefer not to answer | 2 (1.8) |
| **Time living in Culebra (years; N=111)** | |
| 1-10 | 35 (31.5) |
| More than 11 | 76 (68.5) |
| **Do you have a cell phone data plan?** | |
| Yes | 105 (96.3) |
| No | 4 (3.7) |
| **How is your cell phone reception on the island?** | |
| Good | 11 (10) |
| Regular or bad | 99 (90) |
| **How many cell phones do you have at home?** | |
| 1 | 28 (25.5) |
| >1 | 82 (74.6) |
| **Does your home have internet access?** | |
| Yes | 96 (87.3) |
| No | 14 (12.7) |
| **What type of internet do you have in your home?** | |
| Wi-Fi and satellite | 31 (32.3) |
| Data plan | 63 (65.6) |
| **How many hours per day do you spend using electronic devices?** | |
| <1-5 | 89 (80.9) |
| ≥6 | 21 (19.1) |
| **Do you use social media?** | |
| Yes | 105 (95.5) |
| No | 5 (4.5) |
| **Daily use of social media** | |

| Variables | Participants, n (%) |
|---|---|
| WhatsApp | 99 (94.3) |
| Facebook | 89 (84.8) |
| Instagram | 43 (41.3) |
| **Agreement with the following statements** | |
| Technology is a tool to help us daily | 100 (90.9) |
| The use of technologies helps to fix community problems | 66 (60) |
| The use of technologies brings us closer to the community | 37 (33.9) |
| The constant use of technology limits the community work | 43 (39.1) |
| The technology cause problems in the community | 36 (32.7) |

## Technology and Community Characteristics Among ¡mZAP! Survey Participants

As Table 1 shows, every participant reported owning a cell phone, and 96.3% (105/110) of participants also had a data plan as part of their cell phone plan. With respect to technology at home, 87.3% (96/110) of participants reported having internet access at home, with 66% (63/95) using cell phone data plans for connecting to the internet. Most of the participants owned more than one cell phone (62/83, 74.6%), used electronic devices for up to 5 hours daily (89/110, 80.9%), and had negative perceptions about cell phone reception (99/110, 90%). In addition, 95.5% (105/110) of participants reported being social media users, with WhatsApp being the most used daily social media network (99/105, 94.3%), followed by Facebook (89/105, 84.8%) and Instagram (43/105, 40.1%).

In a series of statements about the relationship between technology and the community, 90.9% (100/110) of participants believed that technology is a tool that can be useful daily and 60% (66/110) agreed that technology could facilitate the resolution of problems in the community. Participants disagreed or were undecided about the role of technology in bringing communities closer (72/110, 65.4%), whether using technologies was an impediment for individuals to do community work (67/110, 60.9%), and whether technology caused community problems (74/110, 67.3%).

## Attitudes Toward the Environment and Predictors

On the basis of the responses on the a-EAI scale, participants often disagreed with items reflecting negative statements about protecting the environment. For example, 44.2% (49/116) disagreed with “One of the most important reasons to keep lakes, beaches and rivers clean is so that people have a place to enjoy water sports,” and 46.8% (51/109) disagreed with “Modern science will not be able to solve our environmental problems.” Most participants agreed with or were ambivalent about items that reflected positive inclinations toward the environment; for example, “Protecting the environment is more important than protecting peoples’ jobs” (65/109, 59.6%) and “I would like to join and actively participate in an environmentalist group” (73/110, 66.4%). When comparing a-EAI with sociodemographic variables, as shown in Table 2, having a higher educational level (OR 4.5, 95% CI 1.2-16.3) and having lived fewer total years in Culebra were significantly associated with a stronger positive attitude toward the environment (OR 3.0, 95% CI 1.2-7.6).

**Table 2.** Associations of the a-EAI[a] with sociodemographics, technology attitudes, and knowledge about mosquito-borne diseases among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019.

| Variables | Total participants, n (%) | a-EAI, n (%) | | Odds ratio (95% CI) | *P* value[b] |
|---|---|---|---|---|---|
| | | High (more positive) | Low (less positive) | | |
| **Age (years; n=108)** | | | | | |
| ≤36 | 37 (34.3) | 11 (29.7) | 26 (70.3) | 1.5 (0.6-3.6) | .41 |
| ≥37 | 71 (65.7) | 16 (22.5) | 55 (77.5) | Referent | N/A[c] |
| **Education level (n=104)** | | | | | |
| Higher education | 71 (68.3) | 22 (31) | 49 (69) | 4.5 (1.2-16.3) | .01 |
| Less than or equal to high school | 33 (31.7) | 3 (9.1) | 30 (90.9) | Referent | N/A |
| **Time living in Culebra (years; n=108)** | | | | | |
| >11 | 73 (67.6) | 13 (17.8) | 60 (82.2) | Referent | N/A |
| 1-10 | 35 (32.4) | 14 (40) | 21 (60) | 3.0 (1.2-7.6) | .01 |
| **How is your cell phone reception on the island? (n=108)** | | | | | |
| Regular or bad | 97 (89.8) | 25 (25.8) | 72 (74.2) | 1.5 (0.3-7.7) | .58 |
| Good | 11 (10.2) | 2 (18.2) | 9 (81.8) | Referent | N/A |
| **How many cell phones do you have at home? (n=108)** | | | | | |
| >1 | 80 (74.1) | 22 (27.5) | 58 (72.5) | 1.7 (0.6-5.2) | .31 |
| 1 | 28 (25.9) | 5 (17.9) | 23 (82.1) | Referent | N/A |
| **Does your home have internet access? (n=108)** | | | | | |
| Yes | 95 (87.9) | 24 (25.3) | 71 (74.7) | 1.1 (0.3-4.4) | .86 |
| No | 13 (12.1) | 3 (23.1) | 10 (76.9) | Referent | N/A |
| **What type of internet do you have in your home? (n=95)** | | | | | |
| Wi-Fi and satellite | 33 (34.7) | 6 (18.2) | 27 (81.8) | Referent | N/A |
| Data plan | 62 (65.3) | 18 (29) | 44 (71) | 1.8 (0.7-5.2) | .24 |

[a] a-EAI: adapted environmental attitudes inventory.

[b] *P* value of chi-square test.

[c] N/A: not applicable for reference groups.

## Collective Action Strategies

A series of questions on the importance of community actions to address environmental risks were posed. As shown in Table 3, most participants (79/108, 73.1%) believed that environmental problems are not typically resolved among community members. Similarly, 58.3% (63/108) of participants expected the government to resolve environmental problems. Other participants expressed that these problems would be resolved by the community (29/108, 26.9%) or would be resolved without making any effort to fix them (28/108, 25.9%). No statistically significant associations were found among the series of items on community actions and a-EAI scale scores.

**Table 3.** Associations of the a-EAI[a] with community environmental actions resolutions among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019 (N=108).

| How does generally environmental differences or problems get resolved in your community? | Total participants, n (%) | a-EAI, n (%) | | Odds ratio (95% CI) | *P* value[b] |
|---|---|---|---|---|---|
| | | Hi (more positive) | No (more negative) | | |
| **Resolved between each other** | | | | | |
| No | 79 (73.1) | 20 (25.3) | 59 (74.7) | 1.0 (0.4-2.9) | .90 |
| Yes | 29 (26.9) | 7 (24.1) | 22 (75.9) | Referent | N/A[c] |
| **Go to government authorities (police and mayor's office)** | | | | | |
| Yes | 63 (58.3) | 13 (20.6) | 50 (79.4) | Referent | N/A |
| No | 45 (41.6) | 14 (31.1) | 31 (68.9) | 1.7 (0.7-4.2) | .22 |
| **Ignore and *pass the page*** | | | | | |
| No | 83 (76.8) | 21 (25.3) | 62 (74.7) | 1.0 (0.4-3.0) | .89 |
| Yes | 25 (23.1) | 6 (24) | 19 (76) | Referent | N/A |
| **Think it will eventually get resolved** | | | | | |
| No | 82 (75.9) | 22 (26.8) | 60 (73.2) | 1.5 (0.5-4.6) | .44 |
| Yes | 26 (24) | 5 (19.2) | 21 (80.8) | Referent | N/A |
| **It will get resolved by itself** | | | | | |
| Yes | 28 (25.9) | 8 (28.6) | 20 (71.4) | 1.2 (0.4-3.3) | .61 |
| No | 80 (74) | 19 (23.8) | 61 (76.3) | Referent | N/A |

[a] a-EAI: adapted environmental attitudes inventory.

[b] *P* value of chi-square test.

[c] N/A: not applicable for reference groups.

## Multidimensional Health Locus of Control

Most participants reported moderate beliefs about health control, MHLC internal (76/118, 64.4%), MHLC *powerful others* (78/108, 72.2%), and MHLC chance (71/108, 65.7%; data not shown). No statistically significant associations were found between the a-EAI and MHLC subscales. In contrast, when comparing MHLC subscales with the a-EAI as a scale score, as MHLC *powerful others* scores increased, overall a-EAI scores decreased (*P*=.02; data not shown). This association indicates that attitudes toward the environment become more negative as the sense that one's health is controlled by *powerful others* (eg, parents, doctors, and authorities) increases.

## ¡mZAP! User's Characteristics and Predictors

Of the 14 participants who used the ¡mZAP! app, 5 (36%) participants used the app more than once; 11 (78%) participants used the app to report environmental health risks, including trash, stray animals, mosquito breeding grounds, abandoned structures, and deposits of stagnant water; and 2 (14%) participants commented on reports. Of the 14 users, 3 (21%) used the tool to report the maritime transportation location (location of the ferries). In addition, of the 14 ¡mZAP! users, 9 (64%) expressed the expectation that the government would resolve the environmental challenges in the community.

Equal proportions of women and men used ¡mZAP! As shown in Table 4, most users (11/14, 79%) were aged >35 years, lived in Culebra for >11 years (10/14, 71%), and had completed higher education (10/14, 71%). No statistically significant differences were found between nonusers and users in terms of sociodemographic variables. Although not significant, several variables had substantial ORs (ORs>1.5): age (OR 2.2, 95% CI 0.5-8.4), educational level (OR 1.6, 95% CI 0.4-6.5), cell phone number per household (OR 2.2, 95% CI 0.5-10.6), household type of internet access (OR 1.8, 95% CI 0.7-5.2), daily hours spent using electronic devices (OR 2.8, 95% CI 0.8-9.3), and *the use of technologies helps to fix community problems* (OR 3.1, 95% CI 0.9-10.0).

**Table 4.** Association of ¡mZAP![a] use with sociodemographics, technology attitudes, and knowledge about mosquito-borne diseases among ¡mZAP! study survey participants in Culebra, Puerto Rico, 2018-2019.

| Variables | Total participants, n (%) | ¡mZAP! use |  | Odds ratio (95% CI) | *P* value[b] |
|---|---|---|---|---|---|
|  |  | Yes | No |  |  |
| **Age category (years; n=110)** |  |  |  |  |  |
| ≤36 | 39 (35.5) | 3 (7.7) | 36 (92.3) | Referent | N/A[c] |
| ≥37 | 71 (64.5) | 11 (15.5) | 60 (84.5) | 2.2 (0.5-8.4) | .24 |
| **Education level (n=106)** |  |  |  |  |  |
| Higher education | 72 (67.9) | 10 (13.9) | 62 (86.1) | 1.6 (0.4-6.5) | .46 |
| Less than equal to high school | 34 (32.1) | 3 (8.8) | 31 (91.2) | Referent | N/A |
| **Time living in Culebra (years; n=111)** |  |  |  |  |  |
| >11 | 76 (68.5) | 10 (13.2) | 66 (86.8) | 1.1 (0.3-4.0) | .79 |
| 1-10 | 35 (31.5) | 4 (11.4) | 31 (88.6) | Referent | N/A |
| **How is your cell phone reception on the island? (n=110)** |  |  |  |  |  |
| Regular or bad | 99 (90) | 14 (14.1) | 85 (85.9) | N/A | N/A |
| Good | 11 (10) | 0 (0) | 11 (100) | N/A | N/A |
| **How many cell phones do you have at home? (n=110)** |  |  |  |  |  |
| >1 | 82 (74.5) | 12 (14.6) | 70 (85.4) | 2.2 (0.5-10.6) | .30 |
| 1 | 28 (25.5) | 2 (7.1) | 26 (92.9) | Referent | N/A |
| **Does your home have internet access? (n=110)** |  |  |  |  |  |
| Yes | 96 (87.3) | 12 (12.5) | 84 (87.5) | Referent | N/A |
| No | 14 (12.7) | 2 (14.3) | 12 (85.7) | 1.1 (0.2-5.8) | .85 |
| **What type of internet do you have in your home? (n=95)** |  |  |  |  |  |
| Wi-fi and satellite | 33 (34.7) | 6 (18.2) | 27 (81.8) | Referent | N/A |
| Data plan | 62 (65.3) | 18 (29) | 44 (71) | 1.8 (0.7-5.2) | .25 |
| **How many hours per day do you spend using electronic devices? (n=110)** |  |  |  |  |  |
| ≥6 hours | 21 (19.1) | 5 (23.8) | 16 (76.2) | 2.8 (0.8-9.3) | .90 |
| <1 hour or up to 5 hours | 89 (80.9) | 9 (10.1) | 80 (89.9) | Referent | N/A |
| **a-EAI[d] scores (n=108)** |  |  |  |  |  |
| Higher (top tertile) score and lower (lowest two tertiles) score | N/A | N/A | N/A | N/A | N/A |
| Two-third low score | 81 (75) | 6 (7.4) | 75 (92.6) | Referent | N/A |
| One-third high score | 27 (25) | 8 (29.6) | 19 (70.4) | 5.3 (1.6-17.0) | .003 |
| Uses social media (n=110) | 105 (95.5) | 14 (13.3) | 91 (86.7) | N/A | N/A |
| **How often do you use WhatsApp? (n=105)** |  |  |  |  |  |
| Less than daily | 6 (5.7) | 1 (16.7) | 5 (83.3) | 1.3 (0.1-12.2) | .81 |
| Daily | 99 (94.3) | 13 (13.1) | 86 (86.9) | Referent | N/A |
| **How often do you use Facebook? (n=105)** |  |  |  |  |  |
| Daily | 89 (84.8) | 12 (13.5) | 77 (86.5) | 1.1 (0.2-5.4) | .91 |
| Less than daily | 16 (15.2) | 2 (12.5) | 14 (87.5) | Referent | N/A |
| **How often do you use Instagram? (n=104)** |  |  |  |  |  |
| Daily | 43 (41.3) | 8 (18.6) | 35 (81.4) | 2.1 (0.7-6.6) | .19 |
| Less than daily | 61 (58.7) | 6 (9.8) | 55 (90.2) | Referent | N/A |
| **Technology is a tool to help us daily (n=110)** |  |  |  |  |  |

| Variables | Total participants, n (%) | ¡mZAP! use |  | Odds ratio (95% CI) | *P* value[b] |
|---|---|---|---|---|---|
|  |  | Yes | No |  |  |
| Agree | 100 (90.9) | 14 (14) | 86 (86) | N/A | N/A |
| **The use of technologies helps to fix community problems (n=110)** |  |  |  |  |  |
| Undecided or disagree | 44 (40) | 9 (20.5) | 35 (79.5) | 3.1 (0.9-10.0) | .04 |
| Agree | 66 (60) | 5 (7.6) | 61 (92.4) | Referent | N/A |
| **The use of technologies brings us closer to the community (n=109)** |  |  |  |  |  |
| Undecided or disagree | 72 (66.1) | 10 (13.9) | 62 (86.1) | 1.3 (0.3-4.6) | .65 |
| Agree | 37 (33.9) | 4 (10.8) | 33 (89.2) | Referent | N/A |
| **The constant use of technology limits the community work (n=110)** |  |  |  |  |  |
| Undecided or disagree | 67 (60.9) | 9 (13.4) | 58 (86.6) | 1.2 (0.3-3.8) | .78 |
| Agree | 43 (39.1) | 5 (11.6) | 38 (88.4) | Referent | N/A |
| **The technology cause problems in the community (n=110)** |  |  |  |  |  |
| Undecided or disagree | 74 (67.3) | 10 (13.5) | 64 (86.5) | 1.3 (0.4-4.3) | .72 |
| Agree | 36 (32.7) | 4 (11.1) | 32 (88.9) | Referent | N/A |

[a] ¡mZAP!: Zonas, Acción y Protección.

[b] *P* value of chi-square test.

[c] N/A: not applicable.

[d] a-EAI: adapted environmental attitudes inventory.

No item from the MHLC was statistically significantly associated with the use of the ¡mZAP! app, including when compared with each subscale's total score (MHLC subscale 1 [internal health locus of control], MHLC subscale 2 [powerful others health locus of control], and MHLC subscale 3 [chance health locus of control]; data not shown).

Higher positive attitudes toward the environment were significantly associated with using the ¡mZAP! app (OR 5.3, 95% CI 1.6-17.0). To assess potential confounding factors, we selected variables that were marginally associated with a-EAI and the use of the ¡mZAP! app, although none of the variables met the statistical criteria for remaining in the model. Clinically relevant sociodemographic variables (education level and age) were forced into the logistic regression model and showed no confounding between environmental attitudes and the use of ¡mZAP! app.

A statistically significant ($P<.01$) association persisted when comparing the highest tertile of a-EAI versus the other two tertiles and ¡mZAP! app use (adjusted OR 5.4, 95% CI 1.4-20.4). The Hosmer and Lemeshow goodness-of-fit test [60] was not statistically significant ($P=.89$), validating the tested model.

# Discussion

## Principal Findings

This study was followed by the *Community-Centered Environmental Health Risk Control Model*. This theoretically informed model facilitated the process of learning about *Culebrenses'* environmental health priorities and, more importantly, the implementation of their community perspectives in the development of ¡mZAP! app as an mHealth tool tailored to the community and by the community. This new model contributes to our understanding of behaviors that can lead to better ICT use among community members, and this model can be adopted in future studies worldwide.

The main findings of this study are as follows: (1) ¡mZAP! users were five times more likely to have stronger environmental attitudes than ¡mZAP! nonusers; (2) a negative relationship between environmental attitudes and the MHLC *powerful others* was observed; and (3) an expectation that the government will meet the needs of *Culebrenses* was found.

The study results also suggest that *Culebrenses* have a strong use of ICT, where most participants spend up to 5 hours daily using electronic devices. In contrast, this study also found important aspects of *Culebrenses'* potential interpretations of and implications for their attitudes toward the environment. For example, participants with a low perception of health control were also more likely to have a less positive attitude toward the environment. This association indicates that attitudes toward the environment become more negative as the sense that one's health is controlled by *powerful others* increases. Higher scores on the MHLC *powerful others* subscale have been previously demonstrated to affect health outcomes negatively [61,62]. This relationship could also lead people to become disengaged in efforts to protect their environment. People may hold conflicting beliefs because most participants agreed with or were ambivalent about items that reflected positive inclinations toward the environment.

As described in previous studies, *Culebrenses* expressed a feeling of *being forgotten* by the Puerto Rican government, which could partly explain why *Culebrenses* have a sense of pride and ownership with respect to Culebra [48]. As a result, *Culebrenses* could demonstrate a greater inherent consciousness of protecting natural resources, the environment, and fragile ecosystems, which are also the island's main tourist attractions. The results of this study help to empirically reaffirm Culebra's positive community attitudes and beliefs toward the environment. Although the intentions to protect the island were in the minds of community members, these intentions did not necessarily translate into the use of ¡mZAP! to protect the environment and address environmental health risks that affect islanders.

This study suggests that less time spent living in Culebra was associated with stronger positive attitudes toward the environment. This finding is important, particularly considering that most ¡mZAP! users have lived in Culebra for >11 years. Perhaps these results suggest that the longer one lives in a location, the more likely one is to be disincentivized to engage in behaviors to respond to community environmental challenges. For example, *Culebrenses* have been exposed to many environmental health risks in the past, including solid waste polluting beaches, improper land development, air pollution, and most notably, the US military bombing practices [35,63-67]. Although the community advocated against these environmental injustices in the past [66,67], the ongoing repercussions of these challenges [31,68] have potentially resulted in reduced enthusiasm. ¡mZAP! users had higher positive attitudes toward the environment, which could also translate into an increase in environmental conscientiousness.

Participants' beliefs that government agencies are responsible for responding to environmental risks and community discrepancies related to environmental challenges help us better understand *Culebrenses*' perceptions of lack of control over the community's health. Societal issues that affect the community's poor health outcomes extend into the social and ecological determinants of health [69,70]. The ongoing dependency on government-based solutions to community problems is worrisome, especially after the hurricane experiences of 2017 (which may potentially continue to be an issue subsequent to the more recent 2020 earthquakes).

The catastrophic impact of hurricanes Irma and María and the inappropriate responses of the Puerto Rican and US governments may have cemented this perception of lack of control. Community members experienced the isolation of being disconnected from the main island of Puerto Rico and the exacerbation of an unreliable maritime transportation system, resulting in a societal crisis where health access, common goods, and food became scarce [71,72].

*Culebrenses*' dependency systems and the way social injustices from the past and the present may have resulted in the perception of lack of control, which was associated in this study with negative environmental attitudes, can undermine the good intentions of community members to protect their land. Therefore, implementing community-driven approaches to address environmental health risks or health disparities may not be sufficient to ensure a successful mHealth tool intervention, especially when other social conditions work as oppressors. Socioecological conditions force community priorities to be in constant transformative change. This transformative change was especially true after the 2017 hurricane season in Culebra, where the emergence of new challenges such as rebuilding destroyed properties became a new top priority for *Culebrenses*.

Previous studies suggest that new mobile apps are used between 4% and 20% of the time, with an average session app use of <1 minute. Furthermore, 51% of apps are deleted after the first week of use [73-78]. In this study, only 12.6% (14/111) of participants adopted the use of ¡mZAP!. Although the number of initial users was small, it may be noteworthy that 36% (5/14) of participants used the mHealth tool more than once. These adoption and reuse rates are typical when compared with the adoption rates of other mobile apps. In addition, identifying proper champions on the island to promote ¡mZAP! presented a challenge that could have also affected the use of the mHealth tool. Early identification of community champions has been previously studied as an approach to develop and increase trust between community members and investigators, facilitating an engaging process and a successful community-based research intervention [79-81].

This research study lacked sufficient power to detect some potentially meaningful and theoretically driven associations. The observed associations that were not statistically significant but which had ORs >1.5 can, therefore, be useful for hypothesis generation and to inform future studies. Therefore, in a larger sample, certain relevant predictor variables would have statistically significant associations with the primary outcome, such as age, education level, cell phone number and type of internet per household, daily use of electronic devices, and the perception of how technology contributes to help fixing community problems.

The results of this study provide some support for suggesting that people used ¡mZAP! and were taking self-directed action by using ICTs to address environmental health challenges in Culebra. These results, although limited, confirm the objective of this study to assess the ability of ICTs as tools to engage islanders in collective actions that address environmental health risks. However, for other *Culebrenses*, although they may have similar intentions to protect the environment, existing higher community priorities and potential oppressive challenges, including health care and maritime transportation access, prevent them from translating their intentions into actions.

## Limitations and Strengths

Although this pilot study may offer some insight into directional relationships via statistically significant associations in this sample, interpretation of results is limited because of the small sample size, which restricted power. There may have been a selection bias, including volunteer bias. The data collected for this study reflect a one-time *snapshot* where questions were self-administered by participants, potentially generating respondent bias. The hurricane disasters of 2017 may have substantially affected community perceptions and priorities to the extent that the development of ¡mZAP! was affected in unexpected ways.

The empirical findings concerning perceived lack of control and government dependency systems support conclusions about how these situations could have a stronger influence on people's behaviors. These behaviors could lead to a perpetuation of the challenges associated with this population's unique social determinants of health. The results from this study should be further qualified by the fact that people who are less familiar with technology may not be as represented in this study as those who are more familiar (eg, those who are younger and have higher education). These contextual factors are important in future research to ensure the potential adoption and success of any ICT.

The study results might only apply to the users of the ¡mZAP! app, a tool that proactively focuses on environmental health risks. Anecdotally we know that some community members were positively affected by the tool (social influence from other users). The perceptions and priorities of nonusers are likely to differ. Future studies could explore the relationship between offline and internet-based generations to address communities' environmental priorities.

This study had several strengths. Existing partnerships with collaborators in Puerto Rico facilitated access to the target population and necessary local ethical review and approval. The study addressed key goals and objectives of Healthy People 2030, including "Use health communication strategies and health information technology to improve population health outcomes and health care quality, and to achieve health equity" [82]. This is the first study to provide an overview of *Culebrenses'* intentions to protect the island's environment in the context of mHealth resource use. In addition, this study provides a unique perspective on how people in remote and underresourced communities perceive environmental health risk and how those perceptions affect the use of rapidly advancing mobile technologies, which can help decrease barriers to access to health in rural areas [2,83,84]. In conclusion, the study findings demonstrated the capacity to stimulate collective action by using ICTs as a novel and engaging approach in underresourced rural locations.

### Implications and Future Research

Future research studies should seek to better understand the factors preventing the use of ¡mZAP! and other ICTs, including changes in the community's perceptions and priorities after natural disasters or other major community-wide challenges. It is important to explore how ICT use can support, facilitate, or even drive collective community actions. For example, there are multiple community-based centers, groups, and institutions, including a federally qualified health center, a women's health community organization, and other grassroot-level environmental organizations that can be supplied with ¡mZAP! and other mHealth ICT–related mobile technologies. These tools could be coupled with educational materials and community-based initiatives aimed at increasing positive perceptions of the environment by community members in this rural setting.

Future research should work toward a better understanding of community members' priorities and addressing pre-existing social determinants, such as those found in this study. Although other social determinants of health in this study, including colonialism, were not assessed, this research establishes a contributing baseline to further investigate the relationship between health or environmental disparities and the sociopolitical power imbalances that affect community islands such as Culebra [85,86]. As community-driven interventions have been successful in the past, with the understanding that community work takes time to develop, the findings of this study can serve as a foundation for future community and ICT research in Culebra and other locations with similar environmental health conditions.

## Acknowledgments

The authors would like to thank the community of Culebra, in particular the collaborators in this study from the Amigos de la Reserva Natural Caño Luis Peña organization (Nicolás Gómez Andujar, Abimarie Otaño, and Alfredo Montañez), Dulce María del Río Pineda and MariCarmen Carbonell from the Mujeres de Islas organization, HealthproMed Inc, Escuela Ecológica de Culebra, Abbie White-Thomas (Abbie's School), Foundation for a Better Puerto Rico, José D Pérez from the Municipality of Culebra, Office of Emergencies Management, and Yahaira Rivera, former Culebra's social worker from the Administración de Servicios de Salud y Contra la Adicción. The authors would also like to acknowledge Ahona Shirin, Matthew Quick, Derek Guo, and Astghik Baghinyan for their contributions to this manuscript. The authors are also thankful to The Dye Lab, including the ¡mZAP! (Zonas, Acción y Protección) developing team, Michael Kelly, and Noah Wilson. This study was funded by the University of Rochester, School of Medicine and Dentistry, Department of Obstetrics and Gynecology Pilot Research Award and the University of Rochester, Office of Equity and Inclusion supported journal publication fees. At the time of this work, the corresponding author (JGPR) was enrolled in a doctoral program and accessed REDCap (Research Electronic Data Capture) resources supported by the University of Rochester, Clinical and Translational Science Awards, award numbers TL1TR002000 and UL1TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

None declared.

## References

1. Adibi S. Mobile Health: A Technology Road Map. Berlin, Germany: Springer; 2015.

2. Mbuthia F, Reid M, Fichardt A. mHealth communication to strengthen postnatal care in rural areas: a systematic review. BMC Pregnancy Childbirth 2019 Nov 06;19(1):406 [FREE Full text] [doi: 10.1186/s12884-019-2531-0] [Medline: 31694578]
3. Environment and health in developing countries. Health and Environment Linkages Initiative, World Health Organization. URL: https://www.who.int/heli/risks/ehindevcoun/en/ [accessed 2021-08-11]
4. Neira M, Prüss-Ustün A. Preventing disease through healthy environments: A global assessment of the environmental burden of disease. Toxicol Lett 2016 Oct 10;259:S1. [doi: 10.1016/j.toxlet.2016.07.028]
5. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
6. Norman CD, McIntosh S, Selby P, Eysenbach G. Web-assisted tobacco interventions: empowering change in the global fight for the public's (e)Health. J Med Internet Res 2008 Nov 25;10(5):e48 [FREE Full text] [doi: 10.2196/jmir.1171] [Medline: 19033147]
7. Haenssgen MJ, Charoenboon N, Zanello G. You've got a friend in me: how social networks and mobile phones facilitate healthcare access among marginalised groups in rural Thailand and Lao PDR. World Dev 2021 Jan;137:105156. [doi: 10.1016/j.worlddev.2020.105156]
8. Johnston FH, Wheeler AJ, Williamson GJ, Campbell SL, Jones PJ, Koolhof IS, et al. Using smartphone technology to reduce health impacts from atmospheric environmental hazards. Environ Res Lett 2018 Mar 29;13(4):044019. [doi: 10.1088/1748-9326/aab1e6]
9. Jafari H, Li X, Qian L, Chen Y. Community based sensing: a test bed for environment air quality monitoring using smartphone paired sensors. In: Proceedings of the 36th IEEE Sarnoff Symposium. 2015 Presented at: 36th IEEE Sarnoff Symposium; Sept 20-22, 2015; Newark, NJ, USA. [doi: 10.1109/sarnof.2015.7324635]
10. Van 'T Hooft M, DÍaz S, Swan K. Examining the potential of handheld computers: findings from the Ohio PEP project. J Educ Comput Res 2016 Jun 1;30(4):295-311. [doi: 10.2190/m1w6-a94d-3nkm-kbuu]
11. Swan K, Hooft M, Kratcoski A, Unger D. Uses and effects of mobile computing devices in K–8 classrooms. J Res Technol Educ 2005;38(1):99-112. [doi: 10.1080/15391523.2005.10782451]
12. English P, Richardson M, Garzón-Galvis C. From crowdsourcing to extreme citizen science: participatory research for environmental health. Annu Rev Public Health 2018 Apr 01;39(1):335-350. [doi: 10.1146/annurev-publhealth-040617-013702] [Medline: 29608871]
13. Uzunboylu H, Cavus N, Ercag E. Using mobile learning to increase environmental awareness. Comp Educ 2009 Feb;52(2):381-389. [doi: 10.1016/j.compedu.2008.09.008]
14. Benelli G, Jeffries C, Walker T. Biological control of mosquito vectors: past, present, and future. Insects 2016 Oct 03;7(4):52 [FREE Full text] [doi: 10.3390/insects7040052] [Medline: 27706105]
15. Cañizares JC. The information society: technological, socioeconomic and cultural aspects - prolegomena for a sustainability-oriented ethics of ICT, with glossary and PhD prop. University of Twente Student Theses. 2017. URL: http://essay.utwente.nl/73135/ [accessed 2021-08-11]
16. Hoffman D, Novak T, Schlosser A. The evolution of the digital divide: how gaps in Internet access may impact electronic commerce. J Comput Mediat Commun 2000 Mar 1;5(3):JCMC534. [doi: 10.1111/j.1083-6101.2000.tb00341.x]
17. Nemer D. Online favela: the use of social media by the marginalized in Brazil. Inf Technol Dev 2015 Apr 18;22(3):364-379. [doi: 10.1080/02681102.2015.1011598]
18. Nemer D, Tsikerdekis M. Political engagement and ICTs: internet use in marginalized communities. J Assoc Inf Sci Technol 2017 Apr 28;68(6):1539-1550. [doi: 10.1002/asi.23779]
19. Xu Y, Yen DC, Lin B, Chou DC. Adopting customer relationship management technology. Ind Manag Data Syst 2002 Nov;102(8):442-452. [doi: 10.1108/02635570210445871]
20. Levine AR, Buchner JA, Verceles AC, Zubrow MT, Mallemat HA, Papali A, et al. Ultrasound images transmitted via FaceTime are non-inferior to images on the ultrasound machine. J Crit Care 2016 Jun;33:51-55. [doi: 10.1016/j.jcrc.2016.02.019] [Medline: 27006267]
21. Madon S, Krishna S. The Digital Challenge: Information Technology in the Development Context. Oxfordshire, England: Routledge; Feb 6, 2018:1-386.
22. Taub A, Fisher M. How YouTube misinformation resolved a WhatsApp mystery in Brazil. The New York Times. 2019 Aug 15. URL: https://www.nytimes.com/2019/08/15/the-weekly/how-youtube-misinformation-resolved-a-whatsapp-mystery-in-brazil.html?action=click&module=Briefings&pgtype=Homepage [accessed 2021-08-11]
23. Boyd A, Gatewood J, Thorson S, Dye T. Data diplomacy. Science & Diplomacy. URL: https://www.sciencediplomacy.org/article/2019/data-diplomacy [accessed 2021-08-11]
24. Minkler M. Community Organizing and Community Building for Health and Welfare. New Brunswick, New Jersey: Rutgers University Press; 2012.
25. Qureshi S, Najjar L. Information and communications technology use and income growth: evidence of the multiplier effect in very small island states. Inf Technol Dev 2017 Apr 26;23(2):212-234. [doi: 10.1080/02681102.2016.1173634]
26. Baldacchino G. A World of Islands: An Island Studies Reader. Charlottetown: University of Prince Edward Island; 2007.

27. Dye TD, Sy A, Albert P, Cash H, Hadley J, Tomeing T, et al. Critical medical ecological perspectives on diabetes in the Pacific Islands: colonialism, power, and balance in human-environment interaction over time. Lancet Glob Health 2018 Mar 1;6:S36. [doi: 10.1016/s2214-109x(18)30165-7]
28. Douglas CH. Small island states and territories: sustainable development issues and strategies – challenges for changing islands in a changing world. Sust Dev 2006 Mar 3;14(2):75-80. [doi: 10.1002/sd.297]
29. Davis JS, Hayes-Conroy JS, Jones VM. Military pollution and natural purity: seeing nature and knowing contamination in Vieques, Puerto Rico. GeoJ 2007 Aug 3;69:165-179. [doi: 10.1007/s10708-007-9095-7]
30. Cintrón C. Culebra y la Marina de Estados Unidos. San Juan, Puerto Rico: Editorial Edil; Jan 1, 1989.
31. Rogers C, Cintrón G, Goenaga C. The impact of military operations on the coral reefs of Vieques and Culebra. Report to Department of Natural Resources, San Juan, Puerto Rico. 1978. URL: http://www.reefbase.org/resource_center/publication/main.aspx?refid=4823 [accessed 2021-08-11]
32. Vistazo a la Salud Pública en Culebra. Dialogo UPR. 2015. URL: http://dialogoupr.com/vistazo-la-salud-publica-en-culebra/ [accessed 2021-08-11]
33. Martinez RB. La Independencia de Puerto Rico (Razon y Lucha). San Juan, Puerto Rico: Editorial Linea; Jan 1, 1983.
34. Culebra recibirá desde hoy electricidad por cable submarino desde Vieques. Agencia EFE. URL: https://www.efe.com/efe/usa/puerto-rico/culebra-recibira-desde-hoy-electricidad-por-cable-submarino-vieques/50000110-3930344 [accessed 2021-08-11]
35. Feliciano C. Apuntes y Comentarios de la Colonización y Liberación de la Isla de Culebra. : Self-Published; Jan 1, 2001.
36. Asociación de Ejecutivos de Ventas y Mercadeo de Puerto Rico (SME). Digital & Mobile Behavioral Study. 2016. URL: https://www.smepr.org/index.php?src=news&srctype=detail&category=Highlighted&refno=175 [accessed 2019-05-30]
37. Haider M, Kreps GL. Forty years of diffusion of innovations: utility and value in public health. J Health Commun 2004;9 Suppl 1:3-11. [doi: 10.1080/10810730490271430] [Medline: 14960400]
38. Rawstorn JC, Gant N, Meads A, Warren I, Maddison R. Remotely delivered exercise-based cardiac rehabilitation: design and content development of a novel mHealth platform. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e57 [FREE Full text] [doi: 10.2196/mhealth.5501] [Medline: 27342791]
39. Henrich J. Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. Am Anthropol 2001 Dec;103(4):992-1013. [doi: 10.1525/aa.2001.103.4.992]
40. IDEO (Firm). The Field Guide to Human-centered Design: Design Kit. San Francisco, California: IDEO; 2015.
41. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. London, United Kingdom: Pearson; Nov 1, 1985.
42. Rogers E. Diffusion of Innovations. Mumbai: Free Press; Jul 6, 2010.
43. American Community Survey (ACS). The United States Census Bureau. 2017. URL: https://www.census.gov/programs-surveys/acs [accessed 2021-08-11]
44. Culebra National Wildlife Refuge. United States Fish and Wildlife Service. URL: https://www.fws.gov/refuge/Culebra/wildlife_and_habitat.html [accessed 2021-08-11]
45. Manifestantes llevan reclamos sobre lanchas para Vieques y Culebra a La Fortaleza y al Capitolio. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/notas/manifestantes-llevan-reclamos-sobre-lanchas-para-vieques-y-culebra-a-la-fortaleza-y-al-capitolio/ [accessed 2021-08-11]
46. Hernández-Delgado EA, Acuña AM, Otaño-Cruz A, Suleiman-Ramos SE. Los arrecifes de coral con craters-bomba en Puerto Rico, la historia no contada sobre un hábitat inusual: desde la destrucción de arrecifes hasta la rehabilitación ecológica basada en la comunidad. Rev Biol Trop 2014 Sep 01;62(Suppl 3):183-200. [doi: 10.15517/rbt.v62i0.15913]
47. Rivera-Arguinzoni A. Seis municipios donde no nacen bebés. Elnuevodía.com. URL: https://www.elnuevodia.com/noticias/locales/notas/seis-municipios-donde-no-nacen-bebes/ [accessed 2021-08-11]
48. Perez-Ramos JG, McIntosh S, Barrett ES, Vélez-Vega CM, Dye TD. Qualitative assessment of environmental health risk perceptions and community challenges in a Puerto Rican community: change and continuity in response to hurricanes Irma and María. Behav Med 2020;46(3-4):231-244. [doi: 10.1080/08964289.2019.1700894] [Medline: 31860413]
49. McIntosh S, Perez-Ramos J, Demment MM, Vélez Vega C, Avendaño E, Ossip DJ, et al. Development and implementation of culturally tailored offline mobile health surveys. JMIR Public Health Surveill 2016 Jun 02;2(1):e28 [FREE Full text] [doi: 10.2196/publichealth.5408] [Medline: 27256208]
50. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
51. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res 1981 Nov 1;10(2):141-163. [doi: 10.1177/004912418101000205]
52. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015 Oct 6;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
53. Milfont TL, Duckitt J. The environmental attitudes inventory: a valid and reliable measure to assess the structure of environmental attitudes. J Environ Psychol 2010 Mar;30(1):80-94. [doi: 10.1016/j.jenvp.2009.09.001]

54. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, United Kingdom: Oxford University Press; 2008.
55. De Jesus M, Xiao C. Predicting health care utilization among Latinos: health locus of control beliefs or access factors? Health Educ Behav 2014 Aug 04;41(4):423-430. [doi: 10.1177/1090198114529130] [Medline: 24707031]
56. Tomás-Sábado J, Montes-Hidalgo J. [Spanish version of the Multidimensional health locus of control scale innursing students]. Enferm Clin 2016;26(3):181-187. [doi: 10.1016/j.enfcli.2015.12.005] [Medline: 26922046]
57. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) scales. Health Educ Monogr 1978;6(2):160-170. [doi: 10.1177/109019817800600107] [Medline: 689890]
58. An innovative training program in Latin America for maternal health improvement. MundoComm. URL: http://mundocomm.org/ [accessed 2021-08-11]
59. IBM SPSS Statistics, Version 25. IBM Corp. 2018. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2021-08-11]
60. Hosmer JD, Lemeshow S, Sturdivant R. Applied Logistic Regression, 3rd Edition. Hoboken, New Jersey: Wiley; Mar 2013.
61. Christensen AJ, Turner CW, Smith TW, Holman JM, Gregory MC. Health locus of control and depression in end-stage renal disease. J Consult Clin Psychol 1991 Jun;59(3):419-424. [doi: 10.1037//0022-006x.59.3.419] [Medline: 2071727]
62. Poortinga W, Dunstan FD, Fone DL. Health locus of control beliefs and socio-economic differences in self-rated health. Prev Med 2008 Apr;46(4):374-380. [doi: 10.1016/j.ypmed.2007.11.015] [Medline: 18177930]
63. López-Alicea K. Montaña de basura al lado de playa Tamarindo en Culebra. Montaña De Basura Al Lado De Playa Tamarindo en Culebra. URL: https://www.cesampr.com/blog/montaa-de-basura-al-lado-de-playa-tamarindo-en-culebra [accessed 2021-08-11]
64. Varsavky M. Vivir entre la basura en Culebra, Puerto Rico. MartinVarsavsky.net. URL: https://spanish.martinvarsavsky.net/general/vivir-entre-la-basura-en-culebra.html [accessed 2021-08-11]
65. Siguen los daños ambientales en Playa Flamenco, Culebra. PrimeraHora.Com. URL: https://www.primerahora.com/noticias/puerto-rico/notas/siguen-los-danos-ambientales-en-playa-flamenco-culebra/ [accessed 2021-08-11]
66. Landauer J. Culebrans fire back: islanders seek to end role as a Navy target - shells fall close to governor of Puerto Rico. Wall St J 1970 Jun 10:A [FREE Full text]
67. Pérez-Rodríguez T. Movilización Social y Política Contra la Presencia de la Marina en Culebra 1960-1975. Publicaciones Puertorriqueñas 2013:A [FREE Full text]
68. Carrasco C. Munitions cleanup on Culebra advances, increasing safety for public recreational use. Defense Visual Information Distribution Service. URL: https://www.dvidshub.net/news/333084/munitions-cleanup-culebra-advances-increasing-safety-public-recreational-use [accessed 2021-08-11]
69. Skinner EA. A guide to constructs of control. J Pers Soc Psychol 1996 Sep;71(3):549-570. [doi: 10.1037//0022-3514.71.3.549] [Medline: 8831161]
70. Bosma H, van de Mheen HD, Borsboom G, Mackenbach J. Neighborhood socioeconomic status and all-cause mortality. Am J Epidemiol 2001 Feb 15;153(4):363-371. [doi: 10.1093/aje/153.4.363] [Medline: 11207154]
71. Protestan culebrenses por servicio de lanchas. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/nota/protestanculebrensesporserviciodelanchas-779136/ [accessed 2021-08-11]
72. Ruiz-Kuilan G. Residentes de las islas municipio denuncian que la transportación ha empeorado. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/nota/residentesdelasislasmunicipiodenuncianquelatransportacionhaempeorado-2472156/ [accessed 2021-08-11]
73. Gordon N. 55+ jaw dropping app usage statistics in 2021 [infographic]. Techjury. URL: https://techjury.net/stats-about/app-usage/ [accessed 2021-08-11]
74. Böhmer M, Hecht B, Schöning J, Krüger A, Bauer G. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. 2011 Aug Presented at: MobileHCI '11: 13th International Conference on Human Computer Interaction with Mobile Devices and Services; Aug 30-Sep 2, 2011; Stockholm Sweden p. 47-56. [doi: 10.1145/2037373.2037383]
75. Lim S, Bentley P. Investigating app store ranking algorithms using a simulation of mobile app ecosystems. In: Proceedings of the IEEE Congress on Evolutionary Computation. 2013 Presented at: IEEE Congress on Evolutionary Computation; June 20-23, 2013; Cancun, Mexico. [doi: 10.1109/cec.2013.6557892]
76. Shepard C, Rahmati A, Tossell C, Zhong L, Kortum P. LiveLab: measuring wireless networks and smartphone users in the field. SIGMETRICS Perform Eval Rev 2010 Dec;38(3):15-20. [doi: 10.1145/1925019.1925023]
77. O'Connell C. 23% of users abandon an app after one use. DZone. URL: https://dzone.com/articles/23-of-users-abandon-an-app-after-one-use [accessed 2021-05-11]
78. The State of Mobile 2021 Report. AppAnnie, San Francisco, California, USA. URL: https://www.appannie.com/en/go/state-of-mobile-2021/ [accessed 2021-08-11]
79. Worthy SL, Downey L, Dyk PH, Monroe PA, Tyler-Mackey C, Welborn R. Turning the tide on poverty: community champions as critical elements of success in economically distressed communities. Community Dev 2016 Mar 31;47(3):341-357. [doi: 10.1080/15575330.2016.1167096]

80. Shields-Zeeman L, Pathare S, Walters BH, Kapadia-Kundu N, Joag K. Promoting wellbeing and improving access to mental health care through community champions in rural India: the intervention approach. Int J Ment Health Syst 2017 Jan 4;11:6 [FREE Full text] [doi: 10.1186/s13033-016-0113-3] [Medline: 28066505]
81. Leadbeater BJ. Engaging Community Champions in the Prevention of Bullying: Understanding and Addressing Bullying: An International Perspective. Bloomington: AuthorHouse; 2008.
82. United States Department of Health and Human Services. Healthy people 2030 framework. Office of Disease Prevention and Health Promotion. URL: https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Framework [accessed 2020-04-04]
83. Uddin J, Biswas T, Adhikary G, Ali W, Alam N, Palit R, et al. Impact of mobile phone-based technology to improve health, population and nutrition services in rural Bangladesh: a study protocol. BMC Med Inform Decis Mak 2017 Jul 06;17(1):101 [FREE Full text] [doi: 10.1186/s12911-017-0502-9] [Medline: 28683742]
84. Peprah P, Abalo EM, Agyemang-Duah W, Budu HI, Appiah-Brempong E, Morgan AK, et al. Lessening barriers to healthcare in rural Ghana: providers and users' perspectives on the role of mHealth technology. A qualitative exploration. BMC Med Inform Decis Mak 2020 Feb 10;20(1):27 [FREE Full text] [doi: 10.1186/s12911-020-1040-4] [Medline: 32041608]
85. Czyzewski K. Colonialism as a broader social determinant of health. Int Indig Policy J 2011 May;2(1):5. [doi: 10.18584/iipj.2011.2.1.5]
86. Rodríguez-Díaz CE. Community resistance as public health activism in Puerto Rico. Am J Public Health 2020 Oct;110(10):1454-1455. [doi: 10.2105/AJPH.2020.305884] [Medline: 32903086]

## Abbreviations

**¡mZAP!:** Zonas, Acción y Protección
**a-EAI:** adapted environmental attitudes inventory
**FAITH!:** Fostering African American Improvement in Total Health
**ICT:** information and communication technology
**mHealth:** mobile health
**MHLC:** multidimensional health locus of control
**OR:** odds ratio
**REDCap:** Research Electronic Data Capture
**SCT:** social cognitive theory

*Edited by R Kukafka; submitted 29.09.20; peer-reviewed by Z Quiñones, YC Chen; comments to author 13.02.21; revised version received 27.02.21; accepted 19.07.21; published 23.09.21*

*Please cite as:*
*Perez-Ramos JG, McIntosh S, Barrett ES, Velez Vega CM, Dye TD*
*Attitudes Toward the Environment and Use of Information and Communication Technologies to Address Environmental Health Risks in Marginalized Communities: Prospective Cohort Study*
*J Med Internet Res 2021;23(9):e24671*
*URL: https://www.jmir.org/2021/9/e24671*
*doi: 10.2196/24671*
*PMID:*

Review

# Evolutionary Overview of Consumer Health Informatics: Bibliometric Study on the Web of Science from 1999 to 2019

Wei Ouyang[1,2,3*], PhD; Wenzhao Xie[1,2,3*], PhD; Zirui Xin[1,2,3], PhD; Haiyan He[1,2,3], PhD; Tingxiao Wen[2,3], PhD; Xiaoqing Peng[1,2,3], PhD; Pingping Dai[1,2,3], PhD; Yifeng Yuan[2,3,4], MS; Fei Liu[1,2,3], MS; Yang Chen[1,2,3], MS; Aijing Luo[4], PhD

[1]The Third Xiangya Hospital, Central South University, Changsha, China
[2]School of Life Sciences, Central South University, Changsha, China
[3]Key Laboratory of Medical Information Research, Central South University, College of Hunan Province, Changsha, China
[4]The Second Xiangya Hospital, Central South University, Changsha, China
[*]these authors contributed equally

**Corresponding Author:**
Aijing Luo, PhD
The Second Xiangya Hospital
Central South University
139 Renmin Middle Road
Changsha, 410011
China
Phone: 86 073188618316
Email: luoaijing@163.com

## Abstract

**Background:** Consumer health informatics (CHI) originated in the 1990s. With the rapid development of computer and information technology for health decision making, an increasing number of consumers have obtained health-related information through the internet, and CHI has also attracted the attention of an increasing number of scholars.

**Objective:** The aim of this study was to analyze the research themes and evolution characteristics of different study periods and to discuss the dynamic evolution path and research theme rules in a time-series framework from the perspective of a strategy map and a data flow in CHI.

**Methods:** The Web of Science core collection database of the Institute for Scientific Information was used as the data source to retrieve relevant articles in the field of CHI. SciMAT was used to preprocess the literature data and construct the overlapping map, evolution map, strategic diagram, and cluster network characterized by keywords. Besides, a bibliometric analysis of the general characteristics, the evolutionary characteristics of the theme, and the evolutionary path of the theme was conducted.

**Results:** A total of 986 articles were obtained after the retrieval, and 931 articles met the document-type requirement. In the past 21 years, the number of articles increased every year, with a remarkable growth after 2015. The research content in 4 different study periods formed the following 38 themes: *patient education*, *medicine*, *needs*, and *bibliographic database* in the 1999-2003 study period; *world wide web*, *patient education*, *eHealth*, *patients*, *medication*, *terminology*, *behavior*, *technology*, and *disease* in the 2004-2008 study period; *websites*, *information seeking*, *physicians*, *attitudes*, *technology*, *risk*, *food labeling*, *patient*, *strategies*, *patient education*, and *eHealth* in the 2009-2014 study period; and *electronic medical records*, *health information seeking*, *attitudes*, *health communication*, *breast cancer*, *health literacy*, *technology*, *natural language processing*, *user-centered design*, *pharmacy*, *academic libraries*, *costs*, *internet utilization*, and *online health information* in the 2015-2019 study period. Besides, these themes formed 10 evolution paths in 3 research directions: patient education and intervention, consumer demand attitude and behavior, and internet information technology application.

**Conclusions:** Averaging 93 publications every year since 2015, CHI research is in a rapid growth period. The research themes mainly focus on patient education, health information needs, health information search behavior, health behavior intervention, health literacy, health information technology, eHealth, and other aspects. Patient education and intervention research, consumer demand, attitude, and behavior research comprise the main theme evolution path, whose evolution process has been relatively stable. This evolution path will continue to become the research hotspot in this field. Research on the internet and information technology application is a secondary theme evolution path with development potential.

***(J Med Internet Res 2021;23(9):e21974)*** doi: 10.2196/21974

**KEYWORDS**

consumer health informatics; consumer health information; thematic evaluation; co-word analysis; informatics; SciMAT

## *Introduction*

The concept prototype of consumer health informatics (CHI) was first proposed by Kenneth R. Thornton of the School of Health Information Science (University of Victoria) in 1994. Kenneth R. Thornton also elaborated 4 important research directions: network evolution, automation of the patient record, outcome and other quality-related databases, and consumer health education [1]. In 1995, Ferguson [2] of the Harvard University School of Medicine put forward the concept of consumer health informatics (CHI) for the first time in his paper. He pointed out that CHI is a branch of science that studies the application of computer and wireless communication technology in consumer health care [2]. Eysenbach [3] proposed that CHI is the branch of medical information that does the following: analyses the consumers' needs for information; studies and implements methods of making information accessible to consumers; and models and integrates the consumers' preferences into medical information systems. This definition has been cited by the academic community more than 300 times. The American Medical Informatics Association defines CHI as the field devoted to informatics from multiple consumer or patient views and which includes patient-focused informatics, health literacy, and consumer education [4]. CHI is an interdisciplinary subject that includes nursing information, public health, health promotion, health education, library science, and communication science [4].

With the rapid development of computer and information technology, the continuous popularization of the internet, and the continuous strengthening of people's health awareness, consumption awareness, and information literacy for making health decisions, increasingly more people obtain health-related information through the internet. Because of the interdisciplinary nature of CHI, there is no consensus on the definition of CHI. However, the core content of various definitions can be summarized as follows: CHI emphasizes consumers or users as the center and takes consumers or users and computer and information technology as the research object to explore how to use computer and information technology to meet consumers' health or medical information needs. It is a discipline that helps consumers access relevant health or medical information and make decisions about health care and health promotion.

Previous studies have analyzed and summarized the research progress of CHI. For example, Eysenbach [3] collected literature, internet information, and reports related to CHI before 2000 and summarized the research progress of CHI in the context of health care in the information age, medical knowledge delivery to consumers, the accessibility of electronic health records to patients, decision aids to support the consumers' choices, the quality control of health information on the internet, and other aspects. Kokol et al [5] carried out a bibliometric analysis on the literature related to health informatics and electronic health for the period 1984-2015 and discussed the current research status in this field, including trends in literary production, the geographic and journal distribution, and theoretical analyses. However, their research only clusters all research topics and does not reflect the dynamic changes in research themes at different times. Zhao and Zhang [6] reviewed the literature about consumer health information seeking in social media before 2016 and discussed the characteristics of existing research from the following perspectives: the prevalence of health information seeking in social media, discussion topics emerging from health information in social media, seeking health information from online peers, social and emotional support from social media, concerns of accessing consumer health information in social media, and other aspects. However, they did not discuss the dynamic evolution path and the evolution rules of the research themes of CHI. Further, many of these studies are qualitative research. There is also considerable subjectivity in the selection of literature, identification of important topics, and prediction of research frontiers. The combination of qualitative and quantitative analysis methods can increase the objectivity, accuracy, and comprehensiveness of the research results [7]. Bibliometrics is a measurable informatic method [8], and is often used to discover top journals and authors in a field, identify research progress [9], and predict research trends [10].

Many bibliometrics visualization tools are available, such as SciMAT, CiteSpace, UCINET, HistCite, VOSviewer. The full name of SciMAT is the Science Mapping Analysis tool. In 2012, it was developed by Cobo, López-Herrera, Herrera-Viedma, and Herrera at the Department of Computer and Artificial Intelligence in Granada University, Spain. It can be used for data preprocessing, data network analysis documentation, and result visualization. It can also be used to produce 4 kinds of maps: an overlapping map, an evolution map, a strategic diagram, and a cluster network [11]. Compared with other bibliometric visualization tools, SciMAT is unique in expressing the evolution of the theme and performs excellently in longitudinal timing analysis [7].

Overall, there have been few achievements in the field of CHI research based on bibliometrics, and the dynamic evolution of the research theme of CHI has not been explored. Therefore, this article conducted a bibliometric analysis to present the evolutionary overview of CHI by using SciMAT. The main research questions of this paper are as follows:

- What was the development trend of CHI research in 1999-2019?
- What is the main research direction of CHI research?
- How does the theme of CHI research evolve?
- What are the research trends of CHI?

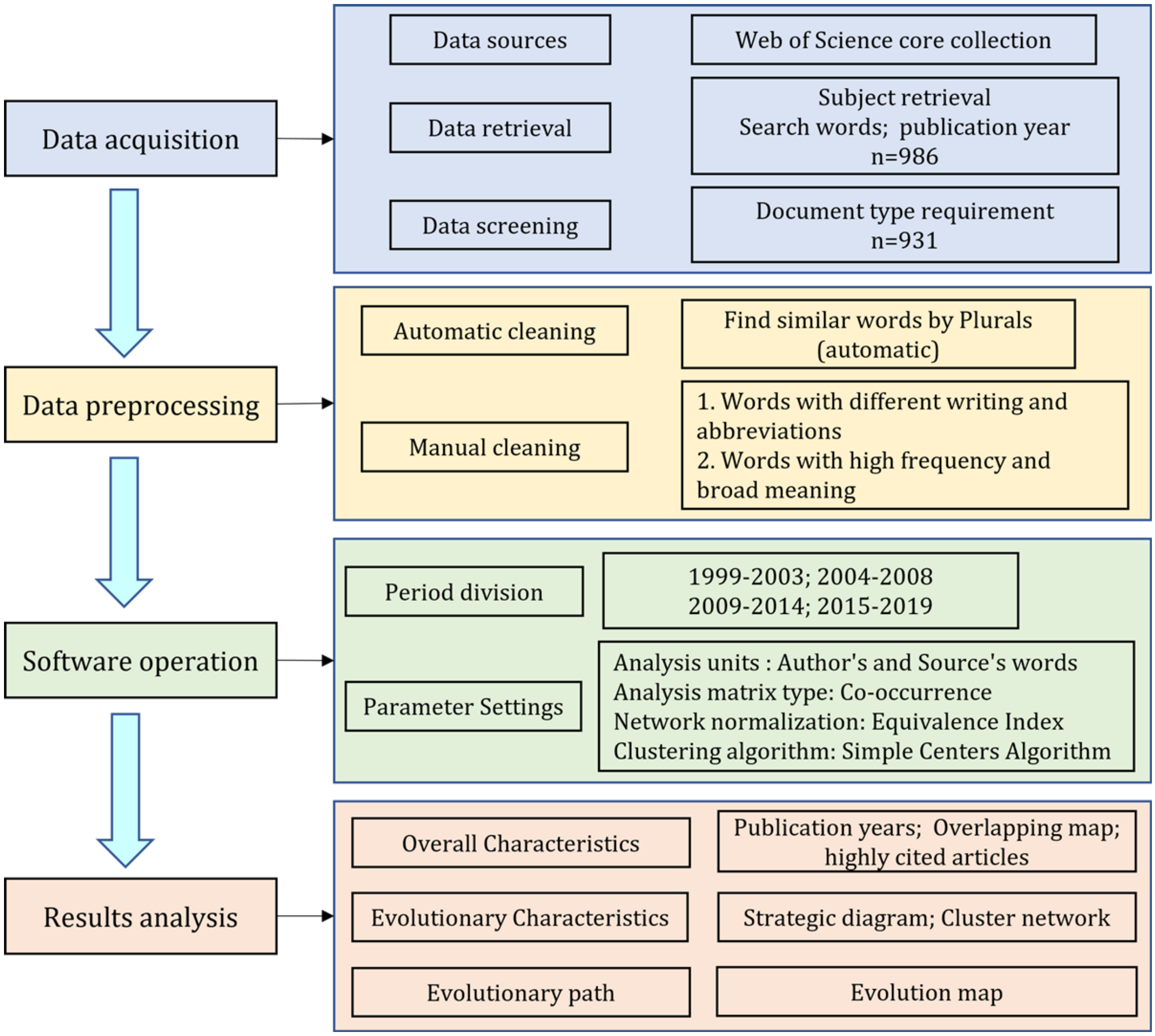
## Methods

### Overview

In bibliometrics, to obtain the research topic of a certain research field in a specific period, it is common to carry out keyword co-occurrence analysis and cluster analysis on the literature collected and to obtain the development path and state of the research topic through a comparative analysis of the split and fusion of research themes in different periods. This paper uses SciMAT tool to draw the research themes' knowledge map, analyzes the research themes and evolution status of CHI in different study periods using a strategic diagram and data flow, and discusses the dynamic evolution path and evolution law of the research themes of CHI.

The analysis framework of this paper is shown in Figure 1, which is mainly divided into 4 parts: data acquisition, data preprocessing, software operation, and results analysis.

**Figure 1.** Analysis framework of consumer health informatics evolutionary.



### Data Sources and Retrieve Strategies

The Web of Science core collection database of the Institute for Scientific Information is an important database for obtaining global academic information [7]. It consists of the Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, and so on. Users can access core academic literature from the fields of natural sciences, social sciences, biomedicine, engineering, arts, and humanities from the database. Therefore, we selected Web of Science as the data source. In this database, subject retrieval was used, and the search terms were “consumer health informatics”, “consumer health information”, “consumer medic * informatics”, “consumer medic * information”. The publication years were limited to 2019 and before. All languages were selected. Selecting all languages provides a more comprehensive coverage of the literature on this topic. Non-English articles generally have English titles, abstracts, and keywords, from which one can understand their main research content. For these articles, we also used translation software to read the full text and obtained the main research content. The literature types were limited to articles, proceeding papers, and reviews. All papers were retrieved on May 2, 2020. A total of 986 documents were obtained, of which 931 met the document-type requirement. The downloaded data were saved in a text format that SciMAT could read directly.

### Data Preprocessing

To accurately obtain the research themes in this field, it was necessary to perform data cleaning. SciMAT was used to clean keyword information. First, SciMAT's cleaning function "find similar words by Plurals (automatic)" automatically merges the singular and plural expressions of keywords, such as "attitude“ and "attitudes" into "attitudes". Then, similar words with different forms and abbreviations were merged by hand. For example, "consumer health information technology," "consumer health IT," "consumer health information technologies," and "consumer health information technology (CHIT)" were merged into " consumer health information technology." synonyms, such as "physician patient relations," "physician–patient relationship," "patient–physician relationship," "doctor–patient relationship," "doctor–patient relations," and "doctor–patient relationships" were merged into "doctor–patient relationship." Some words, such as "consumer health information," "consumer health informatics," "association," and "campaign," which had high frequency and broad meaning and that might cover the association between other micro words, were deleted.

### Parameter Settings

Words (the author's words and the source's words) were selected as the units of analysis. The data reduction thresholds of the 4 study periods are 1, 1, 1, and 2, and the type of analysis matrix is co-occurrence. The network reduction thresholds are 1, 1, 1, and 1, and the network normalization method is the equivalence index. The clustering algorithm used was the simple centers algorithm. The maximum network value was 15 and the minimum network value was 3, and the scale of the cluster network was limited to a reasonable range. The H-index and sum citations were selected as the measurement indexes of clustering quality. Jaccard's index and the Salton index were selected as similarity measurement methods for the evolution map and the overlapping map, respectively.

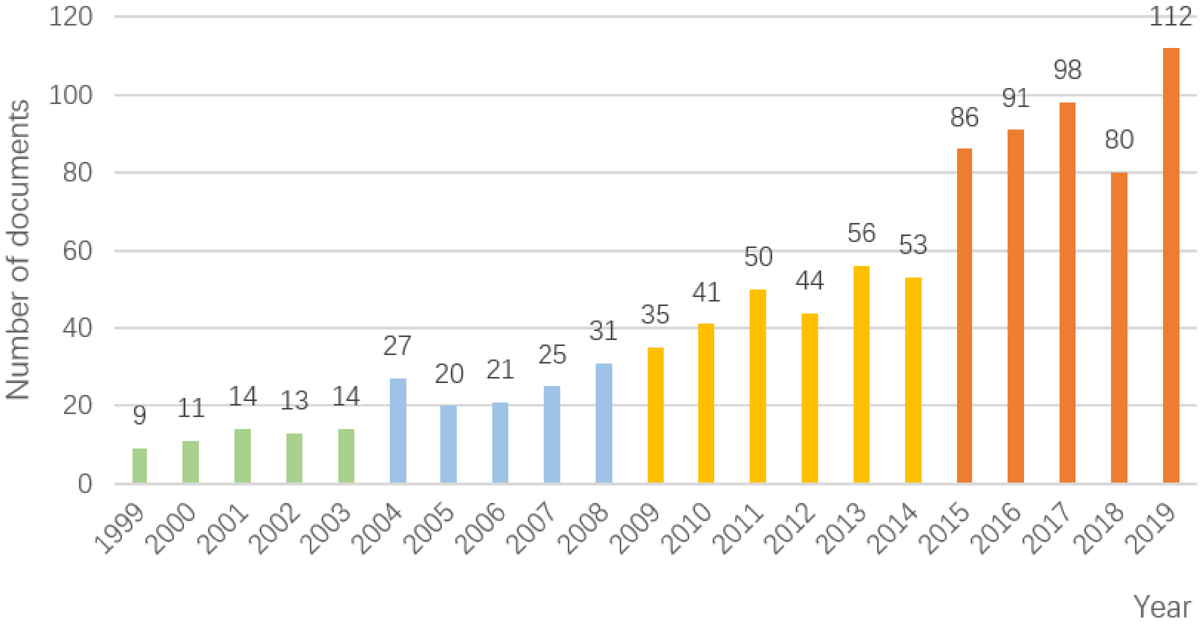
## Results

### Overall Characteristics Analysis

#### *Trends in the Number of Articles*

The number and trend of the articles published each year in a certain field can reflect the scholars' attention to this field. As shown in Figure 2, the amount of literature can be divided into 3 stages: slow growth, stable growth, and rapid growth. The 1999-2003 period was the first stage, with an average of 12.2 papers published each year. The period from 2004 to 2014 was the second stage, with an average of 36.6 papers published each year. The period from 2015 to 2019 was the third stage. After 2015, the number of articles increased substantially, with an average of 93.4 papers published each year. The volume of published articles peaked in 2019 at 112, with an annual total accounting for 12.0% (112/931) of all literature.

**Figure 2.** Consumer health informatic research documents published from 1999 to 2019.

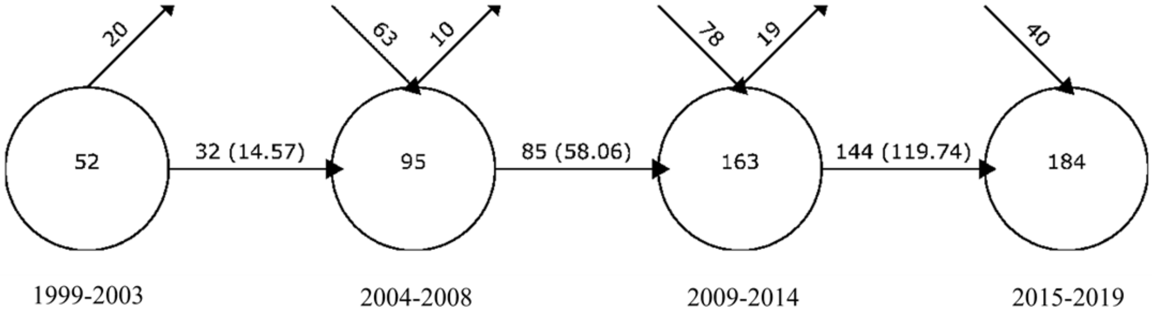


Based on the trend of literature growth, this paper divides the research on CHI into 4 study periods: 1999-2003, the first study period; 2004-2008, the second study period; 2009-2014, the third study period; and 2015-2019, the fourth study period. The number of articles in the 4 study periods is 61, 124, 279, and 467.

#### *Overlapping Map Analysis*

An overlapping map uses the number of keywords to represent the number of themes at each period and shows the stability of a research theme in a certain field in the form of a data flow. Figure 3 is an overlapping map of CHI research for the period from 1999 to 2019 and clearly shows the emerging and declining themes in this field. The 4 circles in the figure represent the 4 study periods. Displayed from left to right, they are as follows: the 1999-2003 study period, the 2004-2008 study period, the 2009-2014 study period, and the 2015-2019 study period. The number of themes in the first, second, third, and fourth study periods is 52, 95, 163, and 184, respectively. As can be seen, the number of themes increases rapidly.

**Figure 3.** Overlapping map of the consumer health informatics research from 1999 to 2019.



From 1999 to 2003, a total of 32 themes were passed to the next study period, and the stability index with the next study period was 14.57. In this study period, the research on CHI remains in a slow-growth stage, which is part of the basic research stage. The number of new themes in the 2004-2008 study period was 63. In the 2009-2014 period, 85 themes were continued from the previous study period, and the stability index between the 2 study periods was 58.06. Reflecting the steady growth of CHI research and the steady increase of research articles in these 2 study periods, many emerging themes and a small number of declining themes were included. There were 40 new themes in the 2015-2019 study period, and the stability index between this and the previous study period was 119.74. At this point, the number of themes is increasing steadily, which reflects that the research area of CHI is gradually expanding, and the research content is becoming increasingly richer.

### Analysis of Highly Cited Articles

Six of the top 10 highly cited articles in this field came from the 1999-2003 study period (Table 1), and the cited frequency of the articles is ranked as the first, second, third, fifth, sixth, and tenth. This reflects that the papers published in this period form the basis of the research on CHI and play a fundamental role in the research in this field. The most frequently cited paper (940 times) is an empirical study of the quality evaluation of online consumer health information. The evaluation of health-related websites is quite different due to the differences in research methods, preciseness, quality standards, research population, and subjects. Therefore, the operability of quality standards needs to be defined [12]. The second most frequently cited (804 times) paper is the one by Cline in 2001 [13]. This paper summarizes, from the perspective of communication, the potential benefits and comprehensive quality of online consumer health information searching and identifies and discusses the criteria of online health information evaluation. In 1999, Charnock et al [14] developed a tool named DISCERN, which can be used by health information providers and consumers to judge the quality of written consumer health information about treatment options. This paper was cited 521 times. At present, DISCERN has become a highly useful tool for the quality evaluation of network health information. Many scholars use this tool to evaluate the quality of online health information about different diseases and published in different languages. For example, Cerminara et al [15] used the DISCERN tool to evaluate the reliability, accuracy, and relevance of the top 50 links for childhood epilepsy (online information) displayed by the Google search engine. Alnaim [16] used the DISCERN tool to assess the quality of information published on websites that share breast cancer information online in Arabic.

Four of the top 10 highly cited articles in the field of CHI research came from the 2004-2008 and 2009-2014 study period (Table 1). The paper “eHealth literacy: Essential Skills for Consumer Health in a Networked World” was published in the *Journal of Medical Internet Research* in 2006. In this article, the concept of eHealth literacy was defined for the first time. The ability to find, discover, understand, and evaluate health information from electronic resources, and the ability to apply the acquired knowledge to solve health problems were also explained in this paper [17]. Ranking fourth in the number of citations, the paper was cited 514 times. eHealth literacy has become a mature research topic, and increasingly more scholars are paying attention to it. For example, Kim et al [18] examined the association among eHealth literacy, perceived benefits, self-efficacy, and health-promoting behaviors in patients with type 2 diabetes. Cherid et al [19] investigated the level of mobile technology acceptance, health literacy, and electronic health literacy of 401 patients aged over 50 years with recent fractures.

**Table 1.** Top 10 cited articles in the consumer health informatics research field.

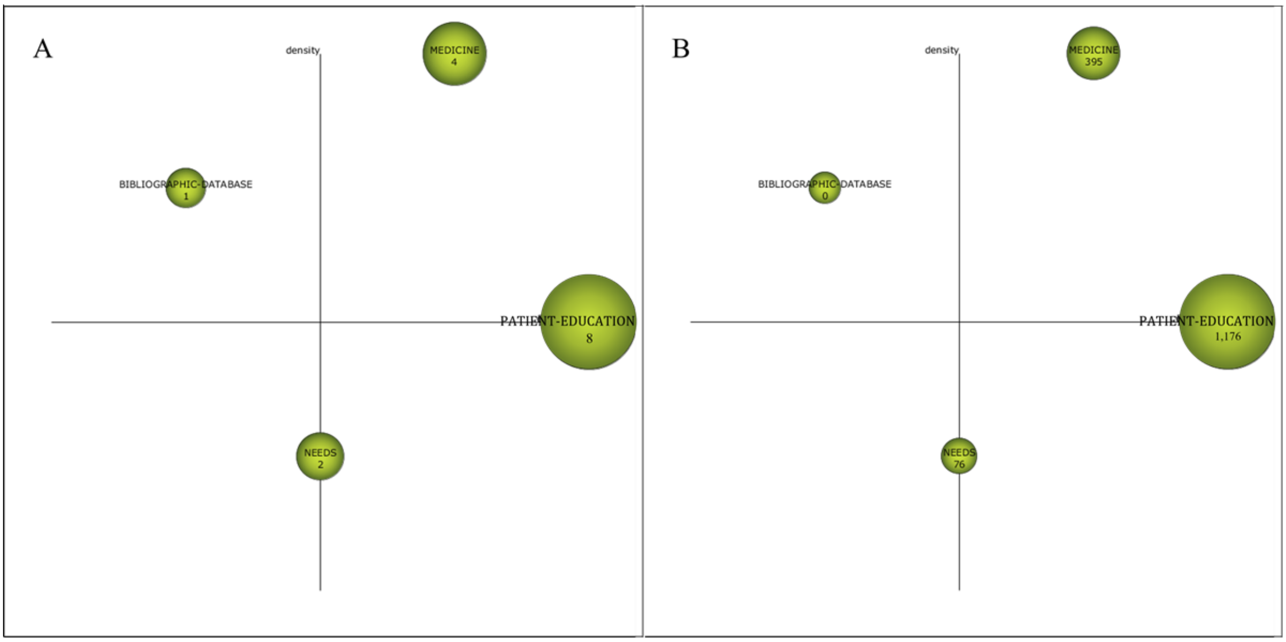
| Reference | Journal | Year | Citations |
|---|---|---|---|
| Eysenbach et al [12] | *Journal of the American Medical Association* | 2002 | 940 |
| Cline and Haynes [13] | *Health Education Research* | 2001 | 804 |
| Charnock et al [14] | *Journal of Epidemiology and Community Health* | 1999 | 521 |
| Norman and Skinner [17] | *Journal of Medical Internet Research* | 2006 | 514 |
| Eysenbach [3] | *British Medical Journal* | 2000 | 336 |
| Gustafson et al [20] | *American Journal of Preventive Medicine* | 1999 | 320 |
| Or and Karsh [21] | *Journal of the American Medical Informatics Association* | 2009 | 247 |
| van den Berg et al [22] | *Journal of Medical Internet Research* | 2007 | 236 |
| Dutta-Bergman [23] | *Health Communication* | 2004 | 229 |
| Hibbard and Peters [24] | *Annual Review of Public Health* | 2003 | 227 |

## Evolutionary Characteristics Analysis

### *Design and Overview*

The strategic map is mainly used to describe the relationship between themes and the relationship strength within the theme to reflect the importance of the theme in the development of the whole field and the development of the theme. The node in Figure 4 is a cluster, and the sizes of the nodes indicate the sizes of the theme clusters. The horizontal axis is the centrality, which measures the relevance of the theme to other themes. With higher centrality, the theme is more important in the whole field. The vertical axis is the density, which measures the relational strength of the cluster keywords within a theme. The higher the density of the theme, the more mature the theme. The strategic map is divided into 4 quadrants. Quadrant 1 (Q1) contains motor themes, quadrant 2 (Q2) contains highly developed and isolated themes, quadrant 3 (Q3) contains emerging or declining themes, and quadrant 4 (Q4) contains basic and transversal themes. The strategic map of the 4 study periods, the relevant bibliometrics indicators, and the evolutionary status of the research themes are as follows.

**Figure 4.** Strategic diagram of themes for the subperiod of 1999-2003 (A: based on the publications, B: based on the citation frequency).



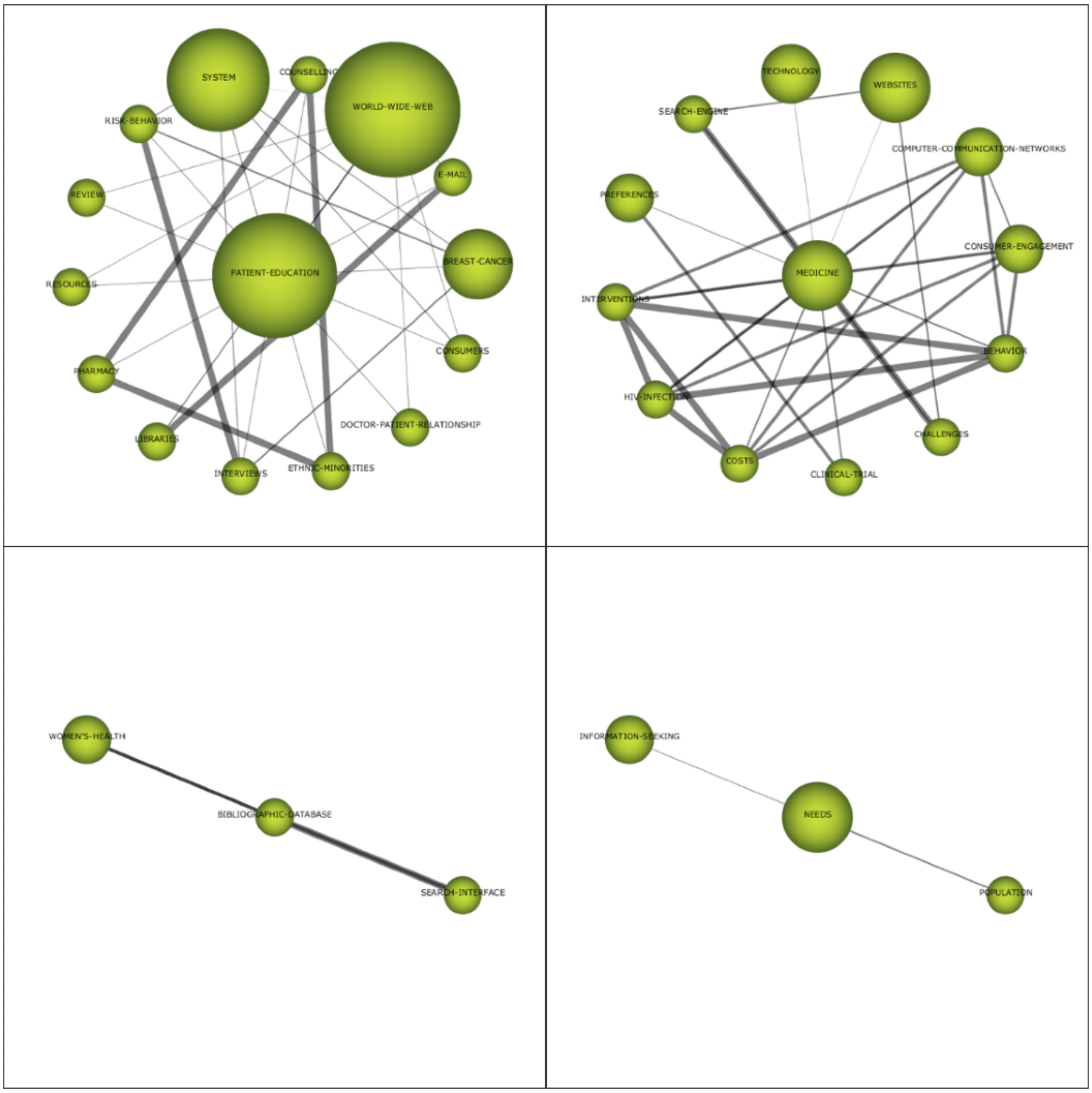
### *Themes (n=4) in 1999-2003: Patient Education, Medicine, Needs, and Bibliographic Database*

The theme of Q1 is *medicine*, which has the highest density value and a higher centrality (Table 2 and Figure 4). The keywords closely connected to *medicine* are *search engine*, *interventions*, and *challenges* (Figure 5). The websites and pages displayed by search engines have become the main source from which people obtain medical and health information [25]. Health information/support systems for health promotion are also emerging, such as the CHESS comprehensive health promotion support system [20].

**Table 2.** Performance measures for the themes of the subperiod 1999-2003.

| Theme | Centrality | Density | Number of documents | H-index | Number of citations |
|---|---|---|---|---|---|
| *Patient education* | 45.99 | 56.26 | 8 | 7 | 1176 |
| *Medicine* | 36.24 | 111.7 | 4 | 3 | 395 |
| *Needs* | 1.85 | 12.5 | 2 | 2 | 76 |
| *Bibliographic database* | 1.25 | 66.67 | 1 | 0 | 0 |

**Figure 5.** Cluster network of the themes (1999-2003).



The theme of Q2 is *bibliographic database*, which has centrality and density values of 1.25 and 66.67, respectively. The theme is closely related internally and has low relevance with other themes. The amount of relevant literature is the least, as the theme has not attracted the attention of the academic circle. The keywords closely connected to *bibliographic database* are *search interface* and *women's health*. For the prototype network interface of the women's health bibliography database, Marton et al [26] performed a comparative analysis from 3 aspects: web interface design, overall visual effect, and hypertext navigation and information organization.

*Patient education* is located on the centrality axis and has the highest centrality. *Patient education* has the largest amount of research literature and has been cited the most times, indicating that it is an area of high concern and has had a great impact on later research. *Patient education* is interconnected with many different keywords, such as world wide web and system, but the connection is not strong. The result of the cluster network shows that the internet and information systems are important tools and approaches in patient education [13] and that individualized customized education can be carried out for patients through computers [27].

*Needs* is located on the density axis and has centrality and density values of 1.85 and 12.5, respectively. The cluster network shows that *needs* is connected with *population* and *information seeking*. The internal demand for health information drives millions of consumers to search for health information on the internet. If the search terms used by consumers do not match the terms set at the information source, the search results will not meet the needs of consumers. Therefore, consumer health information retrieval needs a full range of terminology support [28].

### *Themes (n=9) in 2004-2008: World Wide Web, Patient Education, eHealth, Patients, Medication, Terminology, Behavior, Technology, and Disease*

The themes of Q1 have higher centrality and density (Table 3 and Figure 6) and are *world wide web*, *patient education*, *eHealth*, and *medication*. *World wide web* has the largest amount of research literature and has been cited 635 times. The cluster network shows that the relationship between the *world wide web* and internal keywords is weak. The keywords that are strongly connected to *patient education* are *internet health information*, *record*, and *challenges* (Figure 7). Patient education is an important part of providing health care services and helps to improve the effect of medical care. Doupi and van der Lei [29] discussed the possibility of integrating electronic medical record data and online health information resources to provide personalized patient education for patients. *eHealth* has the highest density and a relatively high centrality; the cluster network reveals that it is closely related to *health communication*, *tailored intervention*, *cell phone*, and *teleconsultation*, which is an important research topic in this field. Tufano and Karras [30] used teleconsultation for an obesity intervention that was expected to achieve mass customization function, interactive function, and national customized electronic health information. The location of *medication* is far from the centrality axis and close to the density axis. The cluster network shows that it has a strong relationship with *health literacy* and *pharmacist*. In the research on drug treatment in the field of CHI, scholars are concerned about the health literacy of consumers and whether they can understand the accompanying instructions of drugs [31]. Webb et al [32] have proposed that the design of patient-centered consumer medication information can improve the comprehensibility of warning labels.

**Table 3.** Performance measures for the themes of the subperiod of 2004-2008

| Theme | Centrality | Density | Number of documents | H-index | Number of citations |
|---|---|---|---|---|---|
| *World wide web* | 77.3 | 28.66 | 14 | 11 | 635 |
| *Patient education* | 64.37 | 26.23 | 10 | 8 | 786 |
| *eHealth* | 41.52 | 87.63 | 3 | 3 | 129 |
| *Patients* | 33.26 | 16.05 | 6 | 5 | 439 |
| *Medication* | 31.25 | 32.5 | 4 | 4 | 80 |
| *Terminology* | 3.85 | 16.67 | 2 | 2 | 45 |
| *Behavior* | 2.38 | 17.59 | 3 | 3 | 77 |
| *Technology* | 18.49 | 41.2 | 2 | 2 | 238 |
| *Disease* | 16.39 | 8.33 | 2 | 2 | 48 |

**Figure 6.** Strategic diagram of themes for the subperiod of 2004-2008 (A: based on the publications, B: based on the citation frequency).

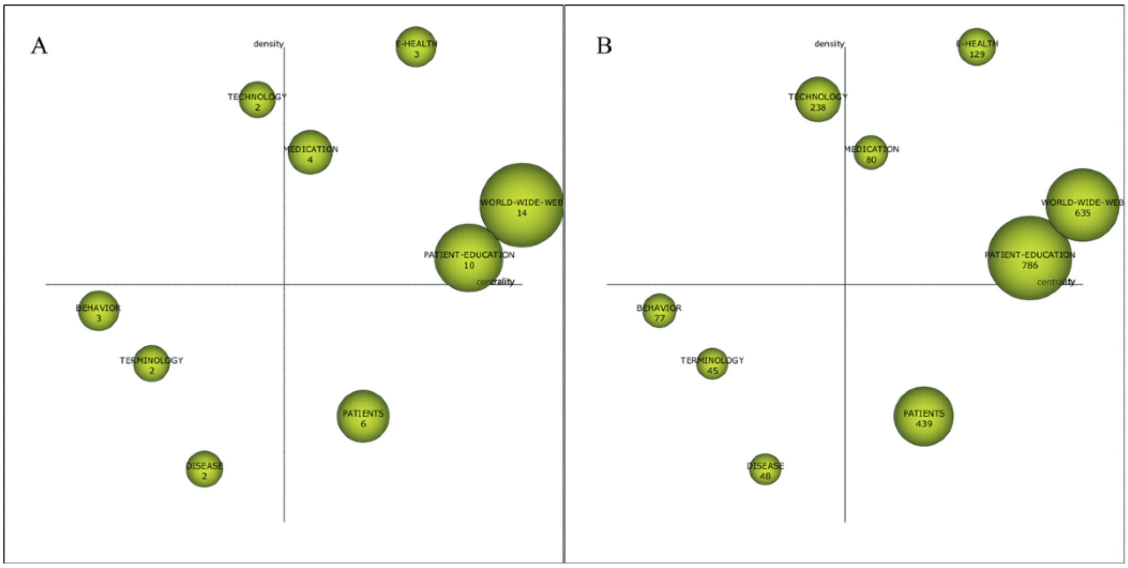
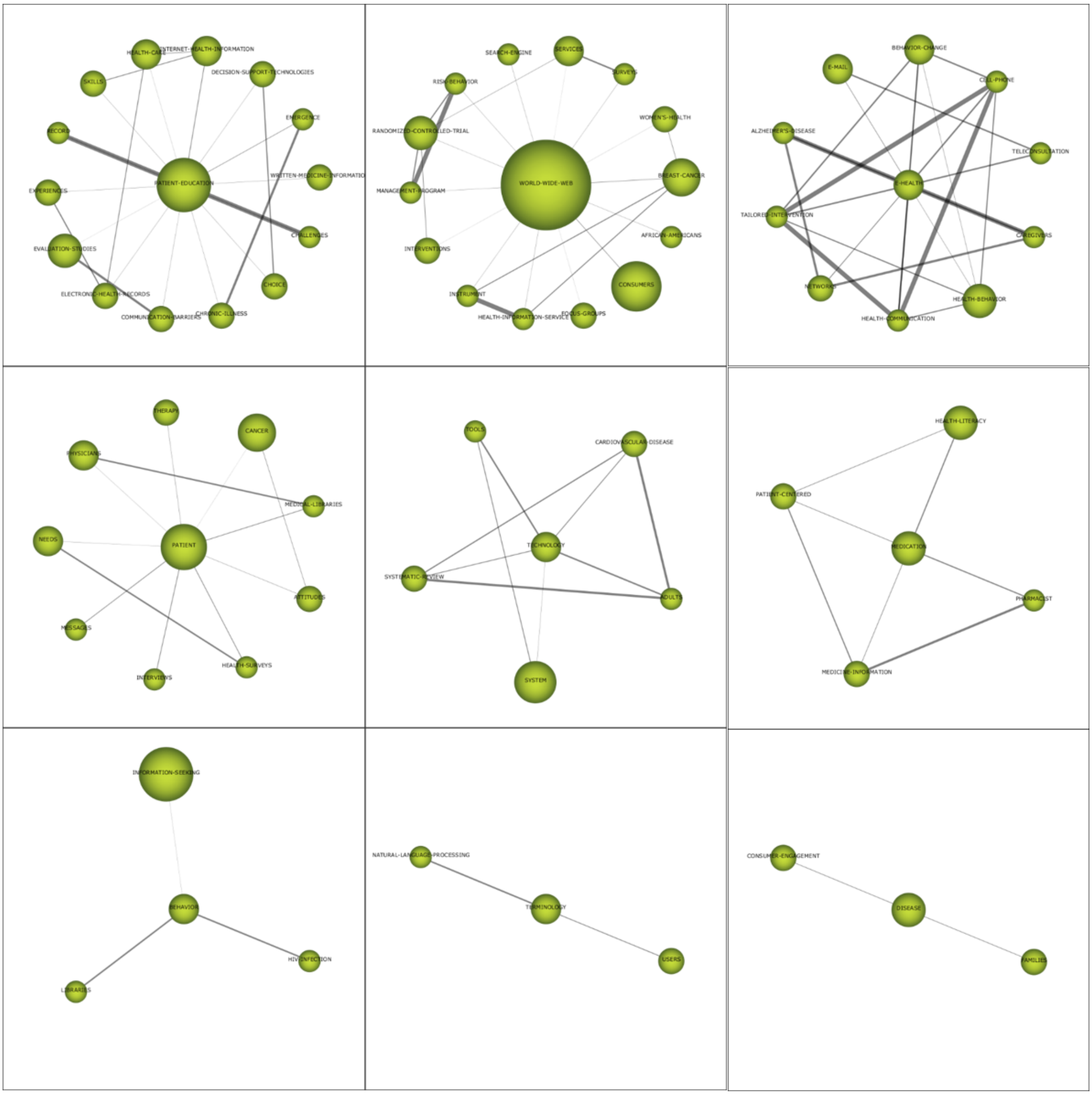


**Figure 7.** Cluster network of the themes (2004-2008).



The theme of Q2 is *technology*, with centrality and density values of 18.49 and 41.2, respectively. The important nodes that are connected internally are *tools* and *adults*. The literature corresponding to cluster nodes reveals that scholars began using internet technology to intervene in people's health activities [22].

The themes of Q3 are *terminology*, *behavior*, and *disease*, which have centralities of 3.85, 2.38 and 16.39, respectively, and density values of 16.67, 17.59, and 8.33, respectively. *Terminology* is only associated with *natural language processing* and *users*. The literature corresponding to *terminology* reveals that the rapid development of consumer health education websites and other applications has promoted research on consumer health vocabulary and that term *recognition* [33] is one of the important research directions. *Disease* has weak internal relations. *Behavior* is associated with *libraries*, *HIV infection*, and *information seeking*. By studying the literature corresponding to *behavior*, we found that researchers had analyzed health information searching behavior according to different populations. For example, Hesse et al [34] analyzed the 2005 Administration of the Health Information National Trends Survey data to explore the information searching behavior of survivors of cancer and found that their information searching behavior was very common and would not decrease with time.

The theme of Q4 is *patients*, with centrality and density values of 33.26 and 16.05, respectively. The internal correlation of the *patients* theme cluster was not strong. An examination of the literature corresponding to cluster nodes revealed that scholars paid more attention to the research of the patients' attitude, cognition, and intention regarding health information [35].

### *Themes (n=11) in 2009-2014: Websites, Information Seeking, Physicians, Attitudes, Technology, Risk, Food*

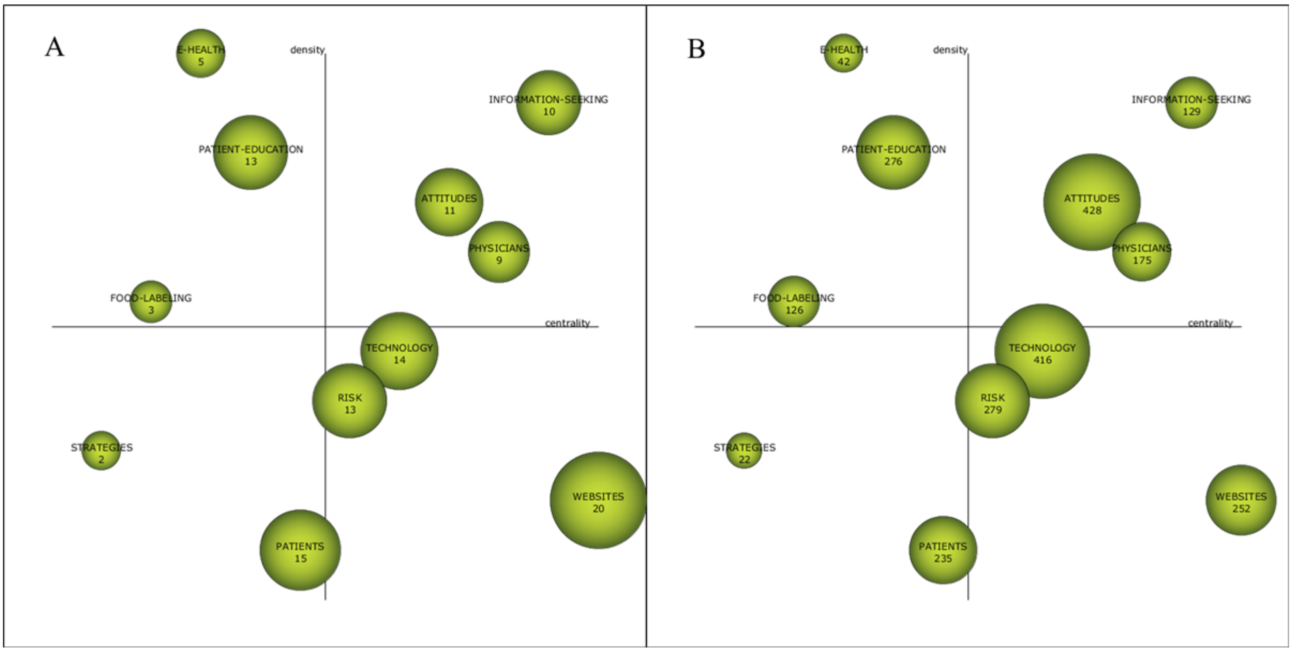
### *Labeling, Patient, Strategies, Patient Education, and eHealth*

The themes of Q1 are *information seeking*, *physicians*, and *attitudes*, which have centralities of 69.64, 67.77, and 64.83, respectively, and density values of 37.42, 19.27, and 28.02, respectively (Table 4 and Figure 8). The cluster network (Multimedia Appendix 1) shows that there are 15 nodes and 25 links in *information seeking* and that the keywords closely related to *information seeking* are *search interface*, *management program*, and *community*. The literature corresponding to *information seeking* revealed that scholars analyzed, from the perspective of consumers, the content of online questions, explored the influencing factors of health information searching, and designed a new connected exploratory navigation interface to improve the effectiveness of health information searching [36]. *Attitudes* had the highest citation frequency, indicating that it had a great influence on the subsequent research. The cluster network shows that *attitudes* is associated with a large number of different keywords, but the relationship is not strong; among these different keywords, the closely related nodes are *community pharmacies*, *information and communication techniques*, and *electronic medical records*. The cluster network reveals that the nodes that are close to *physicians* are *interviews*, *patient centered care*, *patient–provider relationship*, and *teleconsultation*. The literature corresponding to *attitudes* and *physicians* reveals that the consumers' attitude toward, access to, and use of online information related to diseases [37], medicines, diet, and other health issues are the research hotspots in this period. For example, Amicizia et al [38] believes that information and communications technology provide an opportunity for health care workers to use mobile internet to disseminate vaccine-related knowledge interactively and entertainingly and to monitor adolescent attitudes toward vaccination through social media.

**Table 4.** Performance measures for the themes of the subperiod of 2009-2014.

| Theme | Centrality | Density | Number of documents | H-index | Number of citations |
|---|---|---|---|---|---|
| *Websites* | 77.12 | 8.45 | 20 | 9 | 252 |
| *Information seeking* | 69.64 | 37.42 | 10 | 7 | 129 |
| *Physicians* | 67.77 | 19.27 | 9 | 6 | 175 |
| *Attitudes* | 64.83 | 28.02 | 11 | 10 | 428 |
| *Technology* | 61.65 | 15.53 | 14 | 10 | 416 |
| *Risk* | 60.43 | 12.61 | 13 | 10 | 279 |
| *Food labeling* | 6.56 | 15.62 | 3 | 3 | 126 |
| *Patient* | 59.71 | 6.02 | 15 | 10 | 235 |
| *Strategies* | 5.5 | 11.11 | 2 | 2 | 22 |
| *Patient education* | 40.51 | 31.32 | 13 | 9 | 276 |
| *eHealth* | 30.79 | 37.44 | 5 | 3 | 42 |

**Figure 8.** Strategic diagram of themes for the subperiod of 2009-2014 (A: based on the publications, B: based on the citation frequency).



The themes of Q2 are *patient education*, *eHealth*, and *food labeling*, which have centralities of 40.51, 30.79, and 6.56, respectively, and density values of 31.32, 37.44, and 15.62, respectively. The cluster network shows that *patient education* is closely related to *smartphone*, *pamphlet*, *resources*, and *Alzheimer's disease*. Scholars mainly conducted research from 3 aspects: patients' health education needs [39], the effect of health education materials, and health information technology. For example, Kraschnewski et al [40] studied how pregnant women use the internet and mobile phone technology to acquire health care knowledge. They found that women thought that the educational materials they received during prenatal care were not helpful, and therefore, they turned to the internet and smartphone apps to fill the knowledge gap [40]. There are 9 nodes and 12 connections in the *eHealth* theme. Compared with the previous study period, in this study period, there are fewer nodes and connections. The research content of the theme is more concentrated. The keywords closely related to *eHealth* are *tailored intervention*, and *online health information*. Scholars are concerned about how to use information and communication technology to help patients improve their health status and strengthen their self-health management skills [41]. For example, as a health literacy intervention for Hispanic patients with AIDS, Jacobs et al [42] developed a Spanish-language computer application, which can provide information about HIV infection, treatment, and drugs. *Food labeling* is a new theme, for which the cluster network shows the related keywords are *consumer opinions*, *recommendations*, and *ratings*. King et al [43] investigated the consumers' understanding of common terms used to guide food consumption frequency and quantity and found that some terms are highly subjective and that more simple and clear terms need to be developed.

The themes of Q3 are *patients* and *strategies*, which have lower centralities and density values. The cluster network reveals that the keywords strongly related to *patients* were *health information service*, *information behavior*, and *conflict of interest*. *Strategies* is an emerging theme. The cluster network shows that the keywords associated with it are only *evidence-based practice* and *task* and that the weight of the connection is 0.17.

Having higher centralities and lower density values, the themes of Q4 are *websites*, *technology*, and *risk*. The cluster network shows that the keywords closely related to *websites* are *world wide web*, *evaluation studies*, *tools*, *email*. By studying the literature corresponding to *websites*, it is found that the research on the content evaluation of websites and health information intervention through websites are the research hotspots in this period [44]. Thakor et al [45] investigated the information quality of ecommerce websites selling *Hypericum perforatum* and found that most sites received poor reviews and lacked information on drug interactions, contraindications, and adverse reactions. The keywords closely related to *technology* are *hometelehealth* and *aging*. With the aggravation of the aging of the population, the research on home telehealth for the aged group has attracted much attention. Cimperman et al [46] studied the important factors affecting the elderly's adoption of home telehealth services and found that the following factors play an important role in the perception of home telehealth: perceived usefulness, expectation, social influence, perceived security, computer anxiety, convenience, and the doctor's opinion [46,47]. *Risk* is split from *medication*. The cluster network shows that the relationship between internal keywords is weak and that the relatively strong ones are *costs*, *patient decision aid*, and *cardiovascular disease*. For veterans with and without multiple sclerosis, Cameron et al [48] studied the relative risk of falling and requiring medical care and found that the adjusted odds ratio of falling was 3 times higher for female veterans with multiple sclerosis than for female veterans without multiple sclerosis.

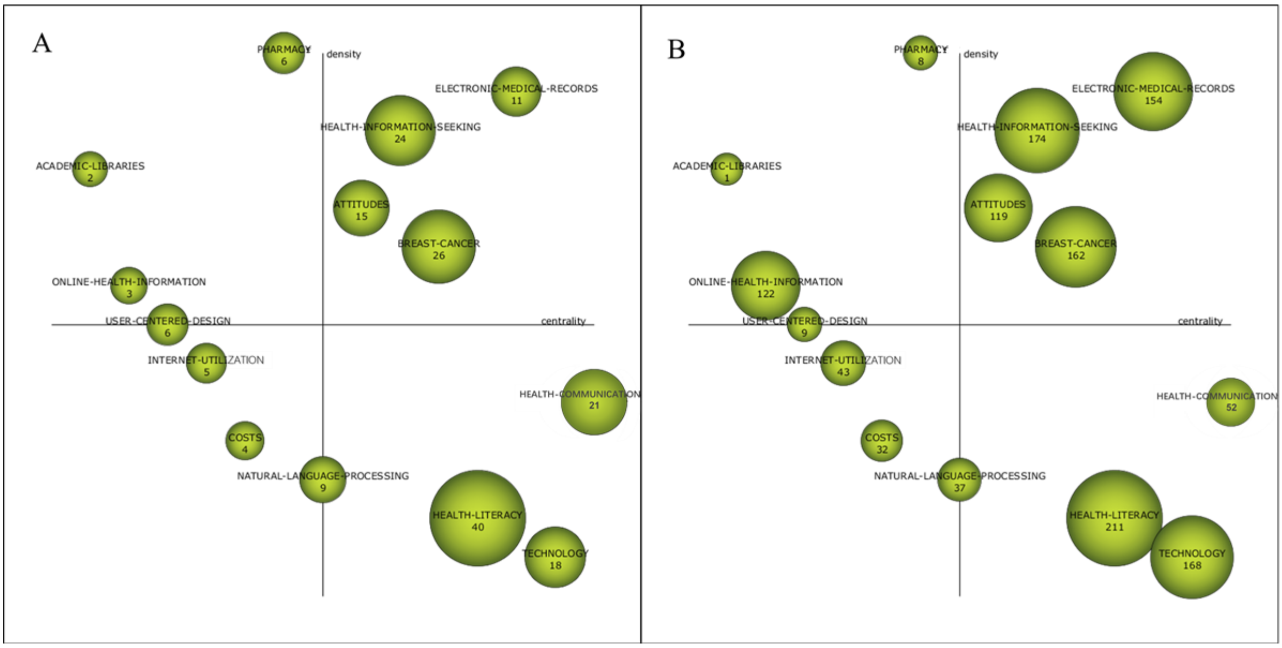
### Themes (n=14) in 2015-2019: Electronic Medical Records, Health Information Seeking, Attitudes, Health Communication, Breast Cancer, Health Literacy, Technology, Natural Language Processing, User-Centered Design, Pharmacy, Academic Libraries, Costs, Internet Utilization, and Online Health Information

The themes of Q1 are *electronic medical records*, *health information seeking*, *attitudes*, and *breast cancer*, with centralities of 53.52, 47.78, 45.18, and 48.08, respectively, and density values of 19.71, 18.14, 8.89, and 7.11, respectively (Table 5 and Figure 9). The cluster network (Multimedia Appendix 2) shows that *electronic medical records* has a strong relationship with internal keywords, which include *personal health information management* and *care partner*. Scholars mainly focus on the patients' health information management, including the sharing of access to the patients' health records, personal health data visualization, and other contents. For example, Wolff et al [49] sent the doctor's medical records to patients and authorized nurses through OpenNotes. The scholars verified the acceptability and effect of this approach, and the results showed that it was acceptable for patients and their nursing staff to view the doctors' medical records; the results also reflected that this method improved the communication with patients and enhanced the patients' confidence in cooperating with nurses. Keywords closely related to *healthinformation seeking* are *users*, *behavior*, *health information needs*, and *consumer health information behavior*. An examination of the literature corresponding to important nodes revealed that the users of library, online health consultation platform, social media, and other media were often considered research objects and that a discussion of these users' health information needs and behaviors was a research hotspot in this period [50,51]. For example, taking the online health consultation case on "Taiwan eDoctor," an online health consultation platform, as the research object, Chiu et al [52] studied the length, time, communication mode, purpose, and identity disclosure of online consultation questions and described the communication mode of patients in the process of health information searching. *Attitudes* was closely related to internal keywords, such as *HPV vaccines*, *risk*, *evaluation studies*, and *physicians*. Systematically studying the rankings, quality, and contents of the web pages related to HPV vaccines, Fu et al [53] classified the included web pages into critical and noncritical ones and found that the quality of the critical web pages was poor but that the critical web pages often obtained higher rankings. Keywords closely related to *breast cancer* are *online* *healthcommunities*, *support groups*, and *preferences*. The scholars mainly focused on the content analysis of the online health community [54] and its supporting role for patients with cancer. For example, through interviews and questionnaires, Huh et al [55] developed online health community roles to reflect the users' needs and requirements when using an online health community. These roles can help provide users with customized social support and patient support. The study found that the roles of online health communities can be divided into 4 categories: managers, opportunists, scientists, and adventurers. These roles reveal user interaction behavior and attitude patterns when using online health communities.

**Table 5.** Performance measures for the themes of the subperiod 2015-2019.

| Theme | Centrality | Density | Number of documents | H-index | Number of citations |
|---|---|---|---|---|---|
| *Electronic medical records* | 53.52 | 19.71 | 11 | 7 | 154 |
| *Health information seeking* | 47.78 | 18.14 | 24 | 5 | 174 |
| *Attitudes* | 45.18 | 8.89 | 15 | 7 | 119 |
| *Health communication* | 57.98 | 4.7 | 21 | 4 | 52 |
| *Breast cancer* | 48.08 | 7.11 | 26 | 7 | 162 |
| *Health literacy* | 50.29 | 3.55 | 40 | 8 | 211 |
| *Technology* | 54.79 | 3.46 | 18 | 7 | 168 |
| *Natural language processing* | 24.1 | 4.18 | 9 | 4 | 37 |
| *User-centered design* | 11.9 | 6.3 | 6 | 2 | 9 |
| *Pharmacy* | 16.52 | 28.49 | 6 | 2 | 8 |
| *Academic libraries* | 4.52 | 9.03 | 2 | 1 | 1 |
| *Costs* | 14.08 | 4.34 | 4 | 4 | 32 |
| *Internet utilization* | 12.67 | 4.72 | 5 | 3 | 43 |
| *Online health information* | 7.82 | 6.49 | 3 | 2 | 122 |

**Figure 9.** Strategic diagram of themes for the subperiod of 2015-2019 (A: based on the publications, B: based on the citation frequency).



The themes of Q2 are *pharmacy*, *online health information*, and *academic libraries*, with centralities of 16.52, 4.52, and 7.82, respectively, and density values of 28.49, 9.03, and 6.49, respectively. A mature professional theme (*pharmacy*) has the highest density and is close to the density axis. The keyword most closely related to *pharmacy* is *patient medication information*. Monkman and Kushniruk [56] studied consumer medication information in pharmacies and found that the organization and presentation of online consumer medication information need to be improved. The content of consumer medication information needs to be improved in order to promote the safety and effective use of drugs [57]. The keywords in the *online health information* cluster network include *video*, *web* 2.0, and *world wide web*. Scholars mainly pay close attention to the use and quality evaluation of online health information [58]. As an emerging theme, *academic libraries* has the lowest centrality, the lowest amount of relevant literature, and the lowest citation frequency.

*User-centered design*, with centrality and density values of 11.9 and 6.3, respectively, is located on the centrality axis. One of the keywords clustering with *user-centered design* is *smartphone apps*. An examination of the key literature corresponding to cluster networks reveals that scholars focus on designing user-centered, smartphone-based mobile health apps, such as nutrition education apps that, for example, provide technical support for adolescents with overweight and obesity [59].

Having lower centralities and density values, the themes of Q3 are *internet utilization* and *costs*. The internal relationship between *internet utilization* and *costs* is weak; both are in an immature development stage and have little impact. The cluster network shows that the keywords in *costs* include *search engine* and *health records*; the keywords in *internet utilization* include *health behavior*, *systematic review*, and *patient education as topic*. The literature corresponding to *internet utilization* reveals that the content of relevant literature involved patients using the internet to search for health information for self-care. For example, Jamal et al [60] studied the online health information searching behavior of patients with type 2 diabetes in the Middle East and its influence on their self-care behavior. The results showed that most internet health information searchers had positive changes in their behaviors after searching and had a stronger awareness of diabetes self-care [60].

*Natural language processing* is located on the density axis, with centrality and density values of 24.1 and 4.18, respectively. The cluster network shows that *natural language processing* is connected with *consumer health vocabulary*, *state*, *online support groups*, and *text*. An examination of the core literature corresponding to the cluster shows that most scholars mined and dealt with the health-related contents of consumers through social media, health websites, health forums, and other ways to obtain the users' health needs, explore the users' health behaviors, and investigate the application of health information technology [61,62].

The themes of Q4 are *health communication*, *health literacy*, and *technology*, with centralities of 57.98, 50.29, and 54.79, respectively, and density values of 4.7, 3.55, and 3.46, respectively. The *technology* cluster network formed 15 nodes and 21 wires, but the internal relationship was very weak. *Health communication* has the highest centrality and has many connections with external themes. It is a key theme with development potential. The cluster network shows that *health communication* is strongly related to keywords such as *medicine*, *telehealth*, and *perspectives*. Benis et al [63] studied the use of communication channels between patients and health care organizations. By matching the communication channels with the patients' personal information, the communication between medical institutions and patients can be transformed into a more active mode, and the patient's participation can be improved. *Health literacy* is an emerging theme; it has the largest number of related literature and the largest citation frequency sum. The keywords closely related to *health literacy* include *skills*, *patient education*, and *public libraries*. Health literacy is the ability of individuals to access and understand health information and to use it to maintain and promote their health. An examination of the literature of important nodes reveals that scholars pay more attention to the assessment and intervention of health literacy [64] and the assessment of internet health information quality [65,66].

## Evolutionary Path Analysis

### *Design and Overview*

The theme evolution map shows, through a data flow, the evolution of the 4 study period themes of CHI research. It can analyze and track the dynamic evolution of the themes in the field of CHI research over a period. Figure 10 shows the CHI research field's theme evolution map. Each column in the figure represents a study period. In the figure, the nodes represent the research themes, and the size of the nodes is proportional to the number of related articles of the research themes. The connection between the nodes represents the theme data flow, the solid line represents the main keywords shared by 2 themes, and the dotted line represents the shared keywords are not the main keywords. The width of the connection is proportional to the Inclusion index [11].

From Figure 10, it can be seen intuitively that *patient education* and *technology* appear in 3 study periods, and that *eHealth*, *attitudes*, and *patients* appear in 2 study periods. With time, the circle in the theme evolution map becomes larger, the number of the research themes increases, and the data flow between the research themes becomes increasingly complex. This reflects the emergence of new research themes and contents from 1999 to 2019. According to the theme evolution map, the size of the research theme circle, and the data flow between themes, 10 theme evolution paths of 3 research directions in the field of CHI were determined.

**Figure 10.** Thematic evaluation map of the consumer health informatics research field (1999-2019).

### *Supply-Side Research of Consumer Health Informatics: Patient Education and Intervention Research*

Consumer health education is the process of assisting one to acquire the correct information and understanding so that one will be able to make wise decisions about a certain health item [67]. *Patient education* → *patient education* → *patient education* → *health literacy*, *pharmacy* (subpath 1). *Patient education* → *world wide web* → *websites* → *health communication* (subpath 2). *Patient education* → *eHealth* → *eHealth* → *health literacy*, *breast cancer*, *health communication* (subpath 3).

The 3 subpaths above evolved from the main and branch directions of the *patient education* theme in the first study period. Over time, the nodes on the evolutionary path become larger, the number of research articles increases, and the data flow from node splitting is greater. The data flow of patient education and intervention research evolution is relatively clear and represents the main path of the evolution theme of CHI. The theme on the path mainly moves between the first quadrant and the fourth quadrant, and the split and fusion of themes are obvious. In the second study period, subpaths 1 and 3 became the focus of research. In the third study period, the theme density increased and the centrality decreased, demonstrating a professional and mature research direction in the field. In the fourth study period, the themes split into several themes with great influence and development potential (*health communication* and *health literacy*). In the third study period, the subpath 2 integrated multiple themes, and the density of themes decreased. In the fourth study period, the theme split into several low-density but high centrality themes (*natural language processing*).

### *Consumer-Side Research of Consumer Health Informatics: Consumer Demand, Attitude, and Behavior*

The evolutionary path of consumer demand, attitude, and behavior research consists of 3 subparts. *Needs → patients → attitudes, physicians → attitudes, health literacy, healthinformation seeking, internet utilization* (subpath 1); *medicine, needs → behavior → information seeking → healthinformation seeking, health literacy* (subpath 2); and *medicine → diseases → patients → electronic medical records* (subpath 3).

These 3 subpaths evolved from the main and branch directions of *medicine* and *needs* in the first study period. With time, the number of nodes on the evolutionary path increases, the number of research articles increases year by year, and the data flow of node splitting also increases. The themes in the evolutionary path of consumer demand, attitude, and behavior research mainly move between the first, third, and fourth quadrants, and the themes are divided and integrated. The theme of subpath 1 has gone through a stage in which centrality and density rise, and in this field, it finally forms many core themes, such as *healthinformation seeking*, *health literacy*, and *attitudes*. In the second study period, subpath 2 splits to form a new theme (*behavior*). Later, this theme attracted the attention of many scholars in the field; its density and centrality increased rapidly, and it became the core theme in the field. In the second and third study periods, subpath 3 is an edge theme that is not of high concern. In the fourth study period, some themes (*attitudes*, *physicians*, *technology*, *patients*, *websites*) split and merge into a high centrality and density key theme (*electronic medical records*).

### *The Technology Research of Consumer Health Informatics: The Application Research of Internet and Information Technology*

Currently, big data, internet plus, artificial intelligence, and other emerging information technologies have been tightly integrated with traditional medical and health industries. The application of internet and information technology in the field of CHI has attracted more attention from computer science and communication science scholars. The evolution path of internet and information technology application research consists of 4 subpaths. *Medicine → technology → technology → technology, health communication* (subpath 1); *patient education → eHealth → eHealth → online health information* (subpath 2); *needs, bibliographic database → patients, world wide web → information seeking → natural language processing, user-centered design, academic libraries* (subpath 3); and *medicine → world wide web, terminology → websites → online health information, natural language processing, breast cancer* (subpath 4).

These 4 subpaths evolved from the main directions of *bibliographic database* in the first study period as well as the branch directions of *patient education* and *medicine*. The number of research articles is gradually increasing, and the phenomenon of node splitting and fusion is also obvious. The theme of the evolution path of the internet and information technology application research mainly moves between the first, second, and fourth quadrants. The evolution path is relatively complex, which may be related to the application of information technology in different research directions. Subpath 1 emerged from the theme of the first study period, and the density decreased in the second study period. In the third and fourth study periods, the centrality increased. The density decreased after the split of multiple themes, and some important themes with strong influence and development potential developed. Subpath 2 divides from the theme of *patient education* in the first study period. In the second study period, a key theme with high centrality and density was formed. In the third and fourth study periods, the centrality decreased, and the density increased; it became a professional theme in the field. The data flow of subpaths 3 and 4 is relatively complex and is formed by the convergence of 2 routes in the first and second study period.

## Discussion

### Overall Development Status of Consumer Health Informatics

In the past 21 years, CHI research has been on the rise and has gone through 3 stages: slow growth, stable growth, and rapid growth. In the early stage, the development was slow, the themes were few, and the intensity was low. In the later stages, the development became diversified. After 2015, the number of research papers on CHI increased substantially, and it became a hotspot of academic research. In 2019, it reached a peak of 112 papers, a research volume accounting for 12.03% of the total literature.

The period 1999-2003 comprises the basic research stage of consumer health information. Six of the top 10 highly cited literature in this field come from this study period. This literature constitutes the cornerstone of CHI research. Consumer health information evaluation tools, such as DISCERN, that were developed in this study period have become recognized mature tools in the field. The period 2004-2014 was one of steady growth in the research on CHI. During this period, fewer keywords disappeared, and more new ones were absorbed. In 2015-2019, the number of keywords was the largest, and the research content involved was increasingly more abundant. A similarity index for the fourth study period reached a higher level, reflecting that the research field of CHI had become more mature, the inheritance of research had become stronger, and researchers were continuing to focus on more research topics.

### Evolutionary Characteristics Analysis

By analyzing the rise, decline, and the change in the density and centrality of the themes in different study periods, the evolution characteristics and status of the research themes can be reflected.

The motor themes include the following: *medicine* and *patient education* in 1999-2003; *world wide web*, *patient education*, *eHealth*, and *medication* in 2004-2008; *information seeking*, *physicians*, and *attitudes* in 2009-2014; and *electronic medical records*, *healthinformation seeking*, *attitudes*, and *breast cancer* in 2015-2019. These themes have high centralities and density values. Besides, the research studies are popular and influential, and the internal relations within these themes are closely relevant. These themes are the core of the research field and their development is mature.

The highly developed and isolated themes include the following: *bibliographic database* in 1999-2003; *technology* in 2004-2008; *eHealth*, *patient education*, and *food labeling* in 2009-2014; and *pharmacy*, *academic libraries*, *online health information*, and *user-centered design* in 2015-2019. Having high-density values and low centralities, these themes do not represent the research center. They are mature and peripheral themes in this field.

The emerging or declining themes include the following: *behavior*, *terminology*, and *disease* in 2004-2008; *patient* and *strategies* in 2009-2014; and *costs* and *internet utilization* in 2015-2019. Because of their low centralities and density values, these themes have been weakly developed and are marginalized themes in this field.

The basic and transversal themes include the following: *needs* in 1999-2003; *patients* in 2004-2008; *websites*, *technology*, and *risk* in 2009-2014; and *health communication*, *breast cancer*, *health literacy*, *technology*, and *natural language processing* in 2015-2019. These themes have high centralities and low-density values, indicating that these have a high influence on the field of CHI but are not well developed.

By analyzing the evolutionary status of the themes in the 2015-2019 study period, the paper expects to predict the future development trend of the field. The centrality and density value of *electronic medical records* are extremely high. Although there are not many related documents, each document has been cited 14 times on average, indicating that it has received extremely high attention and may appear in the next research period. *Healthinformation seeking* is relatively stable because of its high centrality and density value. After an internal keyword splitting and recombination, the *attitudes* theme is still high in centrality and density value, and the research enthusiasm has not decreased. It will continue to become a hot topic in the field of CHI research. *Breast cancer* split from *risk* in the previous study period, with increased centrality and density value (transferred from Q4 to Q1). With the expansion of research influence, it became more mature and eventually became the research core in the field. Although the number of studies related to *online health information* is not large, the average citation frequency of each study is the highest, indicating that it has a great influence on future research. Online health information is one of the foundations of the CHI research field, has great development potential, and may continue to appear in the next study period. *Natural language processing* has been a hot topic in recent years, and it is likely to continue to appear in the next study period. *Health literacy* is an emerging topic, with the largest number of relevant studies and total cited frequency. This indicates that it has a high level of attention and has a strong evolution and development ability. It will continue to be a hot topic in this field. *Technology* has been transferred from Q2 to Q4. Its research influence and development potential have improved, reflecting that it will continue to be the basic theme in this field.

## Evolutionary Path Analysis

By analyzing the evolution characteristics of patient education and intervention research themes, it can be found that the research in the slow growth period mainly focuses on the application and exploration of the internet and information systems and explores the health education for patients through computers and websites [13]. The research in the stable growth period mainly focuses on the use of network health information [41], the design of personalized or customized patient education [29], and the application of intelligent and mobile devices [30]. The research in the rapid growth period mainly focuses on the role and mechanism of the online health community in patient support [54], doctor–patient communication [63], and the quality evaluation of online health information [66].

By analyzing the evolution characteristics of consumer demand, attitude, and behavior research themes, it can be found that the research in the slow growth period mainly focuses on the users' demand for online health information search, as well as demand-driven behavioral research, such as an analysis of the characteristics of health information–retrieval terms [28]. Studies in the stable growth period mainly focus on the patients' attitude and cognition toward health information [23], health information searching behavior, and its influencing factors [68]. The research in the rapid growth period mainly focuses on personal health information management [69], health literacy assessment, and health literacy intervention [70]. The internet health information search behavior of patients with different diseases is still the research hotspot in this period.

By analyzing the evolution characteristics of the research themes regarding the internet and information technology application, it can be found that the research in the slow growth period mainly focuses on the technology applied to patient education [71] and the development of a bibliographic database network interface [26]. The studies in the stable growth period mainly focus on the use of the internet and information technology to intervene in people's health activities [72], help improve the patients' health status, and strengthen the patients' self-care skills [73]. They also examined the use of information technology in the patients' health information search behavior [74], identifying text difficulty of health information and classifying health care webpages using machine learning and natural language processing methods [75,76]. The research in the rapid growth period mainly focuses on the application of natural language processing [77] and the development and evaluation of user-centered mobile medical apps [78]. Wongchaisuwat et al [79] developed an algorithm to automatically answer health-related question by implementing a semisupervised learning algorithm. Park et al [80] analyzed the online discussion content of 3 online health communities employing text mining and k-means machine learning algorithm to compare the discussion topics.

## Limitations

In this paper, we used the SciMAT tool to analyze the keywords of the literature data and constructed the overlapping map, evolution map, strategic diagram, and a cluster network. Owing to the limitation of the scope of our school's database, we only obtained the literature from 1999 to 2019, failed to trace the earliest research literature on CHI, and failed to build a complete evolution path of the CHI research theme.

## Conclusions

In this research, a bibliometric analysis was carried out to explore the dynamic evolution path and evolution laws of CHI research themes on a time dimension and from the perspective of strategic diagrams and data flows. CHI research focuses on themes, such as the patients' education, health information needs, information search behavior, health behavior intervention, health literacy, health information technology, and eHealth. The research content in 4 different study periods formed the 38 themes. These themes formed 10 evolution paths in 3 research directions: patient education and intervention, consumer demand attitude and behavior, and internet information technology application. Patient education and intervention research, consumer demand, attitude, and behavior research comprise the main theme evolution path. The path's evolution process has been relatively stable, and it will continue to represent the research hotspot in this field. Research on the internet and information technology application is a secondary theme evolution path, where obvious absorption, fragmentation, and extinction of themes have occurred. Its complex evolution process has also attracted the attention of many scholars. The research status and influence of this evolution path have gradually increased, making it a research direction with development potential.

According to the existing literature, themes that will continue to appear in the next study period include the following: *electronic medical records*, *online health information*, *health information seeking*, *attitudes*, *health literacy*, *technology*, and *natural language processing*. The first 2 themes are information resource elements *in* the consumer health information ecosystem. Because of the continuous advancement of medical informatization and the rapid development of social media, electronic health records, online health information, and online health question answering data have formed massive health information repositories. These health information big data are like gold mines waiting to be explored by researchers. The middle 3 themes are information subject elements, and the last 2 themes are the information technical elements in the environment. To maintain the sustainable development of the consumer health information ecosystem, it is necessary to start from the supply side and the demand side of health information and to solve the mismatch between service and demand by using computer and information technology.

## Acknowledgments

We thank the State Key Program of National Social Science of China (grant no. 17AZD037) for assistance in this study. We also appreciate the support of all participants.

## Authors' Contributions

AL and WX co-designed this research. WO, ZX, and HH conducted the data retrieval/construction, data collection, and analysis. YY, FL, and YC helped preprocess and analyze the data. WO interpreted the results and drafted the manuscript. TW, XP, and PD made major revisions. All authors reviewed and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Cluster network of the themes (2009-2014).
[PNG File , 5060 KB-Multimedia Appendix 1]

## Multimedia Appendix 2

Cluster network of the themes (2015-2019).
[PNG File , 4909 KB-Multimedia Appendix 2]

## References

1. Thornton K. Directions and Opportunities in Health Informatics in British Columbia. Methods Inf Med 1994 Jun;33(3):299-301. [doi: 10.1055/s-0038-1635019]
2. Ferguson T. Consumer health informatics. Healthc Forum J 1995;38(1):28-33. [Medline: 10154286]
3. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
4. American Medical Informatics Association. Consumer Health Informatics. URL: https://www.amia.org/applications-informatics/consumer-health-informatics [accessed 2020-06-08]
5. Kokol P, Saranto K, Vosner HB. eHealth and health informatics competences: A systemic analysis of literature production based on bibliometrics. Kybernetes 2018 Mar 05;47(5):1018-1030. [doi: 10.1108/k-09-2017-0338]

6. Zhao YH, Zhang J. Consumer health information seeking in social media: a literature review. Health Information and Libraries Journal 2017 Oct 17;34(4):268-283. [doi: 10.1111/hir.12192]
7. Xie HL, Zhang YW, Duan KF. Evolutionary overview of urban expansion based on bibliometric analysis in Web of Science from 1990 to 2019. Habitat International 2020 Jan;95:102100. [doi: 10.1016/j.habitatint.2019.102100]
8. Guo YQ, Hao ZC, Zhao SC, Gong JQ, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. Journal of Medical Internet Research 2020 Jul 29;22(7):e18228. [doi: 10.2196/18228]
9. Wang J, Deng H, Liu BT, Hu AB, Liang J, Fan LY, et al. Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed. Journal of Medical Internet Research 2020 Jan 23;22(1):e16816. [doi: 10.2196/16816]
10. Yonoff RE, Ochoa GV, Cardenas-Escorcia Y, Silva-Ortega JI, Merino-Stand L. Research trends in proton exchange membrane fuel cells during 2008-2018: A bibliometric analysis. Heliyon 2019 May;5(5):e01724. [doi: 10.1016/j.heliyon.2020.e04848]
11. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. SciMAT: A new science mapping analysis software tool. Journal of the American Society for Information Science and Technology 2012 Jul 02;63(8):1609-1630. [doi: 10.1002/asi.22688]
12. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the World Wide Web - A systematic review. JAMA-Journal of the American Medical Association 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691]
13. Cline RJW, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Education Research 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671]
14. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. Journal of Epidemiology and Community Health 1999 Feb 01;53(2):105-111. [doi: 10.1136/jech.53.2.105]
15. Cerminara C, Santarone ME, Casarelli L, Curatolo P, El Malhany N. Use of the DISCERN tool for evaluating web searches in childhood epilepsy. Epilepsy & Behavior 2014 Dec;41:119-121. [doi: 10.1016/j.yebeh.2014.09.053]
16. Alnaim L. Evaluation Breast Cancer Information on The Internet in Arabic. Journal of Cancer Education 2019 Aug;34(4):810-818. [doi: 10.1007/s13187-018-1378-9]
17. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. Journal of Medical Internet Research 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9]
18. Kim KA, Kim YJ, Choi M. Association of Electronic Health Literacy With Health-Promoting Behaviors in Patients With Type 2 Diabetes: A Cross-sectional Study. Cin-Computers Informatics Nursing 2018 Sep;36(9):438-447. [doi: 10.1097/cin.0000000000000438]
19. Cherid C, Baghdadli A, Wall M, Mayo NE, Berry G, Harvey EJ, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture—a survey in orthopedic clinics. Osteoporosis International 2020 Feb 28;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3]
20. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. American Journal of Preventive Medicine 1999;16(1):1-9. [doi: 10.1016/s0749-3797(98)00108-1]
21. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
22. van den Berg MH, Schoones JW, Vlieland TPMV. Internet-Based Physical Activity Interventions: A Systematic Review of the Literature. Journal of Medical Internet Research 2007 Sep 30;9(3):e26. [doi: 10.2196/jmir.9.3.e26]
23. 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]
24. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
25. Wu JS, Li YC, Jian WS. Consumer health information websites in Taiwan. Journal of the Formosan Medical Association 2000 Aug;99(8):663-666. [Medline: 10969514]
26. Marton CF, Kalsbeek K, Thulin K. Evaluating the Web user interface design of a bibliographic database on women's health. 2003 Jul Presented at: 7th World Multiconference on Systemics, Cybernetics and Informatics; Jul 27-30, 2003; Orlando, fl p. 456-461.
27. Lewis D. Computers in Patient Education. Cin-Computers Informatics Nursing 2003 Mar;21(2):88-96. [doi: 10.1097/00024665-200303000-00010]
28. Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of Consumer Terminology for Health Information Retrieval. Methods of Information in Medicine 2002;41(4):289-298. [doi: 10.1055/s-0038-1634490]
29. Doupi P, van der Lei J. Design and implementation considerations for a personalized patient education system in burn care. International Journal of Medical Informatics 2005 Mar;74(2-4):151-157. [doi: 10.1016/j.ijmedinf.2004.04.021]

30. Tufano JT, Karras BT. Mobile eHealth Interventions for Obesity: A Timely Opportunity to Leverage Convergence Trends. Journal of Medical Internet Research 2005;7(5):97-104. [doi: 10.2196/jmir.7.5.e58]
31. Wallace LS, Keenum AJ, Roskos SE, Blake GH, Colwell ST, Weiss BD. Suitability and readability of consumer medical information accompanying prescription medication samples. Patient Education and Counseling 2008 Mar;70(3):420-425. [doi: 10.1016/j.pec.2007.11.017]
32. Webb J, Davis TC, Bernadella P, Clayman ML, Parker RM, Adler D, et al. Patient-centered approach for improving prescription drug warning labels. Patient Education and Counseling 2008 Sep;72(3):443-449. [doi: 10.1016/j.pec.2008.05.019]
33. Zeng QT, Divita G, Keselman A, Crowell J, Browne AC, Goryachev S, et al. Term Identification Methods for Consumer Health Vocabulary Development. Journal of Medical Internet Research 2007 Mar 14;9(1):e4. [doi: 10.2196/jmir.9.1.e4]
34. Hesse BK, Arora NK, Beckjord EB, Rutten LJF. Information support for cancer survivors. Cancer 2008 Jun 01;112(11):2529-2540. [doi: 10.1002/cncr.23445]
35. McInnes DK, Cleary PD, Stein KD, Ding L, Mehta CC, Ayanian JZ. Perceptions of cancer-related information among cancer survivors: A report from the American Cancer Society's Studies of Cancer Survivors. Cancer 2008 Sep 15;113(6):1471-1479. [doi: 10.1002/cncr.23713]
36. Cui LC, Carter R, Zhang GQ. Evaluation of a Novel Conjunctive Exploratory Navigation Interface for Consumer Health Information: A Crowdsourced Comparative Study. Journal of Medical Internet Research 2014 Feb 10;16(2):e45. [doi: 10.2196/jmir.3111]
37. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online Access to Doctors' Notes: Patient Concerns About Privacy. Journal of Medical Internet Research 2013 Sep 26;15(9):186-198. [doi: 10.2196/jmir.2670]
38. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. Human Vaccines & Immunotherapeutics 2013 Dec 01;9(12):2634-2642. [doi: 10.4161/hv.26010]
39. Meesters JJL, Vlieland TPMV, Hill J, Ndosi ME. Measuring educational needs among patients with rheumatoid arthritis using the Dutch version of the Educational Needs Assessment Tool (DENAT). Clinical Rheumatology 2009 May 16;28(9):1073-1077. [doi: 10.1007/s10067-009-1190-3]
40. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging "Dr. Google": Does Technology Fill the Gap Created by the Prenatal Care Visit Structure? Qualitative Focus Group Study With Pregnant Women. J Med Internet Res 2014 Jun 03;16(6):e147. [doi: 10.2196/jmir.3385]
41. Hewitt-Taylor J, Bond CS. What E-patients Want From the Doctor-Patient Relationship: Content Analysis of Posts on Discussion Boards. Journal of Medical Internet Research 2012 Nov 08;14(6):301-309. [doi: 10.2196/jmir.2068]
42. Jacobs RJ, Caballero J, Ownby RL, Kane MN. Development of a culturally appropriate computer-delivered tailored internet-based health literacy intervention for spanish-dominant hispanics living with HIV. BMC Medical Informatics and Decision Making 2014 Nov 30;14:103. [doi: 10.1186/s12911-014-0103-9]
43. King L, Watson WL, Chapman K, Kelly B, Louie JCY, Hughes C, et al. Do We Provide Meaningful Guidance for Healthful Eating? An Investigation into Consumers' Interpretation of Frequency Consumption Terms. Journal of Nutrition Education and Behavior 2012 Sep;44(5):459-463. [doi: 10.1016/j.jneb.2011.12.004]
44. Meesters JJL, de Boer IG, van den Berg MH, Fiocco M, Vlieland TPMV. Evaluation of a website providing information on regional health care services for patients with rheumatoid arthritis: an observational study. Clinical Rheumatology 2012 Apr;31(4):637-645. [doi: 10.1007/s10067-011-1897-9]
45. Thakor V, Leach MJ, Gillham D, Esterman A. The quality of information on websites selling St. John's wort. Complementary Therapies in Medicine 2011 Jun;19(3):155-160. [doi: 10.1016/j.ctim.2011.05.005]
46. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older Adults' Perceptions of Home Telehealth Services. Telemedicine and E-Health 2013 Oct;19(10):786-790. [doi: 10.1089/tmj.2012.0272]
47. Reeder B, Le T, Thompson HJ, Demiris G. Comparing Information Needs of Health Care Providers and Older Adults: Findings from a Wellness Study. 2013 Presented at: 14th World Congress on Medical and Health Informatics (MEDINFO); Aug 20-23, 2013; Copenhagen, DENMARK p. 18-22.
48. Cameron MH, Poel AJ, Haselkorn JK, Linke A, Bourdette D. Falls requiring medical attention among veterans with multiple sclerosis: A cohort study. Journal of Rehabilitation Research and Development 2011;48(1):13-20. [doi: 10.1682/jrrd.2009.12.0192]
49. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. Journal of the American Medical Informatics Association 2017 Apr(E1):E166-E172. [doi: 10.1093/jamia/ocw108]
50. Ntlotlang T, Grand B. The role of libraries in the dissemination of health information in Botswana A study of Mochudi and Molepolole public libraries. Library Review 2016 Jul 04;65(4-5):320-349. [doi: 10.1108/lr-05-2015-0051]
51. Deng ZH, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. International Journal of Medical Informatics 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014]
52. Chiu MHP. An investigation of the questions posted on medical consultation websites. Health Information and Libraries Journal 2016 Jul 06;33(4):283-294. [doi: 10.1111/hir.12153]

53. Fu LY, Zook K, Spoehr-Labutta Z, Hu P, Joseph JG. Search Engine Ranking, Quality, and Content of Web Pages That Are Critical Versus Noncritical of Human Papillomavirus Vaccine. Journal of Adolescent Health 2016 Jan;58(1):33-39. [doi: 10.1016/j.jadohealth.2015.09.016]
54. Lee YJ, Kamen C, Margolies L, Boehmer U. Online health community experiences of sexual minority women with cancer. Journal of the American Medical Informatics Association 2019 Aug;26(8-9):759-766. [doi: 10.1093/jamia/ocz103]
55. Huh J, Kwon BC, Kim SH, Lee S, Choo J, Kim J, et al. Personas in online health communities. Journal of Biomedical Informatics 2016 Oct;63:212-225. [doi: 10.1016/j.jbi.2016.08.019]
56. Monkman H, Kushniruk AW. Consumer Medication Information: Similarities and Differences Between Three Canadian Pharmacies. 2017 Presented at: Conference on Information Technology and Communications in Health (ITCH); Feb 16-19, 2017; Victoria, CANADA p. 238-242.
57. Monkman H, Kushniruk AW. All Consumer Medication Information Is Not Created Equal: Implications for Medication Safety. 2017 Presented at: Conference on Information Technology and Communications in Health (ITCH); Feb 16-19, 2017; Victoria, CANADA p. 233-237.
58. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics Journal 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220]
59. LeRouge C, Durneva P, Sangameswaran S, Gloster AM. Design Guidelines for a Technology-Enabled Nutrition Education Program to Support Overweight and Obese Adolescents: Qualitative User-Centered Design Study. Journal of Medical Internet Research 2019 Jul 29;21(7):e14430. [doi: 10.2196/14430]
60. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of Online Health Information-Seeking Behavior and Self-Care Activities Among Type 2 Diabetic Patients in Saudi Arabia. Journal of Medical Internet Research 2015 Aug 12;17(8):e196. [doi: 10.2196/jmir.4312]
61. Lee J, Kim J, Hong YJ, Piao M, Byun A, Song H, et al. Health Information Technology Trends in Social Media: Using Twitter Data. Healthcare Informatics Research 2019;25(2):99-105. [doi: 10.4258/hir.2019.25.2.99]
62. Gu G, Zhang XT, Zhu XG, Jian Z, Chen K, Wen D, et al. Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach. JMIR Medical Informatics 2019 May 23;7(2):216-224. [doi: 10.2196/12704]
63. Benis A, Harel N, Barkan RB, Srulovici E, Key C. Patterns of Patients' Interactions With a Health Care Organization and Their Impacts on Health Quality Measurements: Protocol for a Retrospective Cohort Study. JMIR Research Protocols 2018 Nov 07;7(11):e10734. [doi: 10.2196/10734]
64. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing Competencies Needed to Engage With Digital Health Services: Development of the eHealth Literacy Assessment Toolkit. Journal of Medical Internet Research 2018 May 10;20(5):e178. [doi: 10.2196/jmir.8347]
65. Koo K, Yap RL. How Readable Is BPH Treatment Information on the Internet? Assessing Barriers to Literacy in Prostate Health. American Journal of Mens Health 2017 Mar;11(2):300-307. [doi: 10.1177/1557988316680935]
66. Haymes AT. The Quality of Rhinoplasty Health Information on the Internet. Annals of Plastic Surgery 2016 Feb;76(2):143-149. [doi: 10.1097/sap.0000000000000660]
67. Abundo BJ. Consumer Health Education (MAPEH). 2015 Sep 15. URL: https://www.slideshare.net/senohj/consumer-health-education-mapeh [accessed 2020-06-08]
68. Boot CRL, Meijman FJ. The public and the Internet: Multifaceted drives for seeking health information. Health Informatics Journal 2010 Jun 23;16(2):145-156. [doi: 10.1177/1460458210364786]
69. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The Invisible Work of Personal Health Information Management Among People With Multiple Chronic Conditions: Qualitative Interview Study Among Patients and Providers. Journal of Medical Internet Research 2015 Jun 04;17(6):e137. [doi: 10.2196/jmir.4381]
70. Huhta AM, Hirvonen N, Huotari ML. Health Literacy in Web-Based Health Information Environments: Systematic Review of Concepts, Definitions, and Operationalization for Measurement. Journal of Medical Internet Research 2018 Dec 19;20(12):e10273. [doi: 10.2196/10273]
71. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. International Journal of Medical Informatics 2002 Nov;65(3):169-177. [doi: 10.1016/s1386-5056(02)00048-5]
72. Chang BL. Internet Intervention for Community Elders. Western Journal of Nursing Research 2016 Jun;26(4):461-466. [doi: 10.1177/0193945904263428]
73. Pemu PE, Quarshie AQ, Josiah-Willock R, Ojutalayo FO, Alema-Mensah E, Ofili EO. Socio-demographic Psychosocial and Clinical Characteristics of Participants in e-HealthyStrides (c): An Interactive ehealth Program to Improve Diabetes Self-Management Skill. Journal of Health Care for the Poor and Underserved 2011;22(4):146-164. [doi: 10.1353/hpu.2011.0162]
74. Puspitasari I, Moriyama K, Fukui K, Numao M. Effects of Individual Health Topic Familiarity on Activity Patterns During Health Information Searches. JMIR Medical Informatics 2015 Mar 17;3(1):e16. [doi: 10.2196/medinform.3803]

75. Wang YL. Automatic recognition of text difficulty from consumers health information. 2006 Presented at: 19th IEEE Symposium on Computer-Based Medical Systems; Jun 22-23, 2006; Salt Lake City, UT p. 131-136. [doi: 10.1109/cbms.2006.58]
76. Chen GC, Warren J, Riddle P. Semantic Space models for classification of consumer webpages on metadata attributes. Journal of Biomedical Informatics 2010 Oct;43(5):725-735. [doi: 10.1016/j.jbi.2010.06.005]
77. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and Analyzing Patient Experiences of Health Care From Social Media. JMIR Research Protocols 2015 Jul 02;4(3):e78. [doi: 10.2196/resprot.3433]
78. Stec MA, Arbour MW, Hines HF. Client-Centered Mobile Health Care Applications: Using the Mobile Application Rating Scale Instrument for Evidence-Based Evaluation. Journal of Midwifery & Women's Health 2019 Mar 18;64(3):324-329. [doi: 10.1111/jmwh.12941]
79. Wongchaisuwat P, Klabjan D, Jonnalagadda SR. A Semi-Supervised Learning Approach to Enhance Health Care Community-Based Question Answering: A Case Study in Alcoholism. JMIR Medical Informatics 2016 Jul 02;4(3):e24. [doi: 10.2196/medinform.5490]
80. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. Computers in Human Behavior 2018 Jan;78:98-112. [doi: 10.1016/j.chb.2017.09.001]

## Abbreviations

**CHI:** consumer health informatics

*Edited by R Kukafka; submitted 30.06.20; peer-reviewed by S Wen, L Alpay; comments to author 04.11.20; revised version received 23.11.20; accepted 13.07.21; published 09.09.21*

*Please cite as:*
*Ouyang W, Xie W, Xin Z, He H, Wen T, Peng X, Dai P, Yuan Y, Liu F, Chen Y, Luo A*
*Evolutionary Overview of Consumer Health Informatics: Bibliometric Study on the Web of Science from 1999 to 2019*
*J Med Internet Res 2021;23(9):e21974*
*URL: https://www.jmir.org/2021/9/e21974*
*doi: 10.2196/21974*
*PMID:*

Review

# Factors Affecting Engagement in Web-Based Health Care Patient Information: Narrative Review of the Literature

Liam Alperen Oktay[1], BSc; Eyad Abuelgasim[1], BSc; Aida Abdelwahed[1], BSc; Nour Houbby[1], BSc; Smaragda Lampridou[2], MSc, RN; Pasha Normahani[2], BSc, MBBS, MSc, MRSC; Nicholas Peters[1], MBBS, MD, FRCP, FHRS; Usman Jaffer[2], BSc, MBBS, MSc, PhD, PGCE, FRCS

[1]Imperial College London, London, United Kingdom

[2]Imperial College NHS Trust, London, United Kingdom

**Corresponding Author:**
Usman Jaffer, BSc, MBBS, MSc, PhD, PGCE, FRCS
Imperial College NHS Trust
Praed Street
London, W2 0NE
United Kingdom
Phone: 44 7968872992
Email: usman.jaffer@nhs.net

## Abstract

**Background:** Web-based content is rapidly becoming the primary source of health care information. There is a pressing need for web-based health care content to not only be accurate but also be engaging. Improved engagement of people with web-based health care content has the potential to inform as well as influence behavioral change to enable people to make better health care choices. The factors associated with better engagement with web-based health care content have previously not been considered.

**Objective:** The aims of this study are to identify the factors that affect engagement with web-based health care content and develop a framework to be considered when creating such content.

**Methods:** A comprehensive search of the PubMed and MEDLINE database was performed from January 1, 1946, to January 5, 2020. The reference lists of all included studies were also searched. The Medical Subject Headings database was used to derive the following keywords: “patient information,” “online,” “internet,” “web,” and “content.” All studies in English pertaining to the factors affecting engagement in web-based health care patient information were included. No restrictions were set on the study type. Analysis of the themes arising from the results was performed using inductive content analysis.

**Results:** The search yielded 814 articles, of which 56 (6.9%) met our inclusion criteria. The studies ranged from observational and noncontrolled studies to quasi-experimental studies. Overall, there was significant heterogeneity in the types of interventions and outcome assessments, which made quantitative assessment difficult. Consensus among all authors of this study resulted in six categories that formed the basis of a framework to assess the factors affecting engagement in web-based health care content: easy to understand, support, adaptability, accessibility, visuals and content, and credibility and completeness.

**Conclusions:** There is a paucity of high-quality data relating to the factors that improve the quality of engagement with web-based health care content. Our framework summarizes the reported studies, which may be useful to health care content creators. An evaluation of the utility of web-based content to engage users is of significant importance and may be accessible through tools such as the Net Promoter score. Web 3.0 technology and development of the field of psychographics for health care offer further potential for development. Future work may also involve improvement of the framework through a co-design process.

***(J Med Internet Res 2021;23(9):e19896)*** doi: 10.2196/19896

**KEYWORDS**

patient education; web-based health information; internet; patient engagement; mobile phone

# Introduction

## Background

In the United Kingdom, up to two-third use the internet to obtain health-related information at some point in their journey [1,2]. The internet has become an important source of education for patients, who are increasingly expected to, and are motivated to, play an active role in making decisions related to their health [3]. Patient education is defined as "the process by which health professionals and others impart information to patients that will alter their health behaviours or improve their health status" [4]. This may include information that is factual or related to patient experience, depending on the issue being addressed [5,6]. Reports suggest that 70% of the patients would like their physicians to recommend a source of web-based information relating to their medical condition, but only 4% of the patients receive such a recommendation [7].

Web-based patient health care information has several potential benefits, including convenient 24-hour access potentially wherever you are, ability to enhance knowledge acquisition [8,9], reduce anxiety [9], and improve the quality of conversations during health-related encounters. However, the effect of web-based content on patient empowerment, self-efficacy, and health attitudes has been found to be variable [9]. This may be due to the absence of an evidence-based framework outlining the factors that should be considered for improving engagement with web-based health care information.

## Objectives

In this narrative review, we aim to identify and evaluate the factors that should be considered when producing engaging and high-quality web-based health care patient information. We also aim to incorporate these findings into a framework that may be useful as a guide to developing web-based health care information.

# Methods

## Overview

A comprehensive search of the PubMed and MEDLINE database was performed from January 1, 1946, to January 5, 2020. The Medical Subject Headings database was used to derive keywords and search term combinations, which included "patient information," "online," "internet," "web," and "content." All studies pertaining to the factors affecting engagement in web-based health care patient information were included. No restrictions were set on the study type. Only studies in English were included. Analysis of themes arising from the results was performed using inductive content analysis. All retrieved abstracts and titles were reviewed by 2 independent investigators (EA and LAO) for relevance pertaining to engagement with web-based health care content. Disagreements between the reviewers were solved by consensus. Manual cross-checking of the reference lists of the identified papers was carried out to identify any other potentially relevant studies.

Analysis of themes arising from the results was performed using inductive content analysis [10]. This involved the reviewing of titles and abstracts by an author (UJ), with free generation of the categories relating to *factors associated with high-quality and engaging web-based content*. The categories were named using content characteristic words, and these were expanded into subcategories. The number of times a category was reported in the articles was totaled, and a list of categories was created in order of the frequency of mentions. All authors reviewed this list to decide which categories would be included in the final list.

## Eligibility Criteria

As this is the first narrative review of its kind, no limitation was placed on study type or on surrogate measures of the outcome described.

## Outcomes

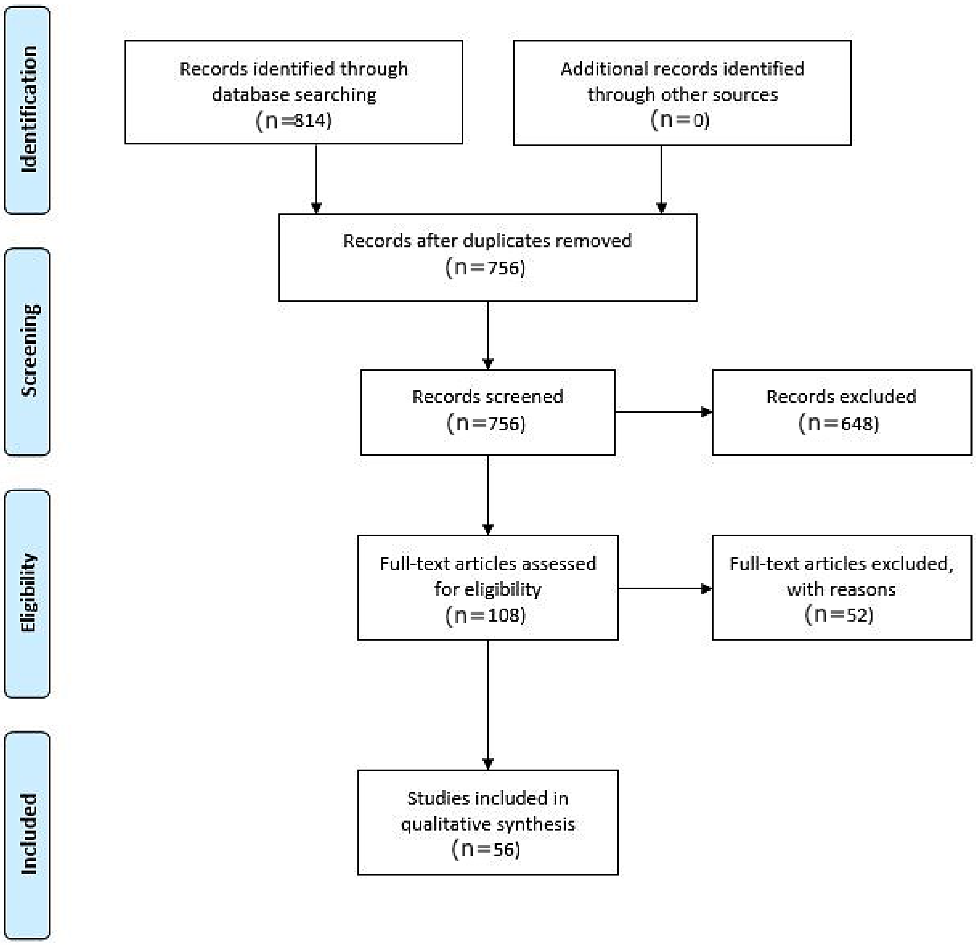
All studies relevant to quality of engagement in web-based health care content were included.

# Results

## Overview

A total of 814 articles were identified, and of these, 108 (13.3%) were selected for full-text review based on their title and abstract. Full-text screening of the 108 articles resulted in the final selection of 56 (51.9%) articles, from which seven categories were derived. Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart which depicts the stages of article selection.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing article selection.



## Description of Studies

### *Development of Categories*

The studies were mostly observational and qualitative. A total of 3 randomized controlled trials and 5 systematic reviews were also identified. Categories relating to *factors associated with high-quality and engaging web-based content* were developed according to themes that were found to arise in the studies identified. The categories settled upon were as follows (with the number of papers reporting included in parentheses): (1) textual information (16 papers); (2) discussion boards or web-based groups (3 papers); (3) video content (11 papers); (4) visuals or pictographs (1 paper); (5) device accessibility (12 papers); (6) stage of patient journey (8 papers); and (7) credibility and completeness of information (4 papers; Table 1).

**Table 1.** Studies pertaining to engagement with web-based content. Studies are according to mode of engagement (N=56).

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| **Textual information** | | | | |
| Chedid et al (2018) [11] | Observational study | Government-hosted website, textual, and visual aids for prenatal health promotion | • Comprehensiveness, evidence-based information, accessibility, and inclusivity.<br>• Minimum of three referenced prenatal health topics necessary to be classed as evidence-based. | • Assessment of prenatal health promotion material revealed that 66.4% of the federal, 84.6% of the provincial or territorial, and 80% of the public health regional unit–hosted websites, and 87.5% of the e-classes were evidence based. Only 25% of the municipal websites met this standard.<br>• No *P* values stated. |
| Ernst et al (2019) [12] | Observational study | Disorders of sex development, affiliated health care system's web-based information | • The SMOGa Readability Formula determined reading level, the PEMATb evaluated content for understandability and actionability, and the DISCERN Tool assessed treatment decision-making information. | • Reading level of webpages as determined with the SMOG Readability Formula met or exceeded high school grade level. The mean PEMAT understandability score for team pages and team links was 68% (SD 6%). On average, the pages met less than 70% of the understandability criteria. The mean PEMAT actionability score was 23% (SD 20%). The DISCERN Tool found that the quality of information relating to hormone treatment and to surgery was poor.<br>• No *P* values stated. |
| Hjelmager et al (2019) [13] | Qualitative study | Web-based information material for patients with low back pain in general practice discussed in the context of app development | • Barriers to and facilitators for future use of the health information technology app for patients with low back pain. | • Eight 1-hour interviews with general practitioners revealed the following: content for lower back pain should be validated by general practitioners; from a trustworthy source; support ongoing treatment plan.<br>• No *P* values stated. |
| Rofaiel et al (2018) [14] | Observational study | Websites that describe the biologic agents used as treatment options for inflammatory bowel disease | • The DISCERN model was used to evaluate the quality of the information content. | • The mean DISCERN score across all websites was 3.21 out of a 5-point scale. No significant difference was found between patient-searched and physician-recommended websites, with a mean score of 3.21 versus 3.63, respectively (*P*=.16). |
| Alfonso et al (2019) [15] | Observational study | American Cleft Palate–Craniofacial Association–approved teams' websites | • Content and readability of team websites. | • The mean reading level 10.7 (SD 1.9) exceeded the American Medical Association-recommended sixth grade reading level.<br>• Children's Hospital–affiliated teams (n=86) yielded significantly higher content scores (14.8 vs 13.5; *P*=.03).<br>• Children's Hospital teams also had better readability as indicated by lower reading grade level (10.5 vs 11.4; *P*=.04). |
| Ayyaswami et al (2019) [16] | Observational study | Web-based cardiovascular disease–related health education articles accessed through Google | • Readability according to 10 readability measures (Flesch Reading Ease, Coleman-Liau Index, Flesch-Kincaid Grade Level, Gunning Fog Index, FORCAST Readability Formula, New Dale-Chall formula, New Fog Count, SMOG Index, Fry Readability Formula, and Raygor Readability Estimate). | • All measures that assessed mean reading grade level found that 196 articles were written at a mean 10.9 (SD 1.8) grade reading level.<br>• 99.5% of the articles were written beyond the fifth to sixth grade reading level.<br>• No *P* values stated. |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Fajardo et al (2019) [17] | Systematic review | Web-based deprescribing patient education materials | - PEMAT and International Patient Decision Aids Standards Inventory.<br>- Readability using Gunning Fog Index and Flesch-Kincaid Grade Level. | - Patient education materials addressing deprescribing of medications for symptom control (81%) were most common.<br>- 37% of deprescribing patient education materials present potential benefits and harms of deprescribing.<br>- Most patient education materials are pitched above average reading levels (average minimum reading level of grade 12).<br>- No *P* values stated. |
| Vivekanantham et al (2017) [18] | Observational study | Web-based health information for patients with polymyalgia rheumatica | - Readability using the Flesch Reading Ease and SMOG Readability Formula tools.<br>- 8-item Credibility Indicator (incorporating authorship, affiliation, editorial team, date of creation, date of update, backing, accreditation, and financing). | - Of the websites identified (n=52), the mean Flesch Reading Ease and SMOG Readability Formula scores were 48 (SD 15) and 10 (SD 2), respectively.<br>- The mean Credibility Indicator was 2 (SD 1).<br>- Of 52 websites, 50 (96%) of the websites were accurate.<br>- No *P* values stated. |
| Harris et al (2018) [19] | Cross-sectional descriptive | Evaluation of leading web-based content on tympanostomy tube placement | - PEMAT understandability and actionability scores | - The PEMAT scores found that all sites (n=10) were understandable (mean 81.9%).<br>- Seven of the sites had a low actionability score (mean 44.6%).<br>- No *P* values stated. |
| Maciolek et al (2017) [20] | Observational study | Prostate biopsy web-based patient education materials | - Readability was assessed using Flesch-Kincaid Grade Level. Understandability and actionability were measured using the PEMAT. | - Of the 148 sites, 31 (20.9%) met the recommended below eighth grade reading level.<br>- The PEMAT understandability score for academic institution–sourced patient education materials was higher than that for patient education materials sourced from the private sector (*P*=.02) and from institutions unaffiliated with urologists (*P*=.01). |
| Siddhanamatha et al (2017) [21] | Observational study | Websites providing educational content for patients with rheumatoid arthritis | - Accuracy, completeness, technical elements, design and aesthetics, readability, usability, and accessibility of the websites | - Of 46 websites in total, 45 (98%) provided accurate information. The mean reading level was grade 12.1 (SD 2.3).<br>- In total, 78% (36/46) were easy to navigate, but only 33% (15/46) were user friendly for people with visual and or hearing impairments.<br>- No *P* values stated. |
| Nielsen-Bohlman et al (2004) [22] | Observational study | An evaluation of health literacy in the United States; formulate solution to overcome associated obstacles | - Current level of readability of web-based content | - More than 300 studies indicate that health-related materials exceed the average reading grade level of US adults.<br>- No *P* values stated. |
| Chin et al (2018) [23] | Observational study | Improve patient understanding of web-based content pertaining to adults with hypertension. | - Information retention and comprehension | - The revised passages yielded improved retention and comprehension, with less reading time required per unit uptake of information also noted. The methods included simplifying language and signaling clear organization.<br>- On average, the participants were found to significantly better remember the revised passages (mean 0.74, SD 0.14) compared with the typical passages (mean 0.70, SD 0.11; *P*<.01). |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Boudewyns et al (2015) [24] | Randomized controlled trial | Web-based information handouts | • Information comprehension and application | • Individuals who received the revised and improved web-based formats had higher comprehension scores than those who received the MedGuide (*P*<.001). |
| Morrow et al (2005) [25] | Observational study | Patient-centered medication instructions to empower patients to plan a medication-taking regimen | • Information comprehension and retention, health-related literacy, and verbal working memory | • Patient-centered instructions were more accurately understood for unfamiliar medications (*P*<.05).<br>• The standard instructions were more accurate for familiar medications (*P*<.05). |
| Pander Maat et al (2010) [26] | Observational study | Revision of patient information leaflets | • Usability, speed of information retrieval, and comprehension. | • Once located, comprehension of the information was approximately 90%.<br>• Revisions led to better performance. Information was found more quickly. Comprehension scores were also improved.<br>• No *P* values stated. |
| **Discussion boards or web-based groups** | | | | |
| Cedars et al (2019) [27] | Qualitative and thematic analysis | Web-based discussion boards for urethral stricture disease and urethroplasty | • To describe the patient experience and chief concerns with urethroplasty to improve physician understanding and patient education<br>• To understand how men use web-based discussion boards and what information is available about urethroplasty | • Problems in navigating the health care system with urethral stricture disease (n=141) and weak urine stream (n=70) were the most frequent preurethroplasty complaints.<br>• The patients participated in web-based discussions to share experiences with urethral stricture disease and urethroplasty, share emotional support, and search for answers.<br>• No *P* values stated. |
| Teaford et al (2019) [28] | Descriptive statistics and qualitative content analysis | New mothers' experiences with web-based postpartum forums | • To explore women's experiences with a web-based forum during the postpartum period | • Five themes were produced through data analysis: (1) social support, (2) anonymity, (3) in-groups, (4) drama, and (5) entertainment or pastime.<br>• The participants found that the discussion board could facilitate sharing of information, act as an entertainment source, and provide community.<br>• No *P* values stated. |
| Castaneda et al (2019) [29] | Qualitative study | eHealth peripheral artery disease community forums | Original posts and related responses were analyzed for thematic content. | • The themes identified included medical advice (41%), personal experiences with peripheral artery disease (33%), and social support (13%). Negative attitudes were discussed in 10 of the 18 (56%) posts related to poor experiences with health care providers; 15.1% of the medical advice was inconsistent with clinical treatment guidelines.<br>• No *P* values stated. |
| **Video** | | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Bae et al (2018) [30] | Observational study | YouTube videos in English as a patient education resource for cataract surgery | • 14 criteria important for educating patients about the procedure. | • The mean number of usefulness criteria satisfied was only 2.28 (SD 1.80) out of 14.<br>• There was no significant difference in view counts between the most useful videos and other videos (*P*=.94).<br>• Videos from medical bodies such as the National Health Service were found more useful in terms of patient education (*P*<.001). |
| Pedersen et al (2019) [31] | Feasibility study | Development of a preventive educational video for patients exposed to whiplash trauma | • The development followed a systematic approach and was theory driven, supplemented with available empirical knowledge. | • The participants (n=4) felt that the content was “relevant, helpful, and reassuring to watch.”<br>• All four preferred video content instead of written material.<br>• No *P* values stated. |
| Finnegan et al (2018) [32] | Case study | A web-based vaccine communication project (textual, videos, and infographics) | • Case study of a provaccine information hub launched in 2011. Vaccines Today provides high-quality information about vaccines and diseases, expert interviews, answers to frequently asked questions, parent or patient stories, and videos or infographics. | • Two categories of informing patients were found to work well: (1) the storytelling approach and (2) answering questions posed by patients.<br>• No *P* values stated. |
| Button et al (2018) [33] | Mixed methods study (qualitative and feasibility study) | A web-based intervention (TRAK[c]) that provides knee patients with health information | • Testing the TRAK intervention in patients undergoing physiotherapy to gain their subjective insights into its use | • The participants reported that TRAK was easy to use overall. Basic internet skills were all that were required.<br>• Using TRAK as an adjunct to physiotherapist management increased the patients’ understanding and confidence.<br>• No *P* values stated. |
| Vogel et al (2018) [34] | User survey | VaPE[d] in anesthesia | • The content of the videos, the technique of video presentation, usefulness of VaPE<br>• Interviews carried out with patients and physicians | • In total, 98% (78/80) of the anesthetists found VaPE useful for patient education.<br>• In total, 93% (74/80) observed time saved for the following interview.<br>• In total, 96% (77/80) stated that watching the video left patients better informed.<br>• Increased anxiety caused by VaPE was noted by 46% (37/80); 54% (43/80) found no such effect.<br>• No *P* values stated. |
| Pithadia et al (2019) [35] | Cross-sectional study | YouTube videos as a source of patient information about phototherapy and excimer laser for psoriasis | • Assess the educational quality of YouTube videos about phototherapy and excimer laser for psoriasis | • In total, 11.2% (15/135) of the videos contained high-quality patient educational information, 2.5% (3/135) were fair quality, and 66.1% (89/135) were low quality.<br>• A total of 28.2% (35/135) of videos provided background information regarding psoriasis. Of these 35 videos, 28 (80%) contained evidence-based content about the epidemiology, systemic involvement, genetics, and immune nature of psoriasis.<br>• Of the 35 videos, 7 (20%) presented nonevidence-based claims and high mortality rates associated with psoriasis.<br>• No *P* values stated. |
| Ferhatoglu et al (2019) [36] | Observational study | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| | | Sleeve gastrectomy videos shared on YouTube | • The popularity of the videos was evaluated with the Video Power Index.<br>• The educational quality of the videos was measured using the DISCERN score, JAMAe benchmark criteria, and GQSf. The technical quality was measured by the SGSSg. | • The DISCERN, JAMA benchmark criteria, GQS, and SGSS evaluations of academic-sourced videos yielded significantly higher scores than patient-sourced videos (*P*<.001, *P*<.001, *P*=.001, and *P*<.001, respectively).<br>• However, the Video Power Index evaluation of patient-sourced videos yielded significantly higher scores than academic- and physician-sourced videos (*P*<.001 and *P*=.003, respectively).<br>• Negative correlations between the Video Power Index and the JAMA, GQS, and SGSS scores were found. |
| Erdem et al (2018) [37] | Observational study | Bariatric surgery videos (n=175) on YouTube | • Usefulness of bariatric surgery videos on YouTube: A usefulness score (very useful, useful, or not useful) | • Of the 175 videos, 94 (53.7%) were useful, and 43 (24.6%) were very useful. No videos were found containing misleading information.<br>• A Spearman rank correlation found no significant correlation between the usefulness score and the number of views (ρ=−0.118; *P*=.12), number of likes (ρ=−0.038; *P*=.61), number of dislikes (ρ=−0.003; *P*=.97) or video length (ρ=−0.106; *P*=.16). |
| Biggs et al (2013) [38] | Observational study | YouTube as a source of information on rhinosinusitis | • Videos (n=100) were graded on their ability to inform the layperson about rhinosinusitis. | • 45% of the videos were deemed to provide some useful information. 55% of the videos contained little or no useful facts, 27% of which contained potentially misleading or even dangerous information. Videos uploaded by medical professionals or those from health information websites contained more useful information than those uploaded by independent users.<br>• No *P* values stated. |
| Kwok et al (2017) [39] | Observational study | Videos available on YouTube pertaining to interventional treatment for varicose veins | • Informational and scientific quality (good, fair, and poor) and stance (for, neutral, against, and unclear) toward the treatment option discussed, treatment type, and video source. | • The largest group of videos (47.3%) received a quality rating of fair, meaning that they discussed one or two aspects of a treatment option, such as procedural technique and indications. Among those videos rated poor (25%), nearly all (98.2%) failed to mention a specific treatment.<br>• No *P* values stated. |
| Bademci et al (2017) [40] | Observational study | YouTube videos on deep venous thrombosis | • Scientific content, accuracy, and currency | • Although most of the videos uploaded (25/111, 22.9%) were created by physicians, the number of views for website-based videos was significantly higher (*P*=.002). When the uploaded videos were assessed in terms of their usefulness, the videos from physicians and hospitals were statistically more useful than the other videos (*P*<.001). |
| **Visuals or pictographs** | | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Christensen et al (2017) [41] | Pilot study | Doodle Health: A crowdsourcing web-based game for the co-design and testing of pictographs to reduce disparities in health care communication | • To test the usability of the game and its appeal to health care consumers in the co-design and evaluation of pictographs. | • Initial testing indicates that crowdsourcing is a promising approach to pictograph development and testing for relevancy and comprehension. More than 596 drawings were collected, and 1758 guesses were performed to date with 70%-90% accuracy.<br>• No *P* values stated. |
| **Device on which content is accessed** | | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Gogovor et al (2017) [42] | Literature review and qualitative focus group study | Development of an internet-based chronic pain self-management program | • Information needs and gaps in chronic pain management as well as technology features to inform the development of an internet-based self-management program | • The gaps identified in terms of chronic pain management included lack of knowledge, limited access to health care, substandard care, and scarce self-management support.<br>• The focus group themes included patient education on chronic pain care, attitude-belief-culture, financial and legal issues, and motivational content.<br>• No $P$ values stated. |
| Lüchtenberg et al (2008) [43] | Observational study | Websites containing medical information addressing laymen or patients (n=139) | • Accessibility using a quantitative checklist based upon the Web Content Accessibility Guidelines of the World Wide Web Consortium | • Of the 139 sites, 25 (17.9%) of the sites were categorized as WAIh guidelines level A or AA.<br>• WAI guidelines level AA was reached by 0.7% (1/139) of website.<br>• None of the websites reached level AAA.<br>• Of the 139 sites, 82% (114) of the assessed consumer websites were not completely accessible to persons who are visually impaired.<br>• No $P$ values stated. |
| Bashi et al (2018) [44] | Systematic review | Patient educational interventions using mobile apps | • The reporting quality of the studies was assessed according to the mHealthi evidence and predefined reporting assessment criteria. | • Of the 15 studies, none reported on the data security, privacy, and confidentiality measures.<br>• No $P$ values stated. |
| Noel et al (2017) [45] | Prospective cohort study | A mobile medical app was developed to improve postoperative care of patients who had undergone plastic surgery | • The content, design, and efficacy of the app were assessed with a questionnaire (n=60). | • The participants reported that the questions regarding postoperative management were addressed effectively, with a mean score of 4.1/5.<br>• Most of the participants recommended the app to patients who had undergone plastic surgery, with a mean score of 4.6/5.<br>• The app's information prevented 12 patients from calling the plastic surgeon or the emergency department unnecessarily.<br>• No $P$ values stated. |
| Nicholas et al (2015) [46] | Systematic review | Mobile apps for bipolar disorder | • The comprehensiveness and quality of information was assessed against core psychoeducation principles and current bipolar disorder treatment guidelines.<br>• The management tools were evaluated with reference to the best practice resources for the specific area.<br>• General app features and privacy and security. | • Informative apps covered more than a third (4/11, 36%) of core psychoeducation principles and 15% (2/13) of best practice guidelines.<br>• A third (10/32, 31%) cited their sources.<br>• "Neither comprehensiveness of psychoeducation information ($\rho=-0.11$; $P=.80$) nor adherence to best practice guidelines ($\rho=-0.02$; $P=.96$) were significantly correlated with average user ratings." |
| Jamison et al (2017) [47] | Randomized controlled trial | To test an app that enables patients with chronic pain to assess, monitor, and communicate their status to their health care provider. | • Frequency of app use and app satisfaction scores | • In total, 78.1% (82/105) of the participants reported daily using the app.<br>• Patient satisfaction survey results:<br>• Ease of use: 1.8/10 (0=very easy to use, 10=unusable)<br>• Willingness to use after the study: 2.4/10 (0=very willing; 10=unwilling.<br>• Participants with more daily assessments reported higher app satisfaction ($P<.05$) than those who used the app less. |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Schulz et al (2007) [48] | Randomized controlled trial | Website designed to enhance self-management in chronic lower back pain | • Change in pain levels, change in knowledge, behavioral changes, and medication use | • Users accessed the website an average of 11.5 times during the 5-month study.<br>• Mean pain levels fell in the control group from 5 to 3.9 (10=most severe pain imaginable, 1=no pain), whereas the mean pain levels in the control group remained largely the same (6.1 to 6.3).<br>• No *P* values stated. |
| Caiata Zufferey et al (2009) [49] | Observational study | Website *Oneself* designed to promote self-management and inform patients on lower back pain management | • Self-comprehension<br>• Improvement of vocabulary, knowledge of exercises, self-confidence, and motivation | • Of the 129 survey participants, 32 (24.8%) reported that Oneself increased their knowledge about back pain.<br>• Successful testimonials indicated that self-management was encouraged.<br>• No *P* values stated. |
| Hagerman et al (2017) [50] | Observational study | DAs[j] for patients with low-risk PCa[k] | • What are the informational needs of patients with low-risk PCa, and what are the essential aspects of treatment DAs that increase the likelihood of physicians recommending them to the patient? | • Semistructured interviews found that "Physicians highlighted the importance of patient education and described the characteristics of a low-risk PCa DA that would increase the likelihood of its use in clinical practice."<br>• Encourage patients to take their time in decision-making.<br>• Frankly inform on posttreatment side effects.<br>• Incorporate physician recommendations on content and mode of delivery.<br>• No *P* values stated. |
| Kim et al (2002) [51] | Observational study | Website comprising a situational approach to the organization of disease-specific patient information | • Interface usability, personal relevance of retrieved information, comprehension of retrieved information. | • Responses (n=37) yielded high ratings for the following: interface usability (4.6/5); personal relevance of information found (4.7/5); comprehension of information (4.8/5).<br>• No *P* values stated. |
| Meppelink et al (2015) [52] | Observational study | Colorectal cancer screening messages divided into high-literacy and low-literacy groups, with and without illustrations | • Information recall, attitudes, intention to undergo screening | • Spoken messages about colorectal cancer screening improved recall ($P$=.03) and attitudes ($P$=.02) compared with written messages in individuals with lower health literacy.<br>• Animations alone failed to improve recall, but when combined with spoken text, they significantly improved recall ($P$=.02). |
| Mayer et al (2003) [53] | Literature review | A theory designed to format multimedia content to optimize patient education | • Overloading, speed of content delivery, and misalignment of textual and visual cues | • Narration has better transfer of information than on-screen text.<br>• Learner-controlled segments increase transfer of information.<br>• Graphics and corresponding text should be aligned visually.<br>• Signals also improve transfer.<br>• No *P* values stated. |
| **Stage of patient journey** | | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Biernatzki et al (2018) [54] | Cross-sectional descriptive | Evaluation of the informational needs of patients with diabetes | • Treatment process, course of disease, abnormalities of glucose metabolism, and diabetes through the life cycle | • Information needs among patients with diabetes is poorly investigated, although in high demand.<br>• No *P* values stated. |
| Boyde et al (2009) [55] | Observational study | An investigation of the learning style and learning needs of patients with HF[l] | • Questionnaire identifying preferred learning modalities | • In total, 64% (55/86) of the participants reported a preference for multimodal learning style; 18% (15/86) preferred textual information; 11% (9/86) preferred auditory; and 7% (6/86) preferred kinesthetic.<br>• Signs and symptoms were ranked as the most important topics to learn about.<br>• No *P* values stated. |
| Hagenhoff et al (1994) [56] | Systematic review | Evaluation of the perceptions of both patients and nurses on the importance of educational content for patients with congestive HF | • Questionnaire evaluating the importance of the following categories: anatomy and physiology; psychology; risk factors; medications, diet, and activity; and other | • Patients and nurses rated all information as moderately to very important to learn.<br>• Patients often rated information as more important than nurses did.<br>• No *P* values stated. |
| Wehby et al (1999) [57] | Descriptive comparative study | Perceptions of RNs[m] and patients concerning educational content on HF were analyzed | • Ranking of categories of HF education in order of importance by patients and RNs | • “Patients perceived the subscales of general HF information, risk factors, medications, prognosis, and signs and symptoms as more important to learn than the RNs ($P<.05$).”<br>• “Patients perceived diet information as less important to learn than the RNs ($P<.05$).”<br>• “Patients perceived all eight subscales as more realistic to learn than the RNs ($P<.05$). Although not in identical order, both groups ranked education related to medication and signs and symptoms as the 2 priority areas.”<br>• “Diet information was ranked eighth by the patients and third by the RNs.”<br>• No *P* values stated. |
| Clark et al (2004) [58] | Descriptive correlational study | Examination of perceived learning needs of patients with heart failure after discharge | • “The Outpatient Heart Failure Learning Needs Inventory was used to rate the participants’ perceptions of the importance of educational topics on a 5-point Likert scale.” | • “The findings indicated that the subjects perceived signs and symptoms and medications as most important to learn and diet, activity, and psychological factors as least important to learn. These findings are consistent with previous research and provide a framework on which to base the development of educational programs for patients with heart failure. A significant finding was that nearly 25% of the screened patients were unable to participate because they were unaware that they had been diagnosed with heart failure.”<br>• No *P* values stated. |
| Kiliç B et al (2015) [59] | Descriptive comparative study | Qualitative analysis of questionnaires examining perceptions of RNs and patients concerning educational content on HF were analyzed. | • Themes related to the educational needs of patients about use of drugs.<br>• Themes related to lifestyle changes.<br>• Themes about the educational needs of the patients related to the characteristics of the disease ranking of categories of HF education in order of importance by patients and RNs. | • “In this study, HF patients stated that they mainly need information about the effects and purposes of the drugs they used. The need for information about the management of the symptoms that affect daily activities are considered 2nd and the educational needs about the disease itself are considered 3rd in importance.”<br>• No *P* values stated. |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Solomon et al (2018) [60] | Qualitative study | To build an evidence-based web-based patient information resource for patients with HIV | - Transcribed interviews of stakeholders underwent qualitative content analysis | - The interviewees suggested that descriptions of all members of the health care team involved with HIV care be included on the website.<br>- It was also suggested to organize the menu into health challenge categories for ease of navigation.<br>- No *P* values stated. |
| Liu et al (2017) [61] | Umbrella review | Aimed to identify the current evidence on health education–related interventions for patients with acute coronary syndrome or type 2 diabetes | - Clinical outcomes, behavioral outcomes, psychosocial outcomes, and medical service use | - Nurses and multidisciplinary teams were the most frequent health care professionals to provide education, and most educational interventions were delivered after discharge.<br>- Face-to-face sessions were the most common delivery formats of the patient educational interventions.<br>- The psychoeducational interventions were found to be effective in reducing smoking and admissions for patients with acute coronary syndrome.<br>- No *P* values stated. |
| **Credibility and completeness** | | | | |

| Study | Study type | Content | Outcomes pertaining to quality of web-based health care patient content | Main results |
|---|---|---|---|---|
| Boyer et al (1998) [62] | Review | Review of *HONcode*[n], a guideline designed to raise the quality of web-based patient education data | - Guidelines to information providers, with the aim of raising the quality of web-based data available and helping to identify websites that are maintained by qualified people and contain reliable data. | - The HONcode mainly includes the following ethical aspects: the author's credentials, the date of the last modification with respect to clinical documents, confidentiality of data, source data reference, funding, and the advertising policy.<br>- No *P* values stated. |
| Priyanka et al (2018) [63] | Observational study | Evaluation and analysis of web-based content pertaining to esophageal duodenoscopy for patients | - GQS, Health on Net, Flesch-Kincaid Reading Ease, and Flesch-Kincaid Grade Level | - Three websites were found to have high-quality, comprehensive, and authentic information: Healthline, Uptodate, and Emedicine.<br>- In total, 13 sites yielded moderate quality of information. The mean Flesch-Kincaid Reading Ease score was 46.92.<br>- The mean Flesch-Kincaid Grade Level was 11th grade.<br>- No *P* values stated. |
| Couper et al (2010) [64] | Cross-sectional survey | Analysis of the perceived importance of sources of health information on the web | - Ranking of sources in terms of reliability and influence; use of the internet in age groups. | - Internet use was more common at younger ages, increasing from 14% among those aged 70 years or older to 38% for those aged 40-49 years.<br>- Internet users rated health care providers as the most influential source of information for medical decisions, followed by the internet, family and friends, and media.<br>- No *P* values stated. |
| Volk et al (2013) [65] | Cross-sectional survey | Evaluation of ongoing studies regarding what the standards for DAs for patients should be | - Voting system to develop criteria for DA standards | - The review comprised 13 manuscripts on topics including current frameworks used to create health care content, health literacy, and the role of patient stories.<br>- No *P* values stated. |

[a] SMOG: Simple Measure of Gobbledygook.

[b] PEMAT: Patient Education Materials Assessment Tool.

[c] TRAK: Taxonomy for the Rehabilitation of Knee Conditions.

[d] VaPE: Video-Assisted Patient Education.

[e] JAMA: Journal of the American Medical Association.

[f] GQS: Global Quality Scores.

[g] SGSS: Sleeve Gastrectomy Scoring System.

[h] WAI: Web Accessibility Initiative.

[i] mHealth: mobile health.

[j] DA: decision aid.

[k] PCa: prostate cancer.

[l] HF: heart failure.

[m] RN: Regional nurse.

[n] HONcode: Health on the Net Foundation Code of Conduct.

### Textual Information

A study by Ernest et al [12] used the *DISCERN* Tool to evaluate the quality of written information regarding hormone treatment and surgery, which was found to be poor [66]. DISCERN is a validated tool developed by an expert panel through a process of panel debate and health care information analysis. It is noted to be the "first standardised index of quality of consumer health information" [66]. The DISCERN Tool comprises 15 key questions that investigate publication reliability and details of treatment choices, followed by overall judgment of quality. Each question is answered on a 5-point scale ranging from no to yes.

The study by Vivenkanatham et al [18] evaluated textual information on polymyalgia rheumatica. The study concluded that for web-based health care content to be effective, readability must be accessible to people of all literacy levels. Similar conclusions on readability were also reached in the study by Maciolek et al [20], which reported that patient education materials are most effective when simple language accessible to a wide patient population is used. A quality assessment study of web-based content on rheumatoid arthritis analyzed the readability, applicability, and accessibility of patient education websites [21]. The mean reading level was found to be 12.1 (SD 2.3), according to the Flesch-Kincaid Readability Tool. This tool gives a measure of how difficult a passage written in English is to understand through analyzing factors such as word length, sentence length, and total number of syllables. It provides a grade level according to the US educational system ranging from fifth grade to college graduate [67]. The same study reported that 78% of the websites assessed were easy to navigate [21]. Importantly, only 33% of the websites were assessed to be user friendly for people who are visually or hearing impaired.

An evaluation of web-based information on disorders of sex development noted strengths, including the tendency of webpages to present focused information in chunks and in a logical sequence [12]. A semistructured interview study of general practitioners conducted by Hjelmager et al [13] revealed that health information technology apps require textual information that is targeted to patients and written with the input of health care professionals.

The study by Rofaiel et al [14] assessed the quality of website information about inflammatory bowel disease using the DISCERN score to assess reliability and the relevance of pertinent details. The mean DISCERN score for patient-searched websites was not statistically different from that for physician-recommended websites (3.21 vs 3.62, respectively; *P*=.16). Numerous studies [12,13,15-17] identified that readability, as analyzed by the Flesch-Kincaid Tool, of web-based information exceeded the recommended sixth to eighth grade reading levels [68].

In all, 3 studies evaluated the understandability and actionability of web-based patient information by using the validated Patient Education Materials Assessment Tool [12,17,20]. The tool comprises inventories (one for print and another for audiovisual content) that list desirable and undesirable characteristics of information and produce a numeric value for understandability as well as actionability (ie, how easy it is to act on given information). The scores range from 0% to 100%, and a higher score indicates that the text is more understandable or actionable. Web-based educational content pertaining to disorders of sexual development and tympanostomy tube placement yielded low actionability scores (mean scores 23% and 44.6%, respectively). Interestingly, a study by Maciolek et al [20] found that the Patient Education Materials Assessment Tool understandability score for academic institution patient educational material exceeded that for content created by private institutions (*P*=.02) and content from websites unaffiliated with a urologist (*P*=.01).

### *Discussion Boards or Web-Based Groups*

The study by Cedars et al [27] reported on a study of patients with urethral strictures who participate in web-based discussions and share experiences to gain emotional support and find answers. Patients participating in these web-based groups were more often than not satisfied with their postoperative outcomes. These findings are supported by the study by Teaford et al [28], which explored the experiences of new mothers using a web-based postpartum forum. The study found that web-based forums provided a sense of community and a platform for sharing information. They identified five themes pertaining to the participants' concept of a web-based community: social support, anonymity, in-groups, drama, and entertainment.

However, the findings of a study by Castaneda et al [29] highlight the importance of exercising caution with web-based forums. The study evaluated the content of peripheral arterial disease eHealth forums and found that 15.1% of the medical advice given on such platforms was inconsistent with guidelines. Furthermore, the study found that 10 of the 18 posts related to negative personal experiences with health care providers.

It is important to note that there may be differences in patient engagement when comparing groups that are run by health care (or affiliated) professional service providers with those set up by patients or lay individuals themselves. A study comparing the difference between the effects of peer-led and moderated groups found that moderated groups were often more active and therefore had higher patient engagement.

### *Video-Based Content*

The study by Pedersen et al [31] found that 4 participants who were interviewed after watching a 14-minute educational video on whiplash injuries felt reassured, particularly because the video aligned with information that they had received at the hospital. The study by Finnegan et al [32] reported that people who visited an information hub where videos were embedded in the webpage spent longer than 2 minutes on that webpage, indicating that visitors with average reading capability watch and listen to the video while browsing the text. The study also reported that this information hub's YouTube channel, which featured videos explaining the concept of herd immunity, was particularly successful, with visitors spending more than 6 minutes on that page [32].

In the context of managing health conditions, the study by Button et al [33] found that video-based content was particularly helpful for patients having physiotherapy for knee injuries because they were able to visualize the correct technique. Patient understanding and confidence were found to be improved by this intervention. In a separate study of 80 anesthetists using a Video-Assisted Patient Education intervention, 96% reported that patients felt that they had a better understanding of the information provided through the Video-Assisted Patient Education intervention, and 97.5% of the anesthetists felt that it was a useful form of patient education [34]. However, 46% noted increased anxiety caused by the intervention [34].

Interestingly, the study by Ferhatoglu et al [36] found in an evaluation of the quality of YouTube content on sleeve gastrectomy that patient experience and advertisement videos were significantly more popular than academic videos created by medical professionals, according to the Video Power Index (*P*<.001 and *P*<.003, respectively). The Video Power Index assesses video performance by comparing the video with leaders in its respective industry, measures its impact and effectiveness across all platforms, and facilitates strategies to cater to target audiences [69]. The study by Erdem and Sisik [37] reported similar findings and found no significant association between video traction (*likes*, *dislikes*, or *views*) and usefulness of the content (Spearman rank correlation ρ=−0.038, *P*=.61; ρ=−0.003, *P*=.97; and ρ=−0.118, *P*=.12, respectively). Interestingly, the study found no significant correlation in usefulness to video length (ρ=−0.106; *P*=.16) in contrast to the findings of the study by Biggs et al [38], which concluded that medical videos categorized as useful had a mean length of 6 minutes and 51 seconds, with the videos rated in the top 10 having a mean length of 14 minutes and 47 seconds.

Similar to previous findings, a study conducted to assess the quality of YouTube videos on cataract surgery concluded that videos created by medical organizations such as the National Health Service were significantly more useful in terms of educating patients about the procedure than videos sourced by independent medical professionals and other sources (*P*<.001) [30]. The study by Bademci et al [40] similarly concluded that medical topic videos on deep vein thrombosis sourced from medical professionals and hospitals were significantly more useful than videos from other sources (*P*<.001).

In a study evaluating YouTube content on varicose veins, it was found that most of the videos were dominated by the private health care sector and that the video content presented a distorted view of treatment options, consequently leading to a skewed patient perception of the therapeutic options available to them [39]. Not only were 32% of these videos found to be of poor quality, but videos from private medical companies were also significantly more likely to favor minimally invasive surgery over ligation and stripping than videos from other sources. Once again, there was no significant association found between quality of content and viewing frequency.

### *Visuals or Pictographs*

Pictographs are pictorial representations of words or phrases. The study by Christensen et al [41] described the value of pictographs in health communication. The study presents the results of building and testing the game *Doodle Health*, which is designed to produce pictographs through crowdsourcing. It found that this method of pictograph creation yielded positive feedback from focus groups with regard to usability and comprehension. Analysis of the feedback found that 62.2% of the participants praised the game, with a respondent describing it as “engaging and easy to use.” However, the study also found that people from diverse cultural backgrounds have different styles of communication, which may render visually presented information ineffective for minority groups. Crowdsourcing pictographs was suggested as a potential solution to this concern.

### *Device Accessibility*

A study by Gogovor et al [42] concluded that the next generation of web-based educational health care programs should integrate apps for reasons of (1) accessibility, (2) flexibility, and (3) security and trustworthiness. Almost all the studies in our review used web-based platforms and required devices that accessed the web, with some studies requiring devices that accessed the video-sharing platform, YouTube, specifically.

A study by Lüchtenberg et al [43] found that only 18% (25/139) of the health information websites evaluated achieved a high standard of accessibility for users who are visually impaired as assessed by the Web Accessibility Initiative guidelines. The study concluded that web-based educational health care content should at least meet the requirements of priority 1 (level A) and preferably priority 2 (level AA) of the Web Accessibility Initiative guidelines. Developers can ensure accessibility by, for instance, having text alternatives for all nontext content and using high contrast ratios of text and images [36].

Few studies used smartphones exclusively to deliver information. Describing the benefits of using a smartphone, the study by Bashi et al [44] noted that smartphone adoption is becoming widespread, meaning more people can be reached; however, studies on how apps should deliver information in terms of interface and content is particularly lacking. The study by Noel et al [45] assessed the impact of a mobile medical app on plastic surgery patient care. A total of 60 patients answered questionnaires pertaining to the app, and the results supported the app’s utility from the patient’s perspective. The mean score for the app’s ability to answer patient questions was 4.1/5, and 20% of the patients were prevented from unnecessarily calling the emergency department. The use of a smartphone also enabled patients to access information from different locations and allowed for real-time disease management. A systematic review by Nicholas et al [46] revealed that of the 11 studies of apps providing patient information, only a third covered core psychoeducational principles. Furthermore, the average user ratings were not correlated with either comprehensiveness or adherence to best practice guidelines (ρ=−0.11, *P*=.80 and ρ=−0.02, *P*=.96, respectively).

### *Stage of Patient Journey*

A qualitative study by Gogovor et al [42] using health care professionals and patient focus groups found that health care professionals suggested stratifying access to information according to *stage of advancement* in terms of their chronic pain management. A qualitative study of *Oneself*, a website designed for use by patients with chronic lower back pain, found it to be an effective adjunct to doctor-patient consultations because patients found that they were unable to have all their queries answered in the limited time of a consultation [49].

At different stages of a patient journey, decision aids may be used as adjunctive decision-making tools to support patients. In the study by Hagerman et al [50], the authors conducted semistructured interviews of 33 physicians to identify the desirable characteristics of decision aids. Of the 33 physicians, 20 (61%) stated that patients should be educated on the lack of urgency with regard to making a treatment decision. Of the 33 physicians, 28 (85%) agreed that decision aids should be provided to the patient after the consultation when the patient is at home. Furthermore, 36% (12/33) of the physicians deemed it *very* or *extremely* important that decision aids are designed to be used during and after consultations.

Tailoring information delivery to the stage of the patient journey is further supported by the results of the study by Kim et al [51] who developed a system to organize web-based disease-specific information according to a situational knowledge base model.

The approach categorizes information about a specific disease (eg, thyroid cancer) into sections corresponding to discrete clinical events (eg, presentation, fine-needle aspiration biopsy, and diagnosis). In all, 75 patients completed a questionnaire evaluating the website, which found mean usability to be 4.6/5, personal relevance of received information 4.7/5, and comprehension of received information 4.8/5.

### *Credibility and Completeness of Information*

A study by Jamison et al [47] found that 86% of the apps for pain conditions reported were created with no involvement by health care professionals. A comprehensive study by Bae et al [30], assessing the quality of the content of YouTube videos for cataract surgery patient education, found that there was an abundance of videos simply showing patients undergoing a live procedure. More than 20% of the educational videos were commercial and hence potentially misleading. This may make it challenging to find high-quality, comprehensive educational videos on the web.

A study by Pithadia et al [35] used the American Academy of Dermatology guidelines as a benchmark to evaluate the accuracy of patient information YouTube videos on psoriasis treatments. It concluded that 12% of the videos contained high-quality patient education content, and most of them were not patient-centric. Similarly, the study by Ferhatoglu et al [36] used the *Journal of American Medical Association* benchmark criteria to assess the educational quality of sleeve gastrectomy YouTube videos and found that this score was significantly higher in university-affiliated physician videos than in other videos ($P<.001$).

The Health on the Net Foundation Code of Conduct (HONcode) presents a set of eight principles designed to set the quality standard for web-based patient information [62] (Textbox 1). A study by Laversin et al [70] compared 165 HONcode-certified websites with 165 noncertified websites. Only 0.6% of the noncertified websites conformed to the principles of the HONcode compared with 89% of the certified sites ($P<.10$). As the study followed certified websites 6 months after certification, the effect of the HONcode certification shows short-term sustainability.

**Textbox 1.** Quality standard (adapted from the study by Laversin et al [70]).

**Eight Principles Designed to Set the Quality Standard for Web-Based Patient Information**

1. Authoritative: qualifications of the authors indicated
2. Complementarity: information should support the doctor-patient relationship
3. Privacy: personal data collected by the site kept private
4. Attribution: cite all references
5. Justifiability: back up claims relating to benefits and performance
6. Transparency: accessible presentation, accurate email contact
7. Financial disclosure: identify funding sources
8. Advertising policy: clearly distinguish advertising from editorial content

## Discussion

### Principal Findings

In today's world of access to knowledge often being initially web-based, it is of importance for health care professionals to be able to create effective content. This is further emphasized in the current environment where minimum contact between patients and health care providers is required. We performed a wide scoping review of the literature to identify the features of web-based content and other telemedicine requirements that may improve quality of engagement with web-based health care content in this growing field. Using these results, we have developed a framework (Figure 2) to facilitate the development of web-based patient health care content. All the categories except for one (visual or pictograph) were reported on by 3 or more papers. We included the visual or pictograph category after discussion because it was felt to be a relevant and important means of communicating information. Although the features identified are, on reflection, intuitive, the framework arms the content creator with the best available strategies in making the content engaging and hence effective.

**Figure 2.** Infographic framework for modes of engagement for web-based health care content. HONcode: Health on the Net Foundation Code of Conduct; PEMAT: Patient Education Materials Assessment Tool.

**ONLINE PATIENT EDUCATIONAL CONTENT FRAMEWORK**

**1 EASY TO UNDERSTAND**

- **Readability**- ensure language is simple. Flesch-Kincaid readability and Simple Measure of Gobbledygook tools can be used
- **Understandability**- The PEMAT tool which assess understanding and actionability can be used
- **Quality Framework**- DISCERN tool assess textual quality according to users' standards

**2 SUPPORT**

- **Online community**- medical questions can be addressed to physicians
- **Emotional support**- users find comfort in confinding in one another

**3 ADAPTABILITY**

- **Situational approach**- organising content according to stage of treatment/condition
- **Adjunct to consultations**- physicians deem it important to use decision aids during and after consultations

**4 ACCESSIBILITY**

- **Web-based accessibility**- minority groups such as visually impaired should be considered when developing content
- **Smart phone compatible**- improves flexibility of access and user reach potential

**5 VISUALS AND CONTENT**

- **Design content with physician input**- content built by healthcare professionals is more useful to patients
- **Create visuals with crowd-sourcing**- improves usability and comprehension amongst users
- **Consider cultural context**- ensure all users can understand visual information by adapting to minority styles of communication
- **Retention, understanding and confidence**- patients feel more informed with video education

**6 CREDIBILITY AND COMPLETENESS**

- **Healthcare Organisation Guidelines**- accuracy and educational quality of content should be judged by guidelines of bodies such as Royal Medical Colleges
- **Quality Standard**- HONcode may be used to validate the credibility of content

Other similar frameworks such as the International Patient Decision Aid Standards (IPDAS) and the Standards for Universal Reporting of Patient Decision Aid Evaluation Studies (SUNDAE) checklists [65,71] have been developed for the evaluation of decision aids. However, to the best of our knowledge, this is the first framework to specifically focus on engagement with web-based content rather than a checklist approach to ensuring that decisions are made with appropriate consideration given to all relevant issues and options. A study design published by Knerr et al [72] aimed to evaluate patient behavior in response to a patient decision aid based on the IPDAS standards but has not reported results. Although patient decision aids can be a form of web-based content, efforts up to now have been directed toward ensuring transparency and trust in imparting information rather than ensuring the ability of web-based material to engage the user [73]. The need to engage people in health care content has been overlooked. We propose that this framework be used to improve engagement, which in turn will improve adherence with medical treatment and hence improve outcomes. Nevertheless, the rigorous process used by the IPDAS and SUNDAE developers is one that we would be interested in following in refining and adapting the framework arrived at through this literature review.

Further work will involve a co-design process with all stakeholders (including patients) to refine the insights we have gained from the studies regarding engagement with web-based content. The personalization of health care content may have bearing on the issue of engagement. Studies have shown that it is possible to *segment* the population according to the likelihood of responding to health care messages [74,75]. Although our framework provides the best evidence available relating to engagement with web-based content, the holy grail may lie in developing further the field of psychographics for health care. Although researchers have investigated the way that segmentation affects a defined intervention, the effect of segmentation itself is yet to be assessed in a meaningful manner [75]. The internet is able to bring together varied but related content using the concept of the semantic web and the application of folksonomies [76]. The confluence of Web 3.0 (to crowdsource content relevant to a desired health care behavior), psychographic segmentation (including segmentation based on the proposed framework), and machine learning may provide a way forward. We have developed a Web 3.0 health care content platform (Health Shared) and intend to use it for this purpose.

### Limitations

The findings of this review should be considered in the context of several limitations. The principal limitation is that most of the components of this suggested framework are not supported by strong evidence. The studies were heterogeneous in their aims, interventions, and outcomes, and some were of poor methodological quality. Few studies discussed the effect of the platform used on patient engagement—for example, smartphone app versus website versus commercial health care information—which may play a role in patient engagement. However, given that most of the studies discussed in the review were general scans of available websites, the framework developed is largely applicable to the website development platform. In addition, few studies describe the differences between informational sites and other modes-of-engagement systems that provide the ability for patient input and enable patients to contact their provider or providers; therefore, we were not able to compare these patient information platforms.

Furthermore, because only the PubMed database was searched, studies are likely to have been omitted from this review. Despite this limitation, the components of the framework are intuitive, and we believe that its application may be beneficial to health care providers and content creators. Evaluation and subsequent validation of the proposed framework by key stakeholders, including patients, clinicians, and content creators, would increase the robustness.

## Conclusions

There is a paucity of high-quality data relating to the factors that improve users' quality of engagement with web-based health care content. Our framework summarizes the reported studies, which may be useful to health care content creators. Evaluation of the utility of web-based content to engage users is of significant importance and may be accessible through tools such as the Net Promoter score. Web 3.0 technology and development of the field of psychographics for health care offer further potential for development [75]. Future work may also involve improvement of the framework through a co-design process.

Although there are often specific health care issues needing to be addressed in response to crisis situations, we believe that this work is more generally important in facilitating patient activation and patient-supported self-management, which are two major pillars in how health care systems need to realign to keep up with increasing demand.

## Authors' Contributions

UJ conceived the idea; LAO, EA, AA, NH, SL, and PN wrote the manuscript; and UJ, LAO, and PN reviewed and edited the manuscript.

## Conflicts of Interest

UJ is the founder of Health Shared.

## References

1. Dutton WH, Blank G. Next generation users: The internet in Britain. SSRN J 2012:1960655. [doi: 10.2139/ssrn.1960655]
2. O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F. User-generated online health content: a survey of internet users in the United Kingdom. J Med Internet Res 2014 Apr 30;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]
3. Pellisé F, Sell P, EuroSpine Patient Line Task Force. Patient information and education with modern media: the Spine Society of Europe Patient Line. Eur Spine J 2009 Aug 19;18 Suppl 3(S3):395-401 [FREE Full text] [doi: 10.1007/s00586-009-0973-1] [Medline: 19381695]
4. Outreville JF, Kongstvedt PR. The managed health care handbook. J Risk Ins 1998 Jun;65(2):351. [doi: 10.2307/253545]
5. Entwistle VA, France EF, Wyke S, Jepson R, Hunt K, Ziebland S, et al. How information about other people's personal experiences can help with healthcare decision-making: a qualitative study. Patient Educ Couns 2011 Dec;85(3):291-298. [doi: 10.1016/j.pec.2011.05.014] [Medline: 21652162]

6. Kamel BM, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
7. Bader JL, Strickman-Stein N. Evaluation of new multimedia formats for cancer communications. J Med Internet Res 2003 Aug 29;5(3):e16 [FREE Full text] [doi: 10.2196/jmir.5.3.e16] [Medline: 14517107]
8. Marcus C. Strategies for improving the quality of verbal patient and family education: a review of the literature and creation of the EDUCATE model. Health Psychol Behav Med 2014 Jan 01;2(1):482-495 [FREE Full text] [doi: 10.1080/21642850.2014.900450] [Medline: 25750796]
9. Dekkers T, Melles M, Groeneveld BS, de Ridder H. Web-based patient education in orthopedics: systematic review. J Med Internet Res 2018 Apr 23;20(4):e143 [FREE Full text] [doi: 10.2196/jmir.9013] [Medline: 29685869]
10. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
11. Chedid RA, Terrell RM, Phillips KP. Best practices for online Canadian prenatal health promotion: a public health approach. Women Birth 2018 Aug;31(4):223-231. [doi: 10.1016/j.wombi.2017.10.005] [Medline: 29113753]
12. Ernst M, Chen D, Kennedy K, Jewell T, Sajwani A, Foley CB, DSD-TRN Psychosocial WorkgroupAccord Alliance. Disorders of sex development (DSD) web-based information: quality survey of DSD team websites. Int J Pediatr Endocrinol 2019;2019:1 [FREE Full text] [doi: 10.1186/s13633-019-0065-x] [Medline: 31149017]
13. Hjelmager DM, Vinther LD, Poulsen SH, Petersen LS, Jensen MB, Riis A. Requirements for implementing online information material for patients with low back pain in general practice: an interview study. Scand J Prim Health Care 2019 Mar 31;37(1):60-68 [FREE Full text] [doi: 10.1080/02813432.2019.1569413] [Medline: 30700191]
14. Rofaiel R, Chande N. Biologic agents in inflammatory bowel disease – quality of internet website information. Saudi J Gastroenterol 2018;24(6):336. [doi: 10.4103/sjg.sjg_55_18]
15. Alfonso AR, DeMitchell-Rodriguez EM, Ramly EP, Noel DY, Levy-Lambert D, Wang MM, et al. Assessment of American cleft palate-craniofacial association-approved teams' websites for patient-oriented content and readability. Cleft Palate Craniofac J 2019 Oct 26;56(9):1213-1219. [doi: 10.1177/1055665619850441] [Medline: 31129984]
16. Ayyaswami V, Padmanabhan D, Patel M, Prabhu AV, Hansberry DR, Agarwal N, et al. A readability analysis of online cardiovascular disease-related health education materials. Health Lit Res Pract 2019 Apr 01;3(2):74-80 [FREE Full text] [doi: 10.3928/24748307-20190306-03] [Medline: 31049489]
17. Fajardo MA, Weir KR, Bonner C, Gnjidic D, Jansen J. Availability and readability of patient education materials for deprescribing: an environmental scan. Br J Clin Pharmacol 2019 Jul 07;85(7):1396-1406 [FREE Full text] [doi: 10.1111/bcp.13912] [Medline: 30848837]
18. Vivekanantham A, Protheroe J, Muller S, Hider S. Evaluating on-line health information for patients with polymyalgia rheumatica: a descriptive study. BMC Musculoskelet Disord 2017 Jan 26;18(1):43 [FREE Full text] [doi: 10.1186/s12891-017-1416-5] [Medline: 28122554]
19. Harris VC, Links AR, Hong P, Walsh J, Schoo DP, Tunkel DE, et al. Consulting Dr. Google: Quality of online resources about tympanostomy tube placement. Laryngoscope 2018 Dec;128(2):496-501. [doi: 10.1002/lary.26824] [Medline: 28842989]
20. Maciolek KA, Jarrard DF, Abel EJ, Best SL. Systematic assessment reveals lack of understandability for prostate biopsy online patient education materials. Urology 2017 Nov;109:101-106. [doi: 10.1016/j.urology.2017.07.037] [Medline: 28780302]
21. Siddhanamatha HR, Heung E, Lopez-Olivo MD, Abdel-Wahab N, Ojeda-Prias A, Willcockson I, et al. Quality assessment of websites providing educational content for patients with rheumatoid arthritis. Semin Arthritis Rheum 2017 Jun;46(6):715-723. [doi: 10.1016/j.semarthrit.2017.01.006] [Medline: 28258768]
22. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer A, Kindig D, editors. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
23. Chin J, Moeller D, Johnson J, Duwe EA, Graumlich JF, Murray MD, et al. A multi-faceted approach to promote comprehension of online health information among older adults. Gerontologist 2018 Jul 13;58(4):686-695 [FREE Full text] [doi: 10.1093/geront/gnw254] [Medline: 28329840]
24. Boudewyns V, O'Donoghue AC, Kelly B, West SL, Oguntimein O, Bann CM, et al. Influence of patient medication information format on comprehension and application of medication information: a randomized, controlled experiment. Patient Educ Couns 2015 Jul 10;98(12):1592-1599. [doi: 10.1016/j.pec.2015.07.003] [Medline: 26198545]
25. Morrow D, Weiner M, Young J, Steinley D, Deer M, Murray M. Improving medication knowledge among older adults with heart failure: a patient-centered approach to instruction design. Gerontologist 2005 Aug;45(4):545-552. [doi: 10.1093/geront/45.4.545] [Medline: 16051918]
26. Pander Maat H, Lentz L. Improving the usability of patient information leaflets. Patient Educ Couns 2010 Jul;80(1):113-119. [doi: 10.1016/j.pec.2009.09.030] [Medline: 19854022]

27. Cedars BE, Cohen AJ, Fergus KB, Baradaran N, Ndoye M, Kamal P, et al. Qualitative analysis of the content found in online discussion boards for urethral stricture disease and urethroplasty. Urology 2019 Aug;130:155-161. [doi: 10.1016/j.urology.2019.03.033] [Medline: 31129196]
28. Teaford D, McNiesh S, Goyal D. New mothers' experiences with online postpartum forums. MCN Am J Matern Child Nurs 2019;44(1):40-45. [doi: 10.1097/NMC.0000000000000489] [Medline: 30444739]
29. Castaneda P, Sales A, Osborne NH, Corriere MA. Scope, themes, and medical accuracy of eHealth peripheral artery disease community forums. Ann Vasc Surg 2019;54:92-102. [doi: 10.1016/j.avsg.2018.09.004] [Medline: 30267913]
30. Bae SS, Baxter S. YouTube videos in the English language as a patient education resource for cataract surgery. Int Ophthalmol 2018 Oct 28;38(5):1941-1945. [doi: 10.1007/s10792-017-0681-5] [Medline: 28849436]
31. Pedersen MM, Fink P, Kasch H, Frostholm L. Development of an internet-delivered educational video for acute whiplash injuries. Pilot Feasibility Stud 2019 Apr 27;5(1):60 [FREE Full text] [doi: 10.1186/s40814-019-0445-2] [Medline: 31061715]
32. Finnegan G, Holt D, English PM, Glismann S, Thomson A, Salisbury DM, et al. Lessons from an online vaccine communication project. Vaccine 2018 Oct 22;36(44):6509-6511. [doi: 10.1016/j.vaccine.2018.05.007] [Medline: 29921491]
33. Button K, Nicholas K, Busse M, Collins M, Spasić I. Integrating self-management support for knee injuries into routine clinical practice: TRAK intervention design and delivery. Musculoskelet Sci Pract 2018 Feb;33:53-60 [FREE Full text] [doi: 10.1016/j.msksp.2017.11.002] [Medline: 29172113]
34. Vogel H, Schaaf W, Jacob M. [Video-assisted patient education in anesthesia : how do medical users assess the procedure?]. Anaesthesist 2018 Nov 9;67(11):829-836. [doi: 10.1007/s00101-018-0496-4] [Medline: 30302517]
35. Pithadia DJ, Reynolds KA, Lee EB, Wu JJ. A cross-sectional study of YouTube videos as a source of patient information about topical psoriasis therapies. J Dermatolog Treat 2019:1-4. [doi: 10.1080/09546634.2019.1597247] [Medline: 30885022]
36. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. Obes Surg 2019 May;29(5):1477-1484. [doi: 10.1007/s11695-019-03738-2] [Medline: 30706318]
37. Erdem H, Sisik A. The reliability of bariatric surgery videos in YouTube platform. Obes Surg 2018 Mar 5;28(3):712-716. [doi: 10.1007/s11695-017-2911-3] [Medline: 28875396]
38. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
39. Kwok TM, Singla AA, Phang K, Lau AY. YouTube as a source of patient information for varicose vein treatment options. J Vasc Surg Venous Lymphat Disord 2017 Mar;5(2):238-243. [doi: 10.1016/j.jvsv.2016.10.078] [Medline: 28214492]
40. Bademci M, Yazman S, Güneş T, Ocakoglu G, Tayfur K, Gokalp O. YouTube as a potential source of information on deep venous thrombosis. Phlebology 2017 Sep 12;32(8):543-547. [doi: 10.1177/0268355516673795] [Medline: 27738240]
41. Christensen C, Redd D, Lake E, Shipman JP, Aiono H, Altizer R, et al. Doodle Health: A crowdsourcing game for the co-design and testing of pictographs to reduce disparities in healthcare communication. AMIA Annu Symp Proc 2017;2017:585-594 [FREE Full text] [Medline: 29854123]
42. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an internet-based chronic pain self-management program. Int J Med Inform 2017 Jan;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]
43. Lüchtenberg M, Kuhli-Hattenbach C, Sinangin Y, Ohrloff C, Schalnus R. Accessibility of health information on the internet to the visually impaired user. Ophthalmologica 2008;222(3):187-193. [doi: 10.1159/000126082] [Medline: 18497528]
44. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mhealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
45. Noel W, Bosc R, Jabbour S, Kechichian E, Hersant B, Meningaud J. Smartphone-based patient education in plastic surgery. Ann Plast Surg 2017 Dec;79(6):529-531. [doi: 10.1097/SAP.0000000000001241] [Medline: 29053521]
46. 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 Aug;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
47. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
48. Schulz PJ, Rubinell S, Hartung U. An internet-based approach to enhance self-managementof chronic low back pain in the italian-speaking population of Switzerland: results from a pilot study. Int J Public Health 2007 Oct;52(5):286-294. [doi: 10.1007/s00038-007-5127-9] [Medline: 18030944]
49. Zufferey MC, Schulz PJ. Self-management of chronic low back pain: an exploration of the impact of a patient-centered website. Patient Educ Couns 2009 Oct;77(1):27-32. [doi: 10.1016/j.pec.2009.01.016] [Medline: 19321286]
50. Hagerman C, Bellini P, Davis K, Hoffman RM, Aaronson DS, Leigh DY, et al. Physicians' perspectives on the informational needs of low-risk prostate cancer patients. Health Educ Res 2017 Apr 01;32(2):134-152 [FREE Full text] [doi: 10.1093/her/cyx035] [Medline: 28380628]
51. Kim MI, Ladenson P, Johnson KB. A situational approach to the design of a patient-oriented disease-specific knowledge base. Proc AMIA Symp 2002:385-389 [FREE Full text] [Medline: 12463852]

52. Meppelink CS, Smit EG, Buurman BM, van Weert JCM. Should we be afraid of simple messages? The effects of text difficulty and illustrations in people with low or high health literacy. Health Commun 2015 Sep 15;30(12):1181-1189. [doi: 10.1080/10410236.2015.1037425] [Medline: 26372031]
53. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Edu Psychol 2003 Mar;38(1):43-52. [doi: 10.1207/S15326985EP3801_6]
54. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. Syst Rev 2018 Dec 14;7(1):27 [FREE Full text] [doi: 10.1186/s13643-018-0690-0] [Medline: 29444711]
55. Boyde M, Tuckett A, Peters R, Thompson DR, Turner C, Stewart S. Learning style and learning needs of heart failure patients (The Need2Know-HF patient study). Eur J Cardiovasc Nurs 2009 Dec;8(5):316-322. [doi: 10.1016/j.ejcnurse.2009.05.003] [Medline: 19520614]
56. Hagenhoff BD, Feutz C, Conn VS, Sagehorn KK, Moranville-Hunziker M. Patient education needs as reported by congestive heart failure patients and their nurses. J Adv Nurs 1994 Apr;19(4):685-690. [doi: 10.1111/j.1365-2648.1994.tb01139.x] [Medline: 8021389]
57. Wehby D, Brenner PS. Perceived learning needs of patients with heart failure. Hear Lung J Acute Crit Care 1999 Jan;28(1):31-40. [doi: 10.1016/s0147-9563(99)70041-5]
58. Clark JC, Lan VM. Heart failure patient learning needs after hospital discharge. Appl Nurs Res 2004 Aug;17(3):150-157. [doi: 10.1016/j.apnr.2004.06.009] [Medline: 15343548]
59. Kiliç B, Sütçü Çiçek H. Assessment of educational needs of heart failure patients in intensive care units: qualitative study. Intensive Care Med Exp 2015 Oct 1;3(S1):A208. [doi: 10.1186/2197-425x-3-s1-a208]
60. Solomon P, Salbach NM, O'Brien KK, Nixon S, Worthington C, Baxter L, et al. Collaborative development of an educational resource on rehabilitation for people living with HIV. Disabil Rehabil 2018 Nov 12;40(22):2671-2676. [doi: 10.1080/09638288.2017.1347719] [Medline: 28701083]
61. Liu X, Shi Y, Willis K, Wu C, Johnson M. Health education for patients with acute coronary syndrome and type 2 diabetes mellitus: an umbrella review of systematic reviews and meta-analyses. BMJ Open 2017 Oct 16;7(10):e016857 [FREE Full text] [doi: 10.1136/bmjopen-2017-016857] [Medline: 29042383]
62. Boyer C, Selby M, Appel R. The health on the net code of conduct for medical and health web sites. In: Cesnik B, McCray A, Scherrer J, editors. MEDINFO '98 - Vol 52 - Studies in Health Technology and Informatics. Amsterdam: IOS Press; 1998.
63. Priyanka P, Hadi Y, Reynolds G. Analysis of the patient information quality and readability on Esophagogastroduodenoscopy (EGD) on the internet. Am J Gastroenterol 2018;113(Supplement):289-290. [doi: 10.14309/00000434-201810001-00504]
64. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2016 Dec 04;30(5_suppl):106-114. [doi: 10.1177/0272989x10377661]
65. Volk RJ, Llewellyn-Thomas H, Stacey D, Elwyn G. Ten years of the International Patient Decision Aid Standards Collaboration: evolution of the core dimensions for assessing the quality of patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S1 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S1] [Medline: 24624947]
66. Discern online. URL: http://www.discern.org.uk/ [accessed 2020-03-24]
67. Flesch–Kincaid readability tests. Wikipedia. URL: https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests#Flesch%E2%80%93Kincaid_grade_level [accessed 2020-03-24]
68. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
69. Improve video performance with Shareablee's new Video Power Index (VPI). Shareablee. 2020. URL: https://www.shareablee.com/blog/2017/06/27/improve-video-performance-with-shareablee-s-new-video-power-index-vpi [accessed 2020-05-05]
70. Laversin S, Boujard V, Gaudinat A, Simonet M, Boyer C. Improving the transparency of health information found on the internet through the honcode: a comparative study. In: Moen A, Andersen S, Aarts J, Hurlen P, editors. User Centred Networked Health Care - Volume 169 - Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2011:654-658.
71. Sepucha KR, Abhyankar P, Hoffman AS, Bekker HL, LeBlanc A, Levin CA, et al. Standards for UNiversal reporting of patient Decision Aid Evaluation studies: the development of SUNDAE checklist. BMJ Qual Saf 2018 May 21;27(5):380-388 [FREE Full text] [doi: 10.1136/bmjqs-2017-006986] [Medline: 29269567]
72. Knerr S, Wernli KJ, Leppig K, Ehrlich K, Graham AL, Farrell D, et al. A web-based personalized risk communication and decision-making tool for women with dense breasts: Design and methods of a randomized controlled trial within an integrated health care system. Contemp Clin Trials 2017 May;56:25-33 [FREE Full text] [doi: 10.1016/j.cct.2017.02.009] [Medline: 28257920]
73. Krol MW, de Boer D, Delnoij DM, Rademakers JJ. The Net Promoter Score--an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109. [doi: 10.1111/hex.12297] [Medline: 25345554]
74. Kannan VD, Veazie PJ. Who avoids going to the doctor and why? Audience segmentation analysis for application of message development. Health Commun 2015 Jul 25;30(7):635-645. [doi: 10.1080/10410236.2013.878967] [Medline: 25062466]

75. Gomez A, Loar R, England Kramer A. The impact of market segmentation and social marketing on uptake of preventive programmes: the example of voluntary medical male circumcision. A literature review. Gates Open Res 2018 Dec 11;2:68 [FREE Full text] [doi: 10.12688/gatesopenres.12888.1] [Medline: 31131368]
76. Jaffer U, Vaughan-Huxley E, Standfield N, John NW. Medical mentoring via the evolving world wide web. J Surg Educ 2013 Jan;70(1):121-128. [doi: 10.1016/j.jsurg.2012.06.024] [Medline: 23337681]

## Abbreviations

**HONcode:** Health on the Net Foundation Code of Conduct
**IPDAS:** International Patient Decision Aid Standards
**SUNDAE:** Standards for Universal Reporting of Patient Decision Aid Evaluation Studies

---

*Edited by G Eysenbach, R Kukafka; submitted 07.05.20; peer-reviewed by M Nomali, R Jamison, E Ross; comments to author 15.06.20; revised version received 06.08.20; accepted 02.11.20; published 23.09.21*

*Please cite as:*
*Oktay LA, Abuelgasim E, Abdelwahed A, Houbby N, Lampridou S, Normahani P, Peters N, Jaffer U*
*Factors Affecting Engagement in Web-Based Health Care Patient Information: Narrative Review of the Literature*
*J Med Internet Res 2021;23(9):e19896*
*URL: https://www.jmir.org/2021/9/e19896*
*doi: 10.2196/19896*
*PMID:*