Original Paper

The Effect of Fear of Infection and Sufficient Vaccine Reservation Information on Rapid COVID-19 Vaccination in Japan: Evidence From a Retrospective Twitter Analysis

Qian Niu[1*], MS; Junyu Liu[2*], ME; Masaya Kato[1], MS; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

[1]Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, Japan

[*]these authors contributed equally

Corresponding Author:
Momoko Nagai-Tanima, PhD
Department of Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Kawahara-Cho, Shogoin
Sakyo-Ku
Kyoto, 606-8507
Japan
Phone: 81 075 751 3964
Email: tanima.momoko.8s@kyoto-u.ac.jp

Abstract

Background: The global public health and socioeconomic impacts of the COVID-19 pandemic have been substantial, rendering herd immunity by COVID-19 vaccination an important factor for protecting people and retrieving the economy. Among all the countries, Japan became one of the countries with the highest COVID-19 vaccination rates in several months, although vaccine confidence in Japan is the lowest worldwide.

Objective: We attempted to find the reasons for rapid COVID-19 vaccination in Japan given its lowest vaccine confidence levels worldwide, through Twitter analysis.

Methods: We downloaded COVID-19–related Japanese tweets from a large-scale public COVID-19 Twitter chatter data set within the timeline of February 1 and September 30, 2021. The daily number of vaccination cases was collected from the official website of the Prime Minister's Office of Japan. After preprocessing, we applied unigram and bigram token analysis and then calculated the cross-correlation and Pearson correlation coefficient (r) between the term frequency and daily vaccination cases. We then identified vaccine sentiments and emotions of tweets and used the topic modeling to look deeper into the dominant emotions.

Results: We selected 190,697 vaccine-related tweets after filtering. Through n-gram token analysis, we discovered the top unigrams and bigrams over the whole period. In all the combinations of the top 6 unigrams, tweets with both keywords "reserve" and "venue" showed the largest correlation with daily vaccination cases ($r=0.912$; $P<.001$). On sentiment analysis, negative sentiment overwhelmed positive sentiment, and fear was the dominant emotion across the period. For the latent Dirichlet allocation model on tweets with fear emotion, the two topics were identified as "infect" and "vaccine confidence." The expectation of the number of tweets generated from topic "infect" was larger than that generated from topic "vaccine confidence."

Conclusions: Our work indicates that awareness of the danger of COVID-19 might increase the willingness to get vaccinated. With a sufficient vaccine supply, effective delivery of vaccine reservation information may be an important factor for people to get vaccinated. We did not find evidence for increased vaccine confidence in Japan during the period of our study. We recommend policy makers to share accurate and prompt information about the infectious diseases and vaccination and to make efforts on smoother delivery of vaccine reservation information.

XSL•FO
RenderX

KEYWORDS

COVID-19; vaccine hesitancy; Japan; social media; text mining

Introduction

COVID-19 has spread worldwide since its first case in December 2019 and has become a public health emergency of international concern [1]. Until September 30, 2021, Japan experienced 5 waves of the COVID-19 pandemic [2]. The surge of COVID-19 in Japan occurred during the Tokyo Olympics, bringing the cumulative number of COVID-19 cases to 1,556,998 when the Games finished. However, with the lifting of the fourth national state of emergency on September 30, 2021, the pandemic was effectively contained nationwide, and the number of new confirmed cases abruptly decreased. The high vaccination rate in Japan was considered to have caused a decline in the community infections during the fifth wave [3].

A high vaccination rate is thought to be promoted by high vaccine confidence [4]. According to the US Centers of Disease Control and Prevention, "Vaccine confidence is the belief that vaccines work, are safe, and are part of a trustworthy medical system" [5]. A global survey that did not include Japan showed that the potential acceptance of a COVID-19 vaccine largely varied among countries [6]. Japan ranks among the countries with the lowest vaccine confidence worldwide according to a survey in 2020 [7]. Another survey conducted before large-scale vaccination in Japan indicated that Japan ranked last with regard to confidence in COVID-19 vaccines among 15 countries [8]. Gordon and Reich [9] explained the historical reasons for low vaccine confidence in Japan. Kunitoki et al [10] proposed that barriers to vaccine access and use mainly result from effective public communication and called for rebuilding vaccine confidence in Japan.

However, Japan's speed of vaccination has been impressive since large-scale vaccination was opened up (May 24, 2021). Japan's first-dose vaccination rate was approximately 6.8% by June 1, 2021, and over 70% of the population accepted at least one dose until September 30, 2021 [11]. Notably, vaccination was not mandatory and was administered only with the recipient's consent. A survey of multiple countries reported the coexistence of a high level of uncertainty about the safety of COVID-19 vaccines and a high willingness to get vaccinated [12], which indicates that Japan may not be a special case. The reason for the contradiction between the rapid growth in the COVID-19 vaccination rate and low vaccine confidence in Japan is worth studying and maybe instructive for propelling worldwide vaccination against infectious diseases.

Twitter is a widespread social media platform that has attracted the increasing attention of public health researchers because of its advantages of large amounts, real-time availability, and ease of public searching and access [13]. With a large amount of real-time COVID-19–related posts, Twitter has been widely used for public opinion mining toward COVID-19 during the pandemic, providing policy makers with substantiated evidence [12,14,15]. Lyu et al [14] reported the trend of topics and sentiments of English tweets for approximately 11 months since the World Health Organization declared COVID-19 a pandemic.

Yousefinaghani et al [12] reported the dominance of positive sentiments and more vaccine objection and hesitancy than vaccine interest. Huangfu et al [15] reported the results of topic modeling and sentiment analysis of tweets between December 8, 2020, and April 8, 2021. Eibensteiner et al [16] reported willingness to vaccinate despite the safety concerns of vaccines, according to a survey on a Twitter poll. Besides, Twitter is the most popular social media platform in Japan [17], owning 58.2 million users as of October 2021 [18], making Twitter analysis more powerful for COVID-19 research in Japan. A Twitter analysis by Niu et al [19] reported that the Japanese public's negative sentiment overwhelmed the positive sentiment toward the COVID-19 vaccine before and at the beginning of the large-scale vaccination campaign.

This retrospective study aimed to identify public sentiments and concerns associated with rapid COVID-19 vaccination in Japan. We hypothesized that the increase in vaccination rates might be due to subjective factors including increased public confidence in vaccines (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery reservation–related vaccine information (O2). To test these hypotheses, we collected vaccine-related tweets posted between February 1 and September 30, 2021. Then, we preprocessed the collected tweets and conducted a unigram token analysis, sentiment analysis, and topic modeling.

Methods

Overview

In previous works of large-scale Twitter analyses, after preprocessing, there are mainly 4 types of natural language processing (NLP) methods: n-gram token analysis [12,15,20,21], sentiment analysis [12,14,15,20-25], topic modeling [12,14,15,20,22-25], and geographical analysis [22,24]. The geographical analysis is less important in our work because the range of our research is a whole country instead of subareas. In this work, we followed previous works in applying n-gram token analysis, sentiment analysis, and topic modeling. Code in this work will be shared on the web [26].

Data Collection and Preprocessing

The data used in this study were obtained from a large-scale public COVID-19 Twitter chatter data set [27] updated by the Georgia State University's Panacea Lab. The data set provided the IDs, posting time, and the languages of all the tweets were provided in the data set. We downloaded COVID-19–related Japanese tweets between February 1, 2021, the month the first person was vaccinated, and September 30, 2021, when the first-dose vaccination rate exceeded 70%. In addition, data on the number of vaccination cases were collected from the official website of the Prime Minister's Office of Japan (PMOJ) [28].

The downloaded tweets were then cleaned and processed. Retweets were filtered using the Python package tweepy. Tweets that included no keywords related to vaccines were deleted. The keywords used in the filtering are listed in Multimedia Appendix

1. It is worth noting that the three vaccine brands (Pfizer, Moderna, and AstraZeneca) that were approved by the Japanese government were included in the keywords. Other vaccine brands were excluded because we attempted to focus more on the brands adopted in the vaccination process. Frequent misspellings (eg, "Modelna") was also included in the keywords. Weblinks, special characters, emojis, and "amp" (ampersands) were removed, and all full-width English characters were converted to half-width, lowercase characters.

For convenience, all Japanese words in our results were directly presented in English translations. The English-Japanese translation table is provided in Multimedia Appendix 1. In order to minimize the influence of difference between languages, all the translations in our results were carried out as the last step by directly replacing the Japanese words in the graphs with corresponding English words; therefore, they would not influence the statistical results.

Unigram and Bigram Token Analysis

Tokenization is necessary before many other NLP tasks, especially for many non-Latin languages, such as Japanese. We removed the predefined English and Japanese stop words in the Python packages NLTK [29] and SpaCy [30] and tokenized all collected vaccine-related tweets using the Python package SpaCy into unigrams or bigrams for statistical analysis, as reported by Kwok et al [27]. We sorted the unigram tokens or bigram tokens in descending order of term frequency over the entire period. Similar to Liu et al [24], we used the pruned exact linear time (PELT) algorithm [31] to find the first change point of the term frequency. Unigrams before the first change point were regarded as top unigrams, and the term frequencies of the unigrams after the change point were significantly lower than those of the top unigrams. Similar processes were carried out for bigrams. To eliminate the difference in the number of days between months, the monthly term frequency was defined by dividing the total term frequency by the number of days each month for each top unigram or bigram.

Correlation coefficients were widely used in social media analysis. In Google Trends analysis, correlations were calculated between reported cases of infectious disease and the trends of search for relevant keywords. In Twitter analyses, correlations between the daily cases of infection or death and the number of related tweets or sentiment scores, were also investigated [24,25]. In this work, correlation analyses were adopted to find out the factors from the top unigrams that are most related to the COVID-19 vaccination campaign. We first calculated the cross-correlations between the number of tweets containing the top unigrams or bigrams and the vaccination cases and then observed the time lags when maximum cross-correlation appeared for each unigram and bigram. Pearson correlation coefficients (r) between top unigrams or bigrams and the vaccination cases were also calculated.

Sentiment Analysis

After n-gram analysis, sentiment analyses were often used to explore the real-time public attitudes in social media analysis related to COVID-19 vaccination, which may reflect the acceptance of COVID-19 vaccines and related policies [12,14,20,22,24]. The trend of negative sentiments may provide potential evidence for vaccine hesitancy [23]. In this work, sentiment analysis was applied to all vaccine-related tweets. Cloud services were used in this study because there were no reliable public models for sentiment analysis in the Japanese language. We selected Amazon Web Services (AWS) for consistency with previous work [29]. The tweets were divided into positive, negative, neutral, or a mixture of positive and negative tweets using the AWS. Fine-grained emotions were also explored using the Japanese version of the NRC Emotion Lexicon [32]. The NRC Emotion Lexicon is a dictionary of words and their associated scores for eight emotions: anticipation, trust, joy, surprise, anger, disgust, fear, and sadness. The positive and negative tweets were tokenized, and the degree of valence (DOV) for the eight emotions was calculated by adding up the scores for the unigrams that appeared in the NRC Emotion Lexicon. Finally, we calculated the daily average DOV by dividing the number of positive and negative tweets on that day to show the trend of each emotion.

Topic Modeling

Topic modeling were applied to identify fine-grained information from tweets of different sentiments [12,15,24]. Based on the sentiment analysis results, we summarized the topics to look deeper into the dominant emotion in the tweets. Latent Dirichlet allocation (LDA) is often used in tweet topic modeling studies [14,15,20,22,23]. In this study, LDA regards tweets as being generated from different topics, and each topic generates tweets with a Dirichlet distribution. A Python package scikit-learn was used to determine the best number of topics. Log likelihood was adopted as the metric for selection, and 5-fold cross correlation was applied to avoid overfitting. As shown in Multimedia Appendix 1, we chose 2 as the number of topics for LDA modeling, which showed the highest log likelihood score. We used scikit-learn for LDA topic modeling and displayed the top 10 keywords and their weights related to each topic. The weights were the pseudocounts of the keywords in a topic. The themes of topics were summarized from the top 10 keywords by 3 volunteers. The volunteers were first asked to work out the themes of the topics independently, and then they had a meeting to finally reach an agreement on the themes.

We then checked the trends of tweets related to different topics. Defining the i-th tweet in all collected tweets as d_i, and the j-th topic of the LDA model as t_j, the probability of a tweet d_i coming from t_j was calculated using the fitted LDA model as p_{ij}. For tweets posted each day, the expectation of the number of tweets generated from topic j was calculated by summing p_{ij} on that day. The ratio between the expected number of tweets generated from each topic was also plotted to show the trend of public attention under dominant emotion.

Ethics Approval

This study used publicly available and accessible tweets collected by Georgia State University's Panacea Lab, allowing free download. We assert that our analysis is compliant with Twitter's usage policy in aggregate form without identifying specific individuals who published the Twitter posts. Furthermore, the number of vaccination cases downloaded from

the PMOJ are open government data. Therefore, the activities described do not meet the requirements of human subject research and did not require review by an institutional review board.

Results

Data Summary

We downloaded 979,636 Japanese tweets posted between February 1 and September 30, 2021, according to the ID and region information in the data set. After filtering, 190,697 vaccine-related tweets were selected. As a result, the total number of vaccine-related tweets increased from 14,758 tweets in February to 34,692 in August and then decreased to 27,824 in September.

Unigram and Bigram Token Analysis

The change point of unigram term frequencies detected by the PELT algorithm was 6, and the top 6 unigrams were Japanese words for "infection," "Japan," "reserve," "Pfizer," "venue," and "mutation." The unigram "side effects," related to the safety of vaccines, ranked eighth overall. The unigrams "infect," "reserve," and "venue" gradually ranked in the top 3 from February to September, as shown in Figure 1.

The change point of bigram term frequencies detected by the PELT algorithm was 5, and the top 5 were Japanese bigrams

for "Astra + Zeneca," "reserve + available," "article + Reuters," "venue + reserve," and "medical-care + workers." The bigrams "reserve + available" and "venue + reserve" ranked in the top-2 from June to September, and the ranking of "Astra + Zeneca" decreased since May, as shown in Figure 2.

Regarding correlation analysis of unigrams, the time lags for "reserve" and "venue" were 0, and the vaccination cases led the number of tweets containing "infection" for 5 days. After calculating r between the daily number of tweets containing each top unigram and vaccination cases, significant r values ($P<.001$) were obtained for all unigrams except "mutation." The largest r value for the daily vaccination cases was from unigrams "infection" ($r=0.746$), "reserve" ($r=0.829$), and "venue" ($r=0.908$). We then checked the daily number of tweets containing all the combinations of the 3 unigrams showing a strong correlation and found the highest r value ($r=0.912$; $P<.001$) for tweets containing both "reserve" and "venue." By randomly selecting 5 days and checking the source of all the tweets on those days, we found that the 95% CI of tweets containing both "reserve" and "venue" posted by official accounts or mainstream media was 96.0%-100%. The trend of tweets containing both unigrams "reserve" and "venue" compared with the daily vaccination cases is shown in Figure 3.

Figure 1. Translation of the top 10 unigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.
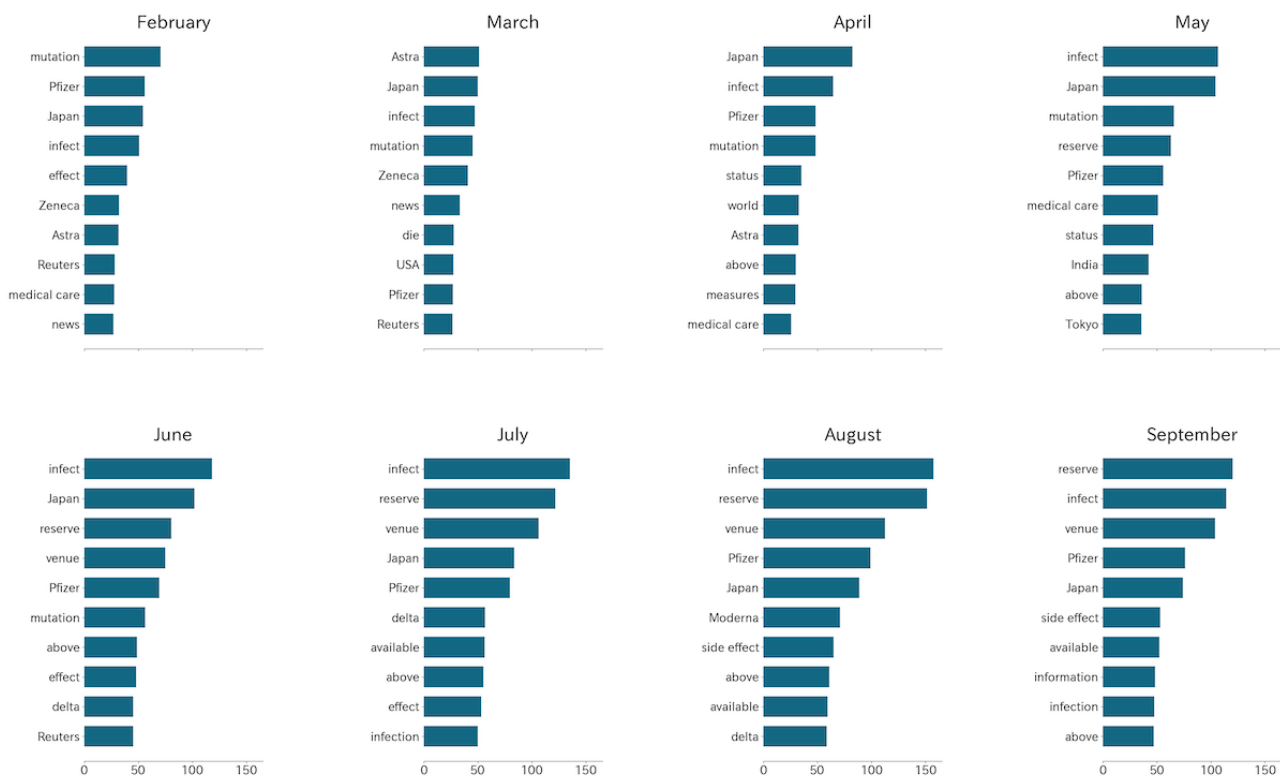
Figure 2. Translation of the top 10 bigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.
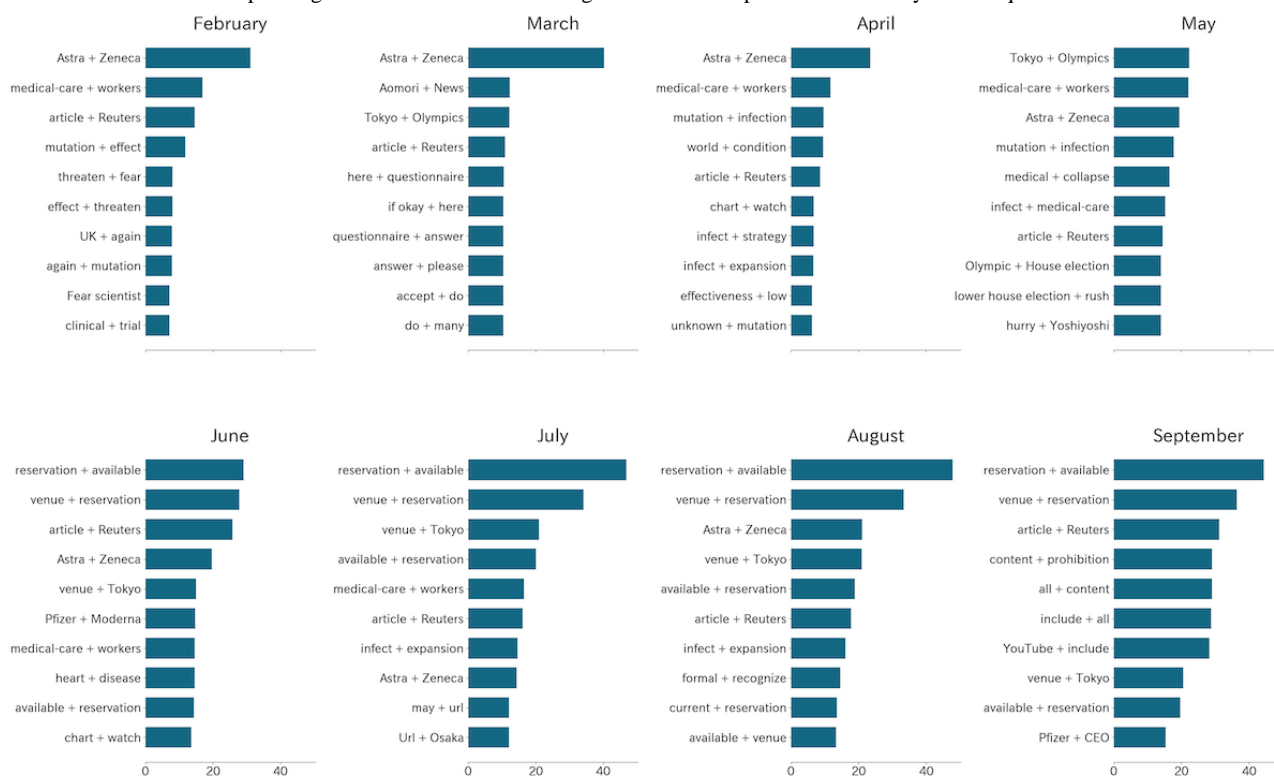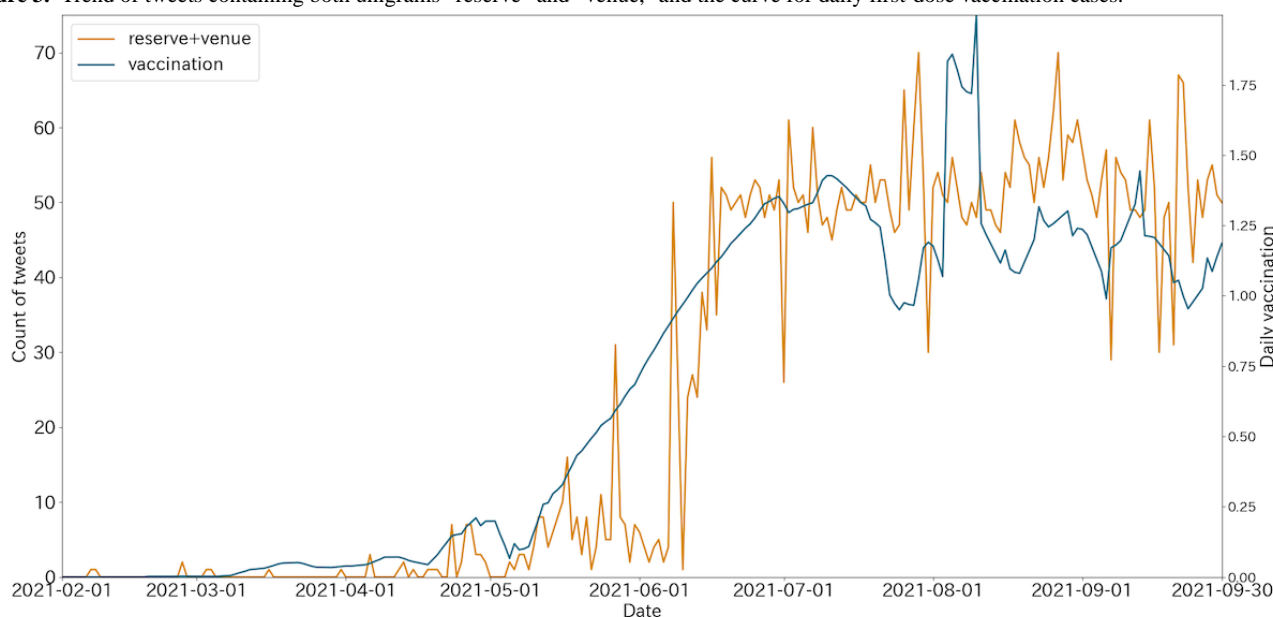

Figure 3. Trend of tweets containing both unigrams "reserve" and "venue," and the curve for daily first-dose vaccination cases.


As for bigrams, the bigram "venue + reserve" overlapped with the unigram analysis and was excluded from this part. The time lags for bigrams "reserve + available" and "article + Reuters" were 0, and vaccination cases led "Astra + Zeneca" and "medical-care + workers" for 116 and 63 days, respectively. The bigrams "reserve + available" and "article + Reuters" had the highest cross-correlations than the others. The bigrams "Astra + Zeneca" ($r=-0.331$), "reserve + available" ($r=0.908$), and "article + Reuters" ($r=0.229$) showed significant correlations ($P<.001$) except for "medical-care + workers" ($r=-0.055$). On manual evaluation by 3 volunteers, we found that 95.4% of the tweets that contain the bigrams "reserve + available" were the same as those of containing the combination of unigrams "venue" and "reserve."

Sentiment Analysis

For all tweets, 4453 (2.3%) were positive, 19,340 (10.1%) were negative, 164,687 (86.4%) were neutral, and 2217 (1.2%) were mixed positive and negative sentiments. A comparison between the daily numbers of tweets marked as positive and negative is shown in Figure 4. Negative sentiments overwhelmed positive sentiments for all days.

Figure 4. Comparison between the daily number of tweets marked positive (orange) and negative (green).
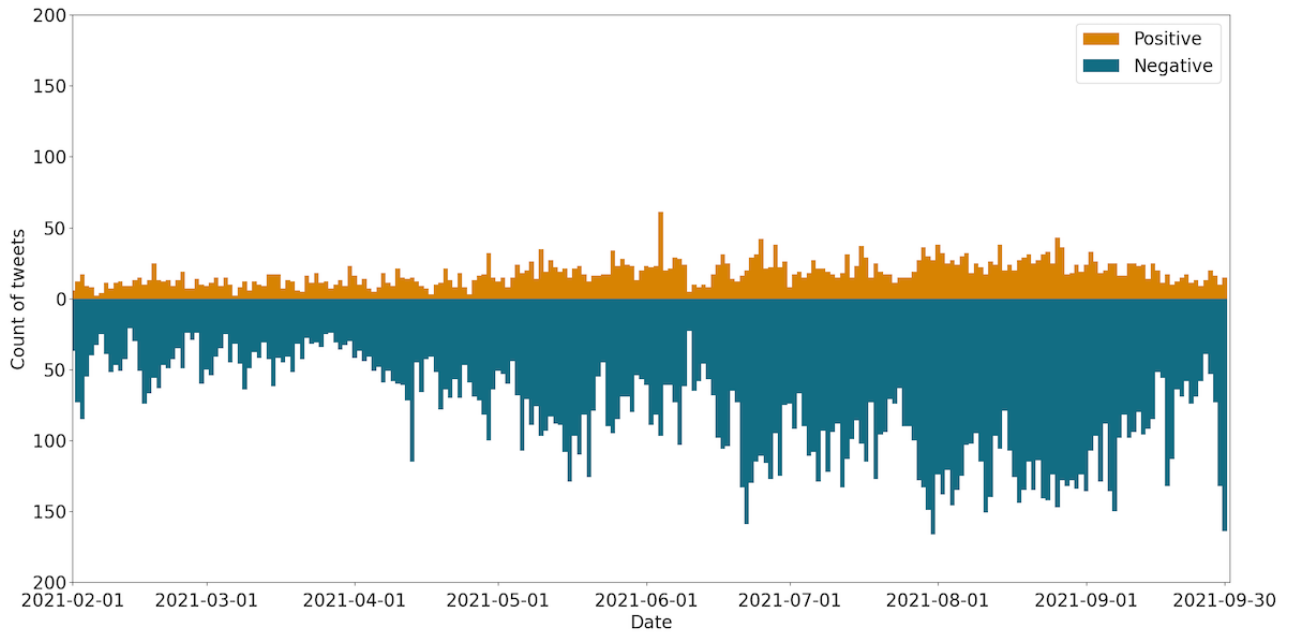
The DOVs for the 8 emotions are shown in Figure 5. The daily average DOV of anger (0.404), disgust (0.268), fear (0.659), sadness (0.486), overwhelmed anticipation (0.163), trust (0.173), joy (0.118), and surprise (0.081). Fear was the dominant emotion during this period. Here, we defined the peaks of emotion as larger than 3 times the daily average DOV for that emotion. Trust peaked (1.114) on February 18, 2021. From May 13 to 18, 2021, there were several peaks of fear.

Figure 5. Daily average degree of valence of 8 emotions in the vaccine-related tweets.
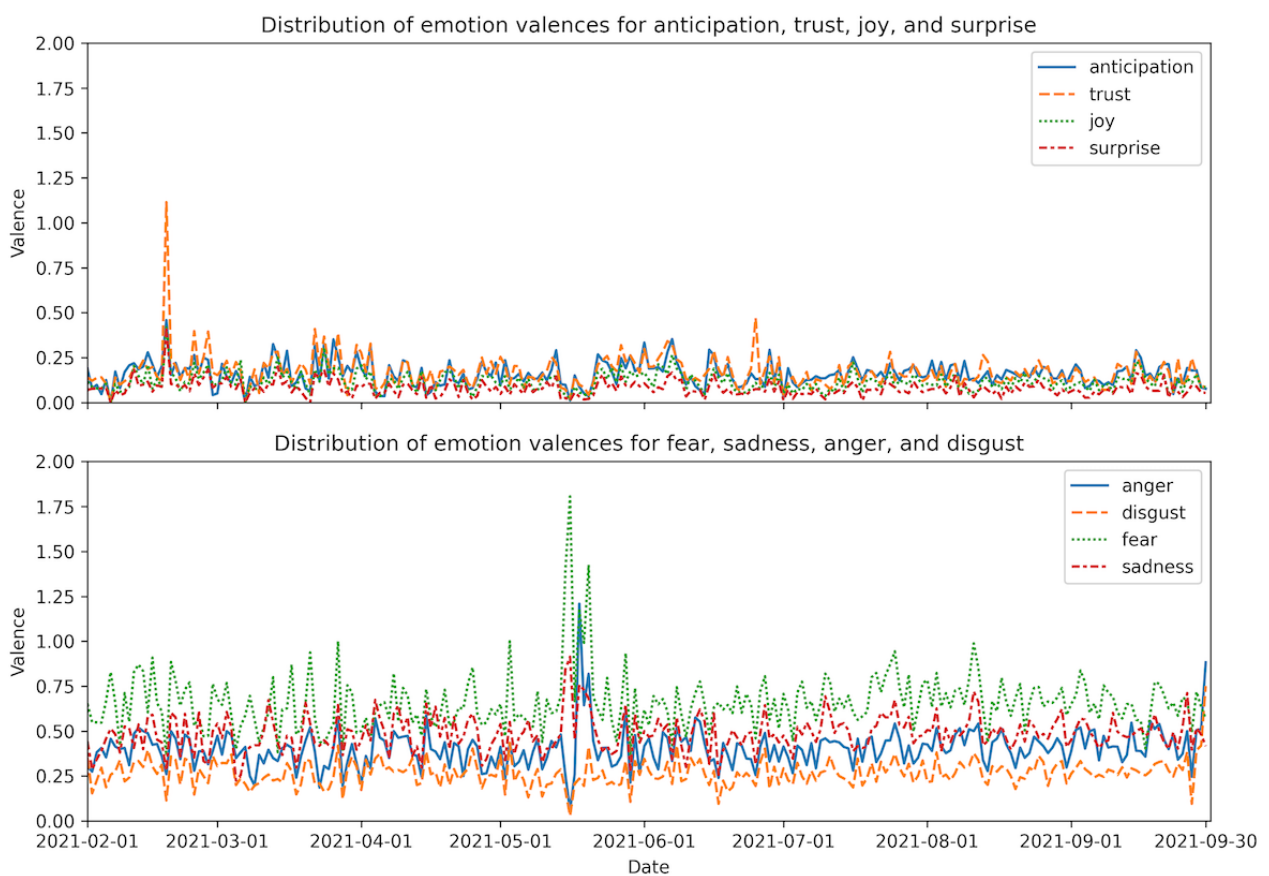
Topic Modeling

The top 10 keywords for each LDA topic are shown in Figure 6. The theme of topic 1 is "infect," and that of topic 2 is "vaccine confidence." It is also noticeable that the weight of "infect" (14,895) in topic-1 was over 3 times that of the second keyword "Japan" (4359), but the weight of "Pfizer" (4348) in topic 2 was only 15.5% larger than the second keyword "die" (3763).

The ratio between the expectation of the number of "infect"-related tweets and "vaccine confidence"–related tweets is shown in Figure 7. The total expectation of the number of tweets generated from topic 1 ("infect," n=30,288) is larger than that generated from topic 2 ("vaccine confidence," n=27,572), and the mean ratio between the expectation of the daily number of tweets generated from topics 1 and 2 is significantly larger than 1 ($P<.01$). On 68.2% of days, the expectation of the number of tweets generated from "infect" was larger than that generated from "vaccine confidence."

Figure 6. Top 10 keywords of 2 topics by latent Dirichlet allocation modeling. The bars represent the weights, which can be regarded as the pseudocounts of the keywords in each topic.
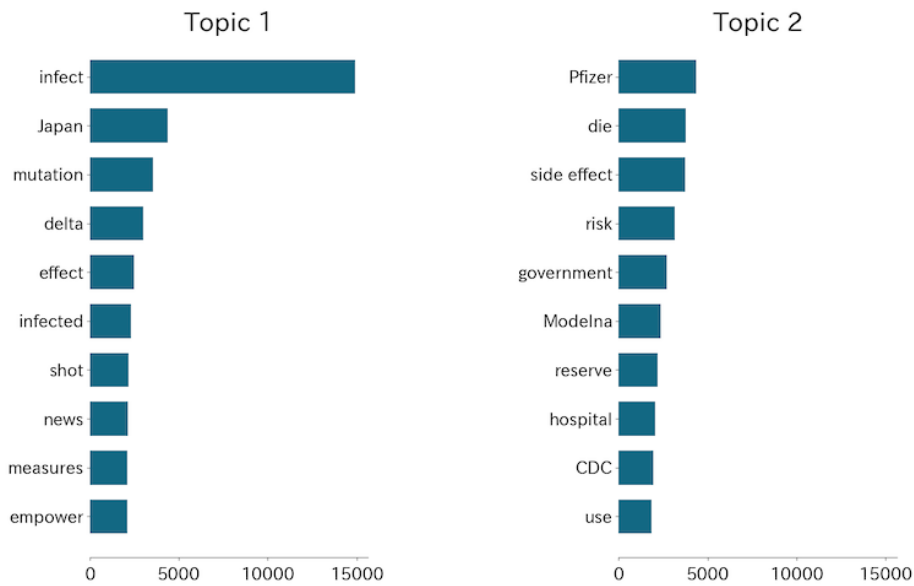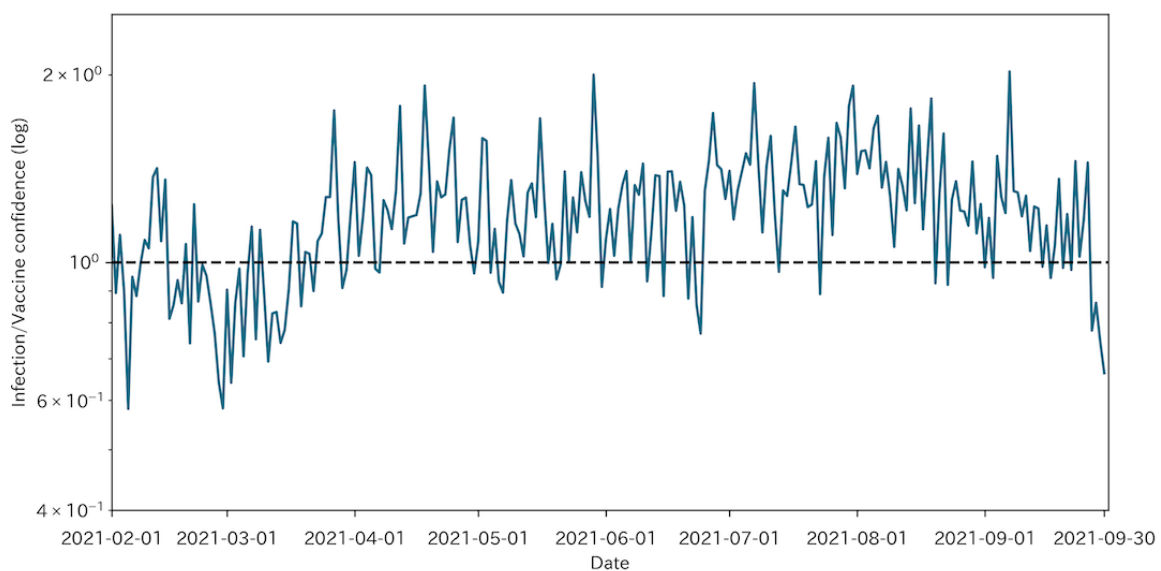

Figure 7. Ratio between the expectation of the number of "infection"-related tweets and "vaccine confidence"–related tweets.

Discussion

Principal Findings

A high vaccination rate is thought to be promoted by high vaccine confidence [16,33-36], but Japan achieved a high vaccination rate in several months, with the lowest vaccine confidence in the world. This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which may be instructive for propelling worldwide vaccination for infectious diseases. Based on previous studies [16,34-37], we hypothesized that subjective factors, including increased

vaccine confidence (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery of reservation-related vaccine information (O2). Our results indicate that hypotheses S2 and O2 might have driven the public to be vaccinated. No evidence supporting hypothesis S1 was found in our results. Evidence for hypothesis O1 can be found in the history of vaccine supply on the official website of the PMOJ (Prime Minister of Japan and his Cabinet) and is not discussed in this paper.

Several results support hypothesis S2. In the unigram token analysis shown in Figure 1, the keyword "infect" ranked among the top 3, except in February, and ranked first from May to August during Japan's fourth and fifth wave infections. The keywords "venue" and "reserve" also ranked up from May. No keywords related to increased vaccine confidence were found. The sentiment analysis shown in Figure 4 showed that negative sentiment overwhelmed positive sentiments, consistent with the results of Chen et al [35] that Japan showed dissatisfaction compared with neighboring countries. Combined with our result that "infect" was the top keyword and "side effect" ranked eighth in the unigram token analysis, our results support that the Japanese public was more concerned about infection than the side effects of COVID-19 vaccines.

More evidence for hypothesis S2 was obtained from the topic modeling results. From the keywords of topic 1 ("infect"), we can see that the public was concerned about the infection and death rate. The mutated virus and empowered cases also led to fear. Willis et al [37] found that less fear of infection may lead to a lower willingness to be vaccinated, which is complementary to our results. From the keywords of topic 2 ("vaccine confidence"), we can see that the side effects of the vaccines were the most concerning, but the following keywords were related to the effectiveness of vaccines on the mutated virus, reservation of vaccines, and medical care conditions. Previous surveys in different countries have indicated that fear of vaccine safety is the key factor for low vaccine acceptance [38,39]. Furthermore, the "side effect" weight in topic 2 was much less than that of "infect" in topic 1. The top keywords in the two topics indicated that people were more concerned about COVID-19 rather than the side effects of vaccines. Bendau et al [40] reported a significant positive correlation between fears of infection and vaccine acceptance and a significant negative correlation between fear of vaccine safety and vaccine acceptance. Therefore, it is important to distinguish the mainstream fear emotion to determine the reason for the high vaccination rate. Figure 7 provides details about the ratio between the expected number of tweets related to "infect" and "vaccine confidence." In most cases, the ratio was larger than 1, indicating that the public was more concerned about infection rather than the safety and effectiveness of vaccines. Higher ratios were observed in April and from July to end-September, which were periods of Japan's fourth and fifth waves of infection. There was also a relatively long period of less than one ratio from mid-February to mid-March, which was the period when the vaccines were less effective against the mutated virus (February 10), severe side effects of the AstraZeneca vaccine were observed (March 12), and several side effects were observed in Japan (February 21, March 7, and March 10).

However, the ratio soon increased because of the fourth wave of infections. This example also proved that fear of infection overcame the vaccine safety concern.

We also provide evidence of a strong relationship between vaccination and hypothesis O2. Bigram analysis in Figure 2 showed that "reservation + available" ranked first since June, shortly after large-scale vaccination started, which might reflect the strong concerns about vaccine reservation by the public. Unigram token analysis in Figure 3 showed that tweets including the keywords "reserve" and "venue" were significantly highly correlated ($r>0.9$; $P<.01$) with the daily number of vaccination cases in Japan, and most of them were from government official accounts. The bigram "reservation + available" also showed a high correlation ($r>0.9$; $P<.01$) with the daily vaccination cases. Because reservation information should always lead to the actual vaccination, this result indicated that in addition to sufficient vaccine supply, reservation information delivery might also be important in large-scale vaccination. Furthermore, the time lag for the maximum cross-correlation was 0, which may indicate the efficiency of the reservation information posted on Twitter. Our results were consistent with Fu's [41] finding that inflexible information systems for vaccine reservation can impair immunization services in the community.

We did not find any evidence for hypothesis S1. Macaraan reported a shift from hesitancy to confidence toward the COVID-19 vaccination program among Filipinos [36]. Okubo [42] reported a shift from hesitancy to confidence but also admitted that the shift might come from the differences in the survey metrics in previous studies [43]. Following these studies, we looked for a similar shift in sentiment or emotions from negative to positive, but negative sentiments overwhelmed positive sentiments as shown in Figure 4, and fear dominates all the emotions in Figure 5. The positive emotions "anticipation," "trust," and "joy" did not increase during the entire period. These two results made it difficult to conclude increased vaccine confidence.

Our results were partially related to the 5 C model (confidence, competence, convenience, calculation, and collective responsibility) measuring vaccine hesitancy [36,44]. Confidence and complacency are two subjective measures that are directly related to individuals. In our work, the LDA theme "vaccine confidence" belonged to "confidence," and "fear of infection" belonged to "complacency." In Japan, fear of infection may drive a high vaccination rate. The delivery of reservation information may be an extension to "convenience," which was previously defined as "physical availability, affordability and willingness-to-pay, geographical accessibility, ability to understand (language and health literacy), and appeal of immunization service affect uptake" [45]. Our work indicates that information about vaccination reservations should also be considered for the convenience of vaccination.

Limitations

We admit that our research might have some potential limitations: (1) the imbalance of the demographics of Twitter users in Japan [46] may cause bias in the results; (2) the status of the user on a certain day (at home or not, other events on that day, etc) may also bias the data set [47]; (3) owing to the lack

of a reliable public model for sentiment analysis in the Japanese language, the cloud service AWS was used for sentiment analysis; (4) filtering keywords may include irrelevant or missing related tweets; (5) antivaccine tweets, especially rumors, were not distinguished or analyzed separately in this study. However, feature works can be combined with classical surveys to train the sentiment analysis model and model to distinguish rumors from tweets to overcome these limitations.

Conclusions

This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which might be instructive for propelling worldwide vaccination toward infectious diseases.

In conclusion, our work indicated that awareness of the danger of COVID-19 increased the willingness to be vaccinated; with a sufficient supply of vaccines, effective reservation information delivery might provide more opportunities for people to be vaccinated. Models measuring vaccine hesitancy might also need to add efficiency in delivering reservation information as a metric. Based on our findings, we recommend public health policy makers and the government to share accurate and prompt information about the infectious diseases and vaccination. Furthermore, efforts on tied cooperation among multilevel relevant organizations and new media operations may help achieve smoother delivery of vaccine reservation information.

Authors' Contributions

QN and JL performed analyses and drafted the manuscript. All authors conceived the study, interpreted the results, and revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The keywords used for the selection of vaccine-related tweets and the corresponding translation; English translations used in our paper and the corresponding original Japanese words; Mean Log likelihood scores for different LDA topic numbers using five-fold cross-validation.

[DOCX File , 29 KB-Multimedia Appendix 1]

References

1. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]

2. Experts warn of 6th wave as COVID cases decrease in Japan but rise overseas. The Mainichi. 2021 Oct 30. URL: https://mainichi.jp/english/articles/20211030/p2a/00m/0na/020000c [accessed 2022-02-22]

3. Fifth wave had fewer big clusters possibly due to vaccinations. The Asahi Shimbun. 2021 Oct 15. URL: https://www.asahi.com/ajw/articles/14461519 [accessed 2022-02-22]

4. Larson H. Vaccine confidence and public trust as drivers of vaccine failure. Int J Infect Dis 2014 Apr;21:50. [doi: 10.1016/j.ijid.2014.03.522]

5. What Is Vaccine Confidence? Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html [accessed 2022-02-22]

6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb 20;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]

7. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. The Lancet 2020 Sep;396(10255):898-908. [doi: 10.1016/s0140-6736(20)31558-0]

8. Mahase E. Covid-19: UK has highest vaccine confidence and Japan and South Korea the lowest, survey finds. BMJ 2021 Jun 04;373:n1439. [doi: 10.1136/bmj.n1439] [Medline: 34088703]

9. Gordon A, Reich MR. The Puzzle of Vaccine Hesitancy in Japan. J Jpn Stud 2021;47(2):411-436. [doi: 10.1353/jjs.2021.0047]

10. Kunitoki K, Funato M, Mitsunami M, Kinoshita T, Reich MR. Access to HPV vaccination in Japan: Increasing social trust to regain vaccine confidence. Vaccine 2021 Oct 01;39(41):6104-6110 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.085] [Medline: 34507858]

11. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E. Coronavirus Pandemic (COVID-19). Our World in Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2022-01-22]

XSL•FO

RenderX

12. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021 Jul;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]

13. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]

14. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]

15. Huangfu L, Mo Y, Zhang P, Zeng DD, He S. COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling. J Med Internet Res 2022 Feb 08;24(2):e31726 [FREE Full text] [doi: 10.2196/31726] [Medline: 34783665]

16. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. J Med Internet Res 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: 10.2196/28973] [Medline: 33872185]

17. Social Media Stats Japan. StatCounter. URL: https://gs.statcounter.com/social-media-stats/all/japan [accessed 2022-01-22]

18. Leading countries based on number of Twitter users as of January 2022. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2022-01-22]

19. Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T, et al. Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. JMIR Infodemiology 2022;2(1):e32335 [FREE Full text] [doi: 10.2196/32335] [Medline: 35578643]

20. Kwok SWH, Vadde SK, Wang G. Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis. J Med Internet Res 2021 May 19;23(5):e26953 [FREE Full text] [doi: 10.2196/26953] [Medline: 33886492]

21. Yang X, Sornlertlamvanich V. Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis. 2021 Presented at: 2021 International Electronics Symposium (IES); September 29-30, 2021; Surabaya. [doi: 10.1109/ies53407.2021.9594036]

22. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, et al. Public Perception of COVID-19 Vaccines on Twitter in the United States. medRxiv Preprint posted online October 18, 2021. [FREE Full text] [doi: 10.1101/2021.10.16.21265097] [Medline: 34704100]

23. Monselise M, Chang C, Ferreira G, Yang R, Yang CC. Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. J Med Internet Res 2021 Oct 21;23(10):e30765 [FREE Full text] [doi: 10.2196/30765] [Medline: 34581682]

24. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. Vaccine 2021 Sep 15;39(39):5499-5505 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.058] [Medline: 34452774]

25. Shim J, Ryu K, Lee SH, Cho E, Lee YJ, Ahn JH. Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea. Int J Environ Res Public Health 2021 Jun 18;18(12):6549 [FREE Full text] [doi: 10.3390/ijerph18126549] [Medline: 34207016]

26. COVID_fear. GitHub. URL: https://github.com/juniorliu95/COVID_fear [accessed 2022-06-06]

27. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. arXiv Preprint posted online April 7, 2020. [Medline: 32550247]

28. Prime Minister of Japan and His Cabinet. URL: https://japan.kantei.go.jp/ [accessed 2022-02-22]

29. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Newton, MA: O'Reilly Media, Inc; 2009.

30. Engebreth G. Bug Fixes. In: PHP 8 Revealed. Berkeley, CA: Apress; 2021:87-99.

31. Dorcas Wambui G. The Power of the Pruned Exact Linear Time(PELT) Test in Multiple Changepoint Detection. AJTAS 2015;4(6):581. [doi: 10.11648/j.ajtas.20150406.30]

32. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. Comput Intell 2013;29:465. [doi: 10.1111/j.1467-8640.2012.00460.x]

33. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]

34. Mori H, Naito T. A rapid increase in the COVID-19 vaccination rate during the Olympic and Paralympic Games 2021 in Japan. Hum Vaccin Immunother 2022 Dec 31;18(1):2010440-2010442. [doi: 10.1080/21645515.2021.2010440] [Medline: 34893009]

35. Chen CW, Lee S, Dong MC, Taniguchi M. What factors drive the satisfaction of citizens with governments' responses to COVID-19? Int J Infect Dis 2021 Jan;102:327-331 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.050] [Medline: 33115678]

36. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. PLoS One 2018 Dec 7;13(12):e0208601 [FREE Full text] [doi: 10.1371/journal.pone.0208601] [Medline: 30532274]

37. Willis DE, Andersen JA, Bryant-Moore K, Selig JP, Long CR, Felix HC, et al. COVID-19 vaccine hesitancy: Race/ethnicity, trust, and fear. Clin Transl Sci 2021 Nov;14(6):2200-2207 [FREE Full text] [doi: 10.1111/cts.13077] [Medline: 34213073]

38.   Pugliese-Garcia M, Heyerdahl LW, Mwamba C, Nkwemu S, Chilengi R, Demolis R, et al. Factors influencing vaccine acceptance and hesitancy in three informal settlements in Lusaka, Zambia. Vaccine 2018 Sep 05;36(37):5617-5624 [FREE Full text] [doi: 10.1016/j.vaccine.2018.07.042] [Medline: 30087047]

39.   Robertson E, Reeve KS, Niedzwiedz CL, Moore J, Blake M, Green M, et al. Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study. Brain Behav Immun 2021 May;94:41-50 [FREE Full text] [doi: 10.1016/j.bbi.2021.03.008] [Medline: 33713824]

40.   Bendau A, Plag J, Petzold MB, Ströhle A. COVID-19 vaccine hesitancy and related fears and anxiety. Int Immunopharmacol 2021 Aug;97:107724 [FREE Full text] [doi: 10.1016/j.intimp.2021.107724] [Medline: 33951558]

41.   Fu C. Milestone and challenges: lessons from defective vaccine incidents in China. Hum Vaccin Immunother 2020 Sep 06;16(1):80-80 [FREE Full text] [doi: 10.1080/21645515.2019.1646579] [Medline: 31339795]

42.   Okubo R, Yoshioka T, Ohfuji S, Matsuo T, Tabuchi T. COVID-19 Vaccine Hesitancy and Its Associated Factors in Japan. Vaccines (Basel) 2021 Jun 17;9(6):662 [FREE Full text] [doi: 10.3390/vaccines9060662] [Medline: 34204465]

43.   Machida M, Nakamura I, Kojima T, Saito R, Nakaya T, Hanibuchi T, et al. Acceptance of a COVID-19 Vaccine in Japan during the COVID-19 Pandemic. Vaccines (Basel) 2021 Mar 03;9(3) [FREE Full text] [doi: 10.3390/vaccines9030210] [Medline: 33802285]

44.   Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batouré O, Anya BM, et al. Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future? Hum Vaccin Immunother 2022 Dec 31;18(1):1-3 [FREE Full text] [doi: 10.1080/21645515.2021.1893062] [Medline: 33684019]

45.   MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]

46.   Statista. URL: https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/ [accessed 2022-04-30]

47.   Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Abbreviations

amp: ampersands
AWS: Amazon Web Services
DOV: degree of valence
LDA: latent Dirichlet allocation
NLP: natural language processing
PELT: pruned exact linear time
PMOJ: Prime Minister's Office of Japan

XSL•FO
RenderX

<u>Original Paper</u>

Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-sectional Survey

Sari Kujala[1], PhD; Iiris Hörhammer[2], DSC; Akseli Väyrynen[1], MSc; Mari Holmroos[3], MSS; Mirva Nättiaho-Rönnholm[3], MSS; Maria Hägglund[4], PhD; Monika Alise Johansen[5], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

[3]Kela, The Social Insurance Institution of Finland, Helsinki, Finland

[4]Healthcare Sciences and e-Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[5]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Sari Kujala, PhD
Department of Computer Science
Aalto University
Konemiehentie 2
Espoo, FI-00076
Finland
Phone: 358 503862768
Email: sari.kujala@aalto.fi

Abstract

Background: Patient portals that provide access to electronic health records offer a means for patients to better understand and self-manage their health. Yet, patient access to electronic health records raises many concerns among physicians, and little is known about the use practices and experiences of patients who access their electronic health records via a mature patient portal that has been available for citizens for over five years.

Objective: We aimed to identify patients' experiences using a national patient portal to access their electronic health records. In particular, we focused on understanding usability-related perceptions and the benefits and challenges of reading clinical notes written by health care professionals.

Methods: Data were collected from 3135 patient users of the Finnish My Kanta patient portal through a web-based survey in June 2021 (response rate: 0.7%). Patients received an invitation to complete the questionnaire when they logged out of the patient portal. Respondents were asked to rate the usability of the patient portal, and the ratings were used to calculate approximations of the System Usability Scale score. Patients were also asked about the usefulness of features, and whether they had discussed the notes with health professionals. Open-ended questions were used to ask patients about their experiences of the benefits and challenges related to reading health professionals' notes.

Results: Overall, patient evaluations of My Kanta were positive, and its usability was rated as good (System Usability Scale score approximation: mean 72.7, SD 15.9). Patients found the portal to be the most useful for managing prescriptions and viewing the results of examinations and medical notes. Viewing notes was the most frequent reason (978/3135, 31.2%) for visiting the portal. Benefits of reading the notes mentioned by patients included remembering and understanding what was said by health professionals and the instructions given during an appointment, the convenience of receiving information about health and care, the capability to check the accuracy of notes, and using the information to support self-management. However, there were challenges related to difficulty in understanding medical terminology, incorrect or inadequate notes, missing notes, and usability.

Conclusions: Patients actively used medical notes to receive information to follow professionals' instructions to take care of their health, and patient access to electronic health records can support self-management. However, for the benefits to be realized, improvements in the quality and availability of medical professionals' notes are necessary. Providing a standard information structure could help patients find the information they need. Furthermore, linking notes to vocabularies and other information sources could also improve the understandability of medical terminology; patient agency could be supported by allowing them to add comments to their notes, and patient trust of the system could be improved by allowing them to control the visibility of the professionals' notes.

XSL•FO

RenderX

Introduction

Patient portals that provide access to electronic health records (EHRs) are becoming increasingly common. Such access to EHRs offers the means for patients to better understand personal health issues, treatment plans, and decisions [1], thus supporting personal health management [2] and informing patients between time- and resource-consuming clinic visits or phone appointments [3].

"Open notes," which are clinical notes that are shared with patients [4], can be considered an essential part of any patient-accessible EHR. In some countries, for example, Sweden [5], Norway [6], and Finland [7], nationwide patient-accessible EHR services, including open notes, are offered to most citizens through national patient portals.

Moreover, the majority of studies in recent reviews [3,8,9] highlighted benefits of patient access to EHRs. Patients were satisfied with the communication and engagement with clinicians, as well as better self-care, achieved as a result of patient access [8]. Improved doctor–patient relationships and patient outcomes were also found to be benefits [3].

Despite these benefits, health care professionals often criticize patient access to EHR [10]; patients, on the other hand, would like more doctors to offer access to their notes [11]. Patient access to EHR changes the physician–patient relationship and power dynamic; physicians have raised concerns [10,12] that such access may worry patients, cause misunderstandings, or cause extra work for physicians [13,14]. Physicians have also been worried that patients who find mistakes or errors would call and ask for corrections to notes which would increase the workload for health care [10].

Many studies [15-18] have also reported lower than anticipated levels of patient uptake of EHR access. Thus, in order to realize the potential of such access to support patient self-management, a better understanding of patient practices, motivations, and challenges is necessary. As de Lusignan et al [15] pointed out, there is still a need to understand how web-based access to EHR might be "redesigned to guide and teach patients in a way that promotes self-management and ultimately improves health."

Patient experiences with access to EHRs have often been explored using surveys, whereby patients were asked to rate usability [19] and attitude [5], usefulness [6], ease of use [20,21], and benefits and risks [22]. In addition, Bell et al [23] used a Likert-scale to study how reading notes affected patient–doctor relationships. Qualitative data have also been collected to understand patient views of access to EHRs. Mishra et al [24] included open-ended questions to identify positive and negative themes related to the usefulness, understandability, and worries caused by patient access; Gerard et al [25] used open-ended questions about the value of reading notes and providing feedback on open notes; Rexhepi et al [26] interviewed patients

with cancer and found that patient access helped them prepare for doctor visits and understand their medical issues; and Eriksson-Backa et al [27] conducted focus groups with older adults and identified the uses, enablers, barriers, and behavioral outcomes of the national My Kanta patient portal.

In Finland, My Kanta, a nationwide patient portal, was introduced in 2010 and varied functions were adopted in a step-by-step manner [28]. Since 2015, the My Kanta patient portal has enabled all citizens using public health care services to access their health records and prescriptions, and to renew the latter [28]. The use of My Kanta is very established, with 63% of Finnish adults having accessed the patient portal during the period from 2010 to 2018 [7], and 92% of adults (from 18 to 65 years) used the patient portal in 2021. The most used functions among pharmacy customers were browsing prescription information (97.4%) and health records (96.3%) [20].

The goal of this study was to understand patients' experiences using My Kanta to access their EHRs. While My Kanta has been available for all patients to use for 7 years, little is known about patient use practices and experiences. Thus, we specifically focused on understanding patients' perceptions related to the usability of the patient portal and the benefits and challenges of reading the clinical notes written by health care professionals.

Methods

Design

We conducted a cross-sectional survey to capture patients' experiences using the My Kanta patient portal.

The My Kanta Patient Portal

My Kanta is a web-based patient portal for all residents with a Finnish personal identity number and access to electronic identification. Patients can view their own or their dependents' health data (consisting of records of health care visits, diagnoses, critical risk factors, laboratory tests, x-ray examinations, referrals, health and care plans, and medical certificates, statements [20], and e-prescriptions), request a prescription renewal, and save living wills and organ donation testaments [29].

My Kanta is a part of national Kanta services that integrate and save medical, health, and prescription data for health care providers, citizens, and pharmacies [28]. All public and private health care providers that use electronic patient record systems are obliged by law to send prescription and health data to Kanta services [7]. Health data, test results, and prescriptions can be used by health care units with patient consent [28], which can be given or withdrawn on My Kanta.

According to international benchmarking, My Kanta provided patients and their caregivers with the best access to their health

record data alongside Korea in 2019 [30] and also provided the most functions in 2016 [31]. However, My Kanta does not allow typical patient portal functions, such as appointment booking or communication with health care professionals.

Questionnaire

The web-based questionnaire included 4 open-ended questions and 11 questions with Likert scale or multiple choice response options (Multimedia Appendix 1). The topics of the questions were (1) reasons for logging into the patient portal and whether the visit was successful or not and why; (2) subjective usability of the patient portal; (3) usefulness of the features of the patient portal; (4) the benefits and challenges of reading health care professionals' notes and discussing their notes with them; (5) improvement ideas for the patient portal; (6) guidance on reading the notes; and (7) background information.

To assess perceived usability, a 2-item questionnaire based on the Usability Metric for User Experience [32]—the UMUX-LITE scale [33]—was used. UMUX-LITE scores were transformed, using a corrective regression formula [33], to System Usability Scale scores. The System Usability Scale is the most frequently used questionnaire for measuring the subjective usability of eHealth apps [34]. Borsci et al [35] tested UMUX-LITE with health care professionals and found it to be appropriate for use in the context of health care technology [35].

Open-ended questions about respondents' experiences of the benefits and challenges of reading health care professionals' notes were used in order to collect qualitative data about the most relevant issues from the patients' perspectives. The web-based questionnaire was dynamic; only respondents who reported having read the notes at least once (ie, had actual use experience) were asked the open-ended follow-up questions. If a respondent rated reading the notes as "not useful," they were only asked about challenges (to avoid unnecessarily asking these respondents questions about benefits). The survey was available in both official languages of Finland: Finnish and Swedish.

The questionnaire was reviewed by 2 researchers in the field and 2 experts from the Social Insurance Institution of Finland, which was the organization responsible for developing My Kanta. In addition, we pilot-tested the questionnaire with 3 patients who filled in the questionnaire and simultaneously talked aloud about how they understood the questions. The questionnaire was subsequently revised to clarify wording.

Conducting the Survey

Data were gathered during the period from June 4, 2021 to June 14, 2021 using a web-based questionnaire. Patient users of My Kanta in Finland received an invitation and a link to the questionnaire when they logged out of the patient portal. Thus, all respondents had used the patient portal just before they responded to the questionnaire. Participation was voluntary and anonymous.

Ethics Approval

The study protocol was reviewed and approved by the Ethical Review Board of Aalto University (ethics approval number D/957/03.04/2020 Nordic eHealth for Patients).

Analysis

Descriptive statistics were calculated for quantitative data (respondents' characteristics: age, gender, and portal usage). We performed content analysis (Atlas.ti, version 8.4.5; ATLAS.ti Scientific Software Development GmbH) on the responses to open-ended questions. One researcher first read through the data and used open coding to identify themes in the data without predefined categories. Short sentences were chosen as the analytical unit; themes were defined using in vivo coding, and to ensure that the themes represented the original meaning of the respondents, we used respondents' sentences to label the themes. The number of respondents who mentioned a theme was calculated, and the themes were categorized. A second researcher then reviewed the results. The researchers discussed similarities and differences in themes and combined categories, until a version was agreed upon as the final version.

Results

Respondents

Of 449,922 users who logged in, 3139 users responded to the survey (response rate 0.7%). Most users reported either weekly (889/3112, 28.6%) or monthly use (1120/3112, 36.0%) (Table 1). The frequency of use was comparable to that of My Kanta in May 2019, when users used My Kanta on an average of 2.4 times per month [7]. The proportion of users over the age of 50 years was high (2681/3135, 85.5%). The proportion was 2-fold that in 2021 (44%). Although the frequency of use may vary notably between users, this may suggest overrepresentation of older age groups among respondents.

Table 1. Respondent characteristics (n=3135).

| Characteristic	Respondents, n (%)
Gender (n=3118)	
Female	2104 (67.5)
Male	962 (30.8)
Other	52 (1.7)
Age (years) (n=3115)	
<18	5 (0.2)
18-35	93 (3.0)
36-50	336 (10.8)
51-65	1082 (34.7)
66-75	1173 (37.6)
76-85	395 (12.7)
>85	31 (1.0)
Frequency of use (n=3112)	
Daily	194 (6.2)
Weekly	889 (28.6)
Monthly	1120 (36.0)
Less than once per month	878 (28.2)
First time user	31 (1.0)
Success of the visit (n=3125)	
Yes	2247 (71.9)
No	766 (24.5)
Do not know	112 (3.6)
Device used (n=3053)	
Computer	1836 (60.1)
Smartphone	690 (22.6)
Tablet	522 (17.1)
Something else	5 (0.2)
Has discussed the notes with a health care professional (n=3039)	
Yes	1046 (34.4)
No	1993 (65.6)

Experiences With the Patient Portal

The total mean score for the System Usability Scale approximation was 72.7 (SD 15.9).

The most common reasons for visiting the My Kanta patient portal were viewing medical notes (978/3135, 31.2%), results of examinations (693/3135, 22.1%) or prescriptions (548/3135, 17.5%). Many people also visited the patient portal to renew a prescription (477/3135,15.2%), because there is no other method for renewing prescriptions electronically. At the time of the survey, COVID-19 vaccinations had started in Finland, and many (229/3135, 7.3%) logged into the patient portal to view their vaccination certificates. Other functions were used by only a few respondents (n=6-21). Some users (n=24) tried to use

functions that did not exist, such as making appointments or checking their appointments (n=18), contact health care professionals (n=4), or looking for information about the reason that their prescription had not been renewed (n=2).

The most used functions were also deemed to be the most useful (Figure 1); for example, 96.4% (2511/2605) of users considered prescription renewal and 91.9% (2749/2992) of users considered viewing health care professionals' notes to be very useful or somewhat useful; however, the majority of users also considered less rarely used functions useful, with the lowest percentage (759/1165, 65.2%) of users considering self-reported wellness data, and the highest percentage (1483/1759, 84.3%) of users considering living will to be very useful or somewhat useful.

Figure 1. Usefulness of My Kanta patient portal features.
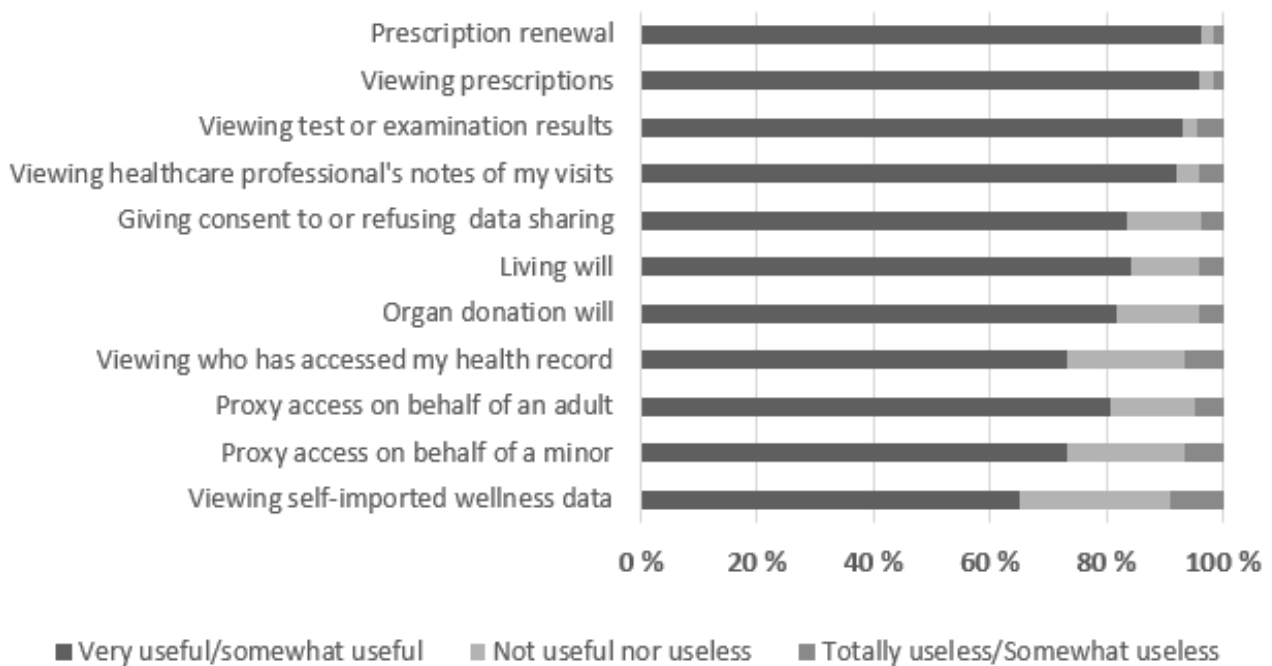
Benefits and Challenges of Reading Notes

Most respondents (2183/3135, 69.6%) answered the open-ended question and mentioned one or more benefits of reading notes (Table 2). Of the perceived benefits, most often respondents stated (560/2178, 25.7%) that notes supported remembering things:

One can recall afterwards what happened in the health care visit and what was discussed about.

Respondents often mentioned that they felt tense or overwhelmed during their appointment, and notes helped in remembering what was said and which instructions were received.

Table 2. Themes of perceived benefits of viewing medical notes.

Benefits	Mentions (n=2178), n (%)
Supports remembering	560 (25.7)
What a doctor or a nurse said	505 (23.2)
The care history	55 (2.5)
Provides information	495 (22.7)
About health and care	223 (15.0)
To check the state of health and remain up-to-date	74 (3.4)
On how I and my disease are perceived	70 (3.2)
On all information concerning myself	43 (2.0)
That is more detailed and was not said during the appointment	34 (1.6)
About what was done during an appointment	30 (1.4)
On diagnoses	21 (1.0)
Convenience of patient portal access	449 (20.6)
Ability to return to view all the saved information	155 (7.1)
Can be checked at leisure	98 (4.5)
No need to call or contact health care	73 (3.4)
Easy of finding information	57 (2.6)
Fast access	56 (2.6)
Clear and reliable information	10 (0.5)
Helps in understanding	339 (15.6)
Own condition or what was said	326 (15.0)
Whether more can be asked if something was unclear	13 (0.6)
Ability to check the notes	234 (10.7)
Identifying potential errors and misunderstandings	142 (6.5)
Asking for error corrections	65 (3.0)
Checking that all essential information was written	21 (1.0)
Increases transparency and reliability	6 (0.3)
Supports self-management	175 (8.0)
Checking the care plan and next steps	63 (2.9)
Following the course of care success	36 (1.7)
Preparing for the next appointment	29 (1.3)
Looking for further information	14 (0.6)
Helps in communicating with health care professionals, learning to express yourself	14 (0.6)
Supports self-care	13 (0.6)
Enables peace of mind	6 (0.3)

Respondents appreciated that notes provided information about their health and care. They were able to follow the course of their care and remain up-to-date. Furthermore, they wanted to identify doctors' perceptions of them and their diseases. Several mentioned that it is important to have all the information concerning themselves:

> *My life and my own information are certainly of primary importance.*

Respondents also noted that the information is provided conveniently in one place, and they can check the information whenever they want. Notes were also perceived as helping them to understand their health conditions and what health care professionals had said during appointments. In addition, many respondents wanted to check the notes to ensure there were no errors or misunderstandings.

Many stated that the reason for accessing the information and remaining up-to-date was to actively self-manage their health. Respondents wanted to be aware of their care plans and to follow

the course of their success. They subsequently prepared themselves for the next appointment and looked for further information related to their condition and care. A few commented that the notes helped in communicating with health care professionals and supported learning to express themselves, and 1046 out of 3135 (33.4%) respondents also discussed the notes with health care professionals.

One-third (1175/3135, 37.5%) of respondents also reported one or more challenges in reading notes (Table 3). The most commonly mentioned challenge was the difficulty in understanding the notes and the medical terminology. For example, one respondent stated:

> Language that I don't understand. Wikipedia may help in translation work, when you don't understand the crucial words.

Many mentioned that they used Google to interpret the unfamiliar terms, codes, and abbreviations, and they wanted plain language to be used instead.

Table 3. Perceived challenges of viewing medical notes.

Challenges	Mentions (n=1175), n (%)
Notes are difficult to understand	707 (60.2)
The medical terminology is difficult to understand	523 (44.5)
Abbreviations are difficult	73 (6.2)
Examination and test results are difficult	44 (3.7)
Notes in general are difficult to understand	44 (3.7)
Diagnoses are not understandable	23 (2.0)
Notes are not available	232 (19.7)
Delay in access	121 (10.3)
Missing information	105 (8.9)
Children's information is not visible	6 (0.5)
Notes are incorrect or inadequate	217 (18.5)
Incorrect information or errors	80 (6.8)
Health care professionals' misinterpretations	28 (2.4)
Imprecise notes	27 (2.3)
Very brief notes	16 (1.4)
Negligent writing	15 (1.3)
Irrelevant or too detailed information	12 (1.0)
Repetition	10 (0.9)
Poor language	9 (0.8)
Wrong language (eg, Finnish instead of Swedish)	7 (0.6)
Too personal	5 (0.4)
Inappropriate	4 (0.3)
Follow-up is unclear	4 (0.3)
Problems with usability	167 (17.4)
Information was difficult to find	85 (7.2)
Errors are difficult or impossible to correct	37 (3.1)
Could be easier to use	25 (2.1)
Disorganized	25 (2.1)
No interactivity	8 (0.7)
The search process is cumbersome	5 (0.4)
Worries about privacy	5 (0.4)
Comparing examination results is difficult	5 (0.4)
Reading on mobile devices is difficult	5 (0.4)
The text is small	4 (0.3)

However, the notes were not always available because there were delays in access and some visits were not recorded or visible. It was mentioned that it could take days or weeks before the notes were available, and some information was not available at all.

Many respondents perceived notes to be incorrect or inadequate. Most commonly, they were seen as having errors—some were not significant, such as a wrong date, but some were more severe, such as having a wrong diagnosis or another patient's information. Respondents described,

> *Mainly the challenge is that the communication has been wrongly recorded or it is misunderstood. People should have possibility to say their views on My Kanta*

and

> *Sometimes there have been erroneous information and diagnoses. For example, a cancer that I don't have.*

Many also reported that the notes differed from what they had experienced themselves. Several also wished for more detailed notes. In contrast, some felt that it was unnecessary to include all personal details that they had mentioned during an appointment or the whole message that they had sent. One person also mentioned that they did not want to talk about certain issues, because they would be recorded and seen by all professionals.

Finally, there were challenges related to the usability of the system. Most commonly, it was mentioned that it was difficult to find information. The information was not always in chronological order, and some examination results were not linked to the appropriate appointments. A few respondents also mentioned that there is no interactivity in the system, and they wanted to comment on the notes or request corrections. Furthermore, it was noted that a patient should receive a notification when new information is available.

Discussion

Principal Results

Respondents evaluated the My Kanta patient portal as useful and usable, which is consistent with the findings of earlier studies [20,21]. The total mean score for the System Usability Scale approximation was 72.7 (SD 15.9), which can verbally be described as good usability, according to Bangor et al [36,37]. Prescription renewal and viewing were indicated to be the most useful functions, but viewing medical notes and the results of examinations were the most frequent reasons for visiting the patient portals, which 91.9% (2749/3135) and 92.9 % (2770/3135) of respondents, respectively, considered useful.

Furthermore, respondents explained in their responses to open-ended questions that they appreciated having access to EHRs and information via a patient portal, which supports earlier findings [6,22]. Because My Kanta has been used nationally for several years, respondents were already familiar with the portal and actively used medical notes to prepare for their communications with health care professionals and to take care of their health.

The qualitative responses provided a rich and versatile description of the benefits of patient access to EHRs. Specifically, reading the notes was described as convenient, because they could be accessed easily and quickly, whenever suitable and at leisure. Therefore, easy access via patient portals may help patients to be engaged in self-management of their health. Reading notes were described as supporting remembering and understanding what health professionals said. They were able to check the state of health and care plans, remain up-to-date, look for further information, prepare for the next appointment, and ask further questions if something was unclear. We suggest that these activities support patients in learning about their disease or care, which motivates them to take care of their health.

Furthermore, reading notes can provide information that is not directly addressed during visits with a health care professional. As previously suggested [38], this may improve patient autonomy by reducing dependence on individual health care professionals and providing the opportunity to consult medical literature or other health care professionals to better understand health status and options for care or treatments.

Many respondents stated that it was important to be able to check the notes to identify potential errors and misunderstandings. They were also interested in professionals' perceptions of their situations. Reading the notes was thus seen to help them understand what health care professionals had said and prepare for the next appointment. Thus, patient access to EHR supports patient–provider communication.

Very few patients were concerned about privacy or felt the notes were too personal or inappropriate. Some patients found incorrect information, and a few mentioned serious errors. It was very rarely mentioned, but a few respondents also felt that the notes included irrelevant information or personal information that was too detailed. Although rarely mentioned, the notes sometimes included information about very personal issues that patients were unwilling to share with all health care personnel. In particular, when a patient portal does not allow patients to correct errors or express their views with a comment, we presume that some patients may feel that their self-determination is violated.

It is noteworthy that respondents did not perceive reading the notes to be harmful per se but that challenges, such as understandability of medical terminology, incorrect or inadequate notes, missing information, or difficulties in finding information, interfered with the benefits of reading the notes. Finnish law requires that professionals' notes are sufficiently comprehensive, clear, and understandable and that only commonly known terms or abbreviations are used [39]. Nevertheless, this is clearly not fulfilled according to the survey results.

In order to realize associated benefits, improvements in the quality and availability of medical professionals' notes are needed. In addition to educating health care professionals, the availability of information can also be supported by providing a standard information structure. Because the information structure was confusing to patients, a standard structure would make finding and reading information easier from patients'

perspectives. It is important that the order of the notes is logical from their point of view and that examination results are clearly linked to corresponding appointments. Linking the notes to vocabularies and other information sources could also improve the understandability of medical terminology without increasing the workload of professionals. In addition, patient agency and trust could be supported by enabling them to add comments to their notes, mark some entries as sensitive, and control the visibility of entries.

Limitations

This was a cross-sectional survey study examining patients' self-reported experiences of the national patient portal in Finland, and the results may not be generalizable to other countries or patient portals. The survey was available only to My Kanta users after logging out of the patient portal. Not all users may have actively logged out the portal or noticed the invitation, which may have contributed to the low response rate. Thus, the results do not represent all My Kanta users or the population of Finland. A similar survey study in Sweden [5] also had a low response rate (0.61%).

In addition, the only demographic information available from the survey was age and gender; health and socioeconomic status of the respondents, literacy, and health literacy remained unknown. It is possible that the survey respondents represented users who were most interested in the patient portal and most capable of using it. Our sample did not include persons who had stopped using the portal or were not able to use it. Thus, nonrespondents may differ in their use of the patient portal and may experience barriers (eg, [40,41]) that were not identified in this study.

Moreover, My Kanta includes many functions that were recently added and, thus, not widely used. Therefore, the usefulness of all the functions could not be reliably evaluated. Furthermore, the portal does not have all the potentially useful functions that users could have experienced. As a few respondents complained, My Kanta does not have much interactivity—patients are not allowed to comment on notes or request corrections in the portal. In addition, the lack of notifications on added content frustrated respondents, because they often logged in to look for notes or test results that were not available yet.

By asking open-ended questions on the benefits and challenges, we improved the reliability of the answers as respondents reported their experiences using their own words and were not guided by having to choose from certain options. Because the number of respondents was high, the data that we collected were rich and versatile. However, respondents may have focused on the most significant benefits and challenges they experienced, and they may not have been able to verbalize those that were more abstract and less obvious. Therefore, we believe that our mixed methods survey study complements previous quantitative studies [5,6,19-23].

Comparison With Prior Work

The main benefits experienced by patients were very similar to those identified in a smaller study [25] conducted at a single institution in the United States over a 12-month pilot period, in which participants reported that reading notes helped them to better remember next steps, provided positive emotions, and gave them faster access and results. The participants in the study [25] also valued the opportunity to correct any possible misunderstandings and give feedback to their providers, which are functions that patients also wished were available on My Kanta. In addition, Rexhepi et al [26] and Pyper et al [42] identified similar benefits, in studies in Sweden and the United Kingdom, respectively. Thus, our study provides further details in understanding self-management practices that patient access to EHR can support.

Moreover, several studies [6,20,24,26,27,42-44] have identified that some parts of medical records are difficult to understand. In addition, Johansen et al [14] found that 25.6% of administrative staff and 15.4% of health care professionals had received feedback from patients or their relatives regarding mistakes or missing information in their EHR. In our study, the number of serious mistakes was seldom mentioned, which was not the case in a recent US survey [45] with 29,656 respondents, in which 1 in 5 patients reported finding a mistake, and 40% perceived the mistake to be serious. It is possible that the number of respondents who found errors would have been higher in our study if we had specifically asked about this. This should be explored in future studies.

Pyper et al [42] also found that patients identified errors and omissions and had differences of opinion when they accessed electronic records for the first time. In this context, our study shows that unclear or inadequate notes are very common even when patients are familiar with use of the EHR and the challenges do not disappear when health care professionals gain experience in conveying information to patients. Thus, the findings support the need for applications that provide explanations of medical terms in EHR notes (eg, [46,47]).

Thus, patients' basic needs and self-management processes seem to be similar regardless of the context—benefits and challenges experienced by patients are remarkably similar across countries, different health care systems, and EHRs.

Conclusions

Our findings indicate that patient access to EHR can support self-management—patients actively used medical notes to understand and remember what health care professionals said and to take care of their health. The challenges interfered with the benefits of reading the notes. In order to realize benefits, improvements in the quality and availability of medical professionals' notes are needed, and patients should be encouraged to discuss their concerns with them. In addition, the availability of information can also be supported by using a standard information structure. Specifically, linking the notes to vocabularies and other information sources could also improve the understandability of medical terminology.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

The questionnaire.
[DOCX File , 15 KB-Multimedia Appendix 1]

References

1. Mahnke M, Nielsen M. eHealth platforms as user–data communicationxamining patients' struggles with digital health data. Nordicom Review 2021;42:58. [doi: 10.2478/nor-2021-0040]

2. Shah SGS, Fitton R, Hannan A, Fisher B, Young T, Barnett J. Accessing personal medical records online: a means to what ends? Int J Med Inform 2015 Feb;84(2):111-118 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.005] [Medline: 25453275]

3. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]

4. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]

5. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]

6. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]

7. Jormanainen V, Parhiala K, Niemi A, Erhola M, Keskimäki I, Kaila M. Half of the Finnish population accessed their own data: comprehensive access to personal health information online is a corner-stone of digital revolution in Finnish health and social care. Finish Journal of eHealth and eWelfare 2019 Nov 02;11(4):298-310. [doi: 10.23996/fjhw.83323]

8. Mold F, de LS, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]

9. Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E. How the use of a patient-accessible health record contributes to patient-centered care: scoping review. J Med Internet Res 2021 Jan 11;23(1):e17655 [FREE Full text] [doi: 10.2196/17655] [Medline: 33427683]

10. Cajander Å, Grünloh C. Electronic health records are more than a work tool: conflicting needs of direct and indirect stakeholders. : AC; 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-13. [doi: 10.1145/3290605.3300865]

11. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]

12. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" patient participation through ehealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]

13. Grünloh C, Cajander Å, Myreteg G. "The record is our work tool!"-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]

14. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]

15. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]

16. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inform Decis Mak 2020 Jul 08;20(1):153 [FREE Full text] [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]

XSL•FO
RenderX

17. Zhao J, Song B, Anand E, Schwartz D, Panesar M, Jackson G, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]

18. Fraccaro P, Vigo M, Balatsoukas P, Buchan I, Peek N. Patient portal adoption rates: a systematic literature review and meta-analysis. 2017 Presented at: MEDINFO: Precision Healthcare through Informatics; August 21-25; Hangzhou, China p. 83. [doi: 10.3233/978-1-61499-830-3-79]

19. Hägglund M, Scandurra I. User evaluation of the Swedish patient accessible electronic health record: System Usability Scale. JMIR Hum Factors 2021 Jul 27;8(3):e24927 [FREE Full text] [doi: 10.2196/24927] [Medline: 34313596]

20. Sääskilahti M, Ahonen R, Timonen J. Pharmacy customers' experiences of use, usability, and satisfaction of a nationwide patient portal: survey study. J Med Internet Res 2021 Jul 16;23(7):e25368 [FREE Full text] [doi: 10.2196/25368] [Medline: 34269687]

21. Lämsä E, Timonen J, Mäntyselkä P, Ahonen R. Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland. Int J Med Inform 2017 Jan;97:221-228. [doi: 10.1016/j.ijmedinf.2016.10.014] [Medline: 27919380]

22. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]

23. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2016 May 18:262-270. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]

24. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18. [doi: 10.1055/s-0038-1676588] [Medline: 30602196]

25. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]

26. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J 2018 Jun 19;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]

27. Eriksson-Backa K, Hirvonen N, Enwald H, Huvila I. Enablers for and barriers to using My Kanta - a focus group study of older adults' perceptions of the National Electronic Health Record in Finland. Inform Health Soc Care 2021 Dec 02;46(4):399-411. [doi: 10.1080/17538157.2021.1902331] [Medline: 33787438]

28. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. FinJeHeW 2018 Dec 04;10(4):74511. [doi: 10.23996/fjhw.74511]

29. My Kanta pages. Kanta. URL: https://www.kanta.fi/en/web/guest/my-kanta-pages [accessed 2021-11-25]

30. Ammenwerth E, Duftschmid G, Al-Hamdan Z, Bawadi H, Cheung NT, Cho K, et al. International comparison of six basic ehealth indicators across 14 countries: an ehealth benchmarking study. Methods Inf Med 2020 Dec;59(S 02):e46-e63 [FREE Full text] [doi: 10.1055/s-0040-1715796] [Medline: 33207386]

31. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy and Technology 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]

32. Finstad K. The usability metric for user experience. Interact Comput 2010 Sep;22(5):323-327. [doi: 10.1016/j.intcom.2010.04.004]

33. Lewis J, Utesch B, Maher D. UMUX-LITE: when there's no time for the SUS. 2013 Presented at: CHI 2013 Conference; April 27–May 2, 2013; Paris, France p. 2099-2102. [doi: 10.1145/2470654.2481287]

34. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]

35. Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Appl Ergon 2020 Apr;84:103007. [doi: 10.1016/j.apergo.2019.103007] [Medline: 31785449]

36. Bangor A. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]

37. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]

38. Blease C, Delbanco T, Torous J, Ponten M, DesRoches C, Hagglund M, et al. Sharing clinical notes, and placebo and nocebo effects: can documentation affect patient health? J Health Psychol 2022 Jan;27(1):135-146. [doi: 10.1177/1359105320948588] [Medline: 32772861]

39. Sosiaali- ja terveysministeriön asetus potilasasiakirjoista (kumottu). FinLex. URL: https://www.finlex.fi/fi/laki/ajantasa/2009/20090298 [accessed 2022-02-04]

XSL•FO
RenderX

40. Valeur HS, Lie AK, Moen K. Patient rationales against the use of patient-accessible electronic health records: qualitative study. J Med Internet Res 2021 May 28;23(5):e24090 [FREE Full text] [doi: 10.2196/24090] [Medline: 34047711]

41. Kainiemi E, Vehko T, Kyytsönen M, Hörhammer I, Kujala S, Jormanainen V, et al. The factors associated with nonuse of and dissatisfaction with the national patient portal in Finland in the era of COVID-19: population-based cross-sectional survey. JMIR Med Inform 2022 Apr 22;10(4):e37500 [FREE Full text] [doi: 10.2196/37500] [Medline: 35404831]

42. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 2004 Jan;54(498):38-43 [FREE Full text] [Medline: 14965405]

43. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? a retrospective observational study. J Med Internet Res 2020 Feb 06;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]

44. Lee EH, Patel JP, Fortin AH. Patient-centric medical notes: identifying areas for improvement in the age of open medical records. Patient Educ Couns 2017 Aug;100(8):1608-1611. [doi: 10.1016/j.pec.2017.02.018] [Medline: 28242141]

45. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]

46. Chen J, Druhl E, Polepalli RB, Houston TK, Brandt CA, Zulman DM, et al. A natural language processing system that links medical terms in electronic health record notes to lay definitions: system development using physician reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]

47. Lalor JP, Woolf B, Yu H. Improving electronic health record note comprehension with NoteAid: randomized trial of electronic health record note comprehension interventions with crowdsourced workers. J Med Internet Res 2019 Jan 16;21(1):e10793 [FREE Full text] [doi: 10.2196/10793] [Medline: 30664453]

Abbreviations

EHR: electronic health record

XSL•FO
RenderX

<u>Original Paper</u>

Deep Phenotyping of Chinese Electronic Health Records by Recognizing Linguistic Patterns of Phenotypic Narratives With a Sequence Motif Discovery Tool: Algorithm Development and Validation

Shicheng Li[1,2*], MM; Lizong Deng[1,2*], PhD; Xu Zhang[1,2*], MM; Luming Chen[1,2,3], MD; Tao Yang[3,4], MM; Yifan Qi[1,2], MM; Taijiao Jiang[1,2,3], PhD

[1]Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Suzhou Institute of Systems Medicine, Suzhou, China

[3]Guangzhou Laboratory, Guangzhou, China

[4]Guangzhou Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Taijiao Jiang, PhD
Institute of Systems Medicine
Chinese Academy of Medical Sciences & Peking Union Medical College
No.100 Chongwen Road, Suzhou, Jiangsu Province, China
Beijing, 215000
China
Phone: 86 13366191184
Email: taijiaobioinfor@ism.cams.cn

Abstract

Background: Phenotype information in electronic health records (EHRs) is mainly recorded in unstructured free text, which cannot be directly used for clinical research. EHR-based deep-phenotyping methods can structure phenotype information in EHRs with high fidelity, making it the focus of medical informatics. However, developing a deep-phenotyping method for non-English EHRs (ie, Chinese EHRs) is challenging. Although numerous EHR resources exist in China, fine-grained annotation data that are suitable for developing deep-phenotyping methods are limited. It is challenging to develop a deep-phenotyping method for Chinese EHRs in such a low-resource scenario.

Objective: In this study, we aimed to develop a deep-phenotyping method with good generalization ability for Chinese EHRs based on limited fine-grained annotation data.

Methods: The core of the methodology was to identify linguistic patterns of phenotype descriptions in Chinese EHRs with a sequence motif discovery tool and perform deep phenotyping of Chinese EHRs by recognizing linguistic patterns in free text. Specifically, 1000 Chinese EHRs were manually annotated based on a fine-grained information model, PhenoSSU (Semantic Structured Unit of Phenotypes). The annotation data set was randomly divided into a training set (n=700, 70%) and a testing set (n=300, 30%). The process for mining linguistic patterns was divided into three steps. First, free text in the training set was encoded as single-letter sequences (P: phenotype, A: attribute). Second, a biological sequence analysis tool—MEME (Multiple Expectation Maximums for Motif Elicitation)—was used to identify motifs in the single-letter sequences. Finally, the identified motifs were reduced to a series of regular expressions representing linguistic patterns of PhenoSSU instances in Chinese EHRs. Based on the discovered linguistic patterns, we developed a deep-phenotyping method for Chinese EHRs, including a deep learning–based method for named entity recognition and a pattern recognition–based method for attribute prediction.

Results: In total, 51 sequence motifs with statistical significance were mined from 700 Chinese EHRs in the training set and were combined into six regular expressions. It was found that these six regular expressions could be learned from a mean of 134 (SD 9.7) annotated EHRs in the training set. The deep-phenotyping algorithm for Chinese EHRs could recognize PhenoSSU instances with an overall accuracy of 0.844 on the test set. For the subtask of entity recognition, the algorithm achieved an F1 score of 0.898 with the Bidirectional Encoder Representations from Transformers–bidirectional long short-term memory and

conditional random field model; for the subtask of attribute prediction, the algorithm achieved a weighted accuracy of 0.940 with the linguistic pattern–based method.

Conclusions: We developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs with limited fine-grained annotation data. Our work will promote the second use of Chinese EHRs and give inspiration to other non–English-speaking countries.

Introduction

Currently, electronic health records (EHRs) are increasingly becoming an important source for clinical data mining and analysis [1]. Phenotype information that describes patients' clinical manifestations is one of the most valuable clinical information types in EHRs [2]. However, phenotype information in EHRs is mainly recorded in free text, which computers have difficulty using directly [3,4]. Therefore, it is important to develop natural language processing (NLP) technology to effectively structure phenotype information in free text. The NLP technology for structuring phenotype information in EHRs is called EHR-based phenotyping [5].

There are two key factors involved in EHR-based phenotyping [6]. The first factor is the development of an information model that can define the normalized target of phenotyping [7]. The second factor is the development of a phenotyping algorithm that can process phenotype information into a predefined information model [8]. In recent years, the focus of EHR-based phenotyping methods has shifted from the coarse-grained level to the fine-grained level [9,10]. Compared with coarse-grained phenotyping, fine-grained phenotyping can capture more phenotype details, including the phenotype concept and its associated attributes [11]. For example, in the free-text description "a sudden severe pain in the right-lower abdomen," a fine-grained deep-phenotyping method not only considers the phenotype "pain" but also its associated attributes of body location ("abdomen"), temporal pattern ("acute"), and severity ("severe"). EHR-based phenotyping that can characterize phenotype details at a fine-grained level is called EHR-based deep phenotyping [12].

Deep-phenotyping methods can characterize phenotype information in a high-fidelity way, which can potentially improve the accuracy of EHR-based applications, such as disease diagnosis and treatment [13]. Hence, deep phenotyping has become the focus of medical informatics. In recent years, a series of deep-phenotyping methods for English EHRs have been developed. For example, Peterson et al [14] used the MetaMap tool [15] to recognize phenotype concepts in EHRs, along with a neural network model to predict attribute values associated with phenotypes. They finally characterized English EHRs with the Fast Healthcare Interoperability Resources (FHIR) model [16]. Xu et al [17] developed a bidirectional long short-term memory and conditional random field (Bi-LSTM-CRF) model to recognize phenotype concepts in EHRs, together with a machine learning method to predict attribute values, and finally represented the phenotype

information in English EHRs with the clinical element model (CEM) [18]. Compared to the progress of deep-phenotyping English EHRs, the method for deep-phenotyping Chinese EHRs is still in its infancy. Regarding the existence of linguistic differences, the established strategies [14,17,19,20] for deep-phenotyping English EHRs cannot be directly used for Chinese EHRs. Moreover, developing a deep-phenotyping algorithm requires fine-grained annotation data. However, it is hard to obtain a large volume of annotation data because of the high annotation cost. This means that the development of a deep-phenotyping algorithm for Chinese EHRs suffers from the challenge of low-resource scenarios [8], so it is worth considering how to develop a generalized algorithm for deep-phenotyping Chinese EHRs with limited fine-grained annotated data.

In previous work, we developed a fine-grained information model named PhenoSSU (Semantic Structured Unit of Phenotypes) [21], which can accurately characterize phenotype information from medical guidelines with 12 attributes from SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms). To explore an effective strategy for deep-phenotyping Chinese EHRs, we tried to annotate some Chinese EHRs with the PhenoSSU model. During the annotation process, some linguistic patterns of PhenoSSU instances were found to frequently occur in the free text of Chinese EHRs. For example, there is a linguistic pattern of "attribute + attribute + attribute + phenotype" in a given Chinese sentence "患者反复出现(attribute)剧烈(attribute)腹部(attribute)疼痛(phenotype)" (English translation: "patients with repeated severe abdominal pain"). If the linguistic patterns of PhenoSSU instances could be effectively learned from the corpus of Chinese EHRs, it would be possible to perform deep phenotyping of Chinese EHRs by scanning linguistic patterns of PhenoSSU instances. Therefore, how to effectively learn linguistic patterns of PhenoSSU instances from the corpus of Chinese EHRs has become an important question.

Although linguistic patterns of PhenoSSU instances can be observed and summarized manually, this is a time-consuming process that depends on experienced experts. In the field of linguistic pattern mining, the Apriori-based method is one of the most representative algorithms, which was based on the principle of frequency counts of keyword occurrences [22]. The Apriori algorithm is well suited to simple linguistic pattern mining based on word co-occurrence. For example, a recent study used the Apriori algorithm to learn linguistic patterns of cyberbullying behaviors in a social networking service [23]. When two keywords co-occur frequently, they are considered

to constitute a potential linguistic pattern, such as the co-occurrence of "foolish" and "abuse." However, the linguistic patterns of the PhenoSSU instances are more complicated. Thus, Apriori-based methods are not competent at mining linguistic patterns of PhenoSSU instances because they cannot handle the co-occurrence of a phenotype and several attribute values simultaneously. Inspired by the work of Ofer et al [24], which considered biological sequences, such as DNA sequences, as human language and used advanced NLP tools to tackle biological tasks, we aimed to model Chinese EHRs as DNA-like sequences and mine linguistic patterns with advanced bioinformatics tools. In a recent review, Castellana et al [25] surveyed 16 classic DNA motif discovery tools and evaluated their ability to discover sequence motifs nested in 29 simulated sequence data sets. The MEME (Multiple Expectation Maximums for Motif Elicitation) motif discovery tool performed best among the 16 classic DNA motif discovery tools. In this study, we characterized phenotypes as "P" and attributes as "A" to transform the free text into a single-letter sequence that could be analyzed with the MEME motif discovery tool. The sequence motifs discovered in this single-letter sequence could be viewed as linguistic patterns of PhenoSSU instances in Chinese EHRs. Based on the linguistic patterns discovered in EHRs, we could identify PhenoSSU instances by recognizing linguistic patterns in free text. To summarize, the task of deep phenotyping of Chinese EHRs could be converted into two consecutive steps of sequence motif discovery and linguistic pattern recognition.

Following this idea, we aimed to identify linguistic patterns of PhenoSSU instances in Chinese EHRs with a biological sequence motif discovery tool and develop a deep-phenotyping algorithm for Chinese EHRs by scanning linguistic patterns in free text. The rest of this paper is organized as follows. The first section introduces the composition of the PhenoSSU model and its common linguistic patterns in free text. The second section introduces the method for using a biological sequence motif discovery tool to learn linguistic patterns from the corpus of Chinese EHRs. The third section introduces the method for recognizing PhenoSSU instances from Chinese EHRs based on linguistic patterns. The final section introduces a case study to illustrate the potential application of the deep-phenotyping algorithm. Although the deep-phenotyping algorithm developed in this study can only deal with Chinese EHRs, the underlying methodology can also be illuminating for other non–English-speaking countries.

Methods

Overview

In this study, a data-driven approach was proposed for learning linguistic patterns from Chinese EHRs. By using a pipeline of encoding the training set as a single-letter sequence and analyzing the sequence with the MEME motif discovery tool, we learned of six regular expressions and then introduced them into our pattern recognition–based algorithm for attribute prediction. The whole pipeline for the linguistic pattern–learning method is shown in Figure 1.

Figure 1. The pipeline for the linguistic pattern–learning method. A: attribute; C: punctuation; EHR: electronic health record; MEME: Multiple Expectation Maximums for Motif Elicitation; O: other information; P: phenotype; PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.
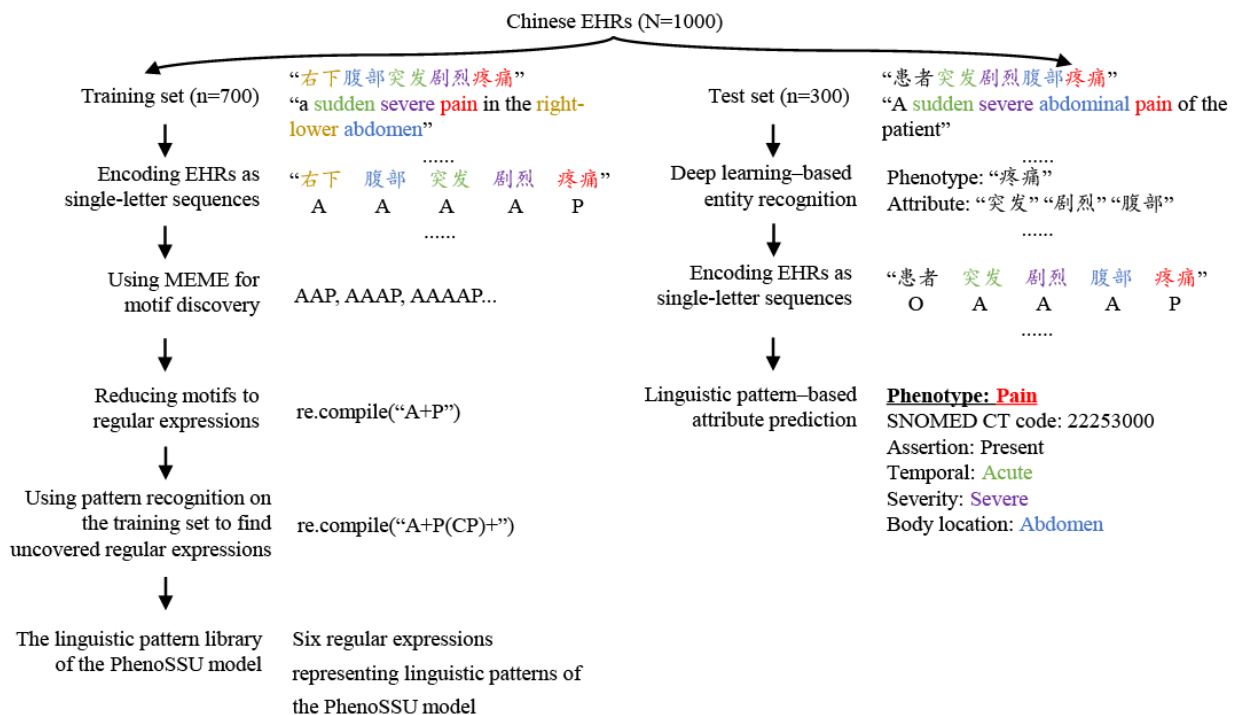
The Design of the PhenoSSU Model for Representing Phenotype Information in Chinese EHRs

PhenoSSU is essentially an entity-attribute-value model consisting of phenotype terms along with standardized attributes from SNOMED CT. Compared with two commonly used information models named CEM and FHIR, the PhenoSSU model is more suitable for the task of deep phenotyping for two reasons. First, it has been shown that the PhenoSSU model is better at representing phenotype information in medical text than CEM and FHIR models [21]. Second, the PhenoSSU model puts more focus on characterizing phenotype traits with standardized attribute and value sets; as well, the attribute and value sets of the PhenoSSU model are easier to adjust according to the study-specific corpus.

To develop a fine-grained annotated corpus, 1000 Chinese EHRs of respiratory system diseases were manually annotated based on the PhenoSSU model, whose design was based on infectious diseases with a large proportion of respiratory diseases [21]. These 1000 Chinese EHRs were obtained from the EHR database of the Iiyi website [26]; all of the patients' private information in these EHRs have been masked by the Iiyi website.

During manual annotation, we optimized the attributes included in the PhenoSSU model to make them suitable for Chinese EHRs. The optimized PhenoSSU model contained 10 attributes, which could be further divided into two subtypes: (1) attributes for phrase-based phenotypes, such as "heavy cough" or "fever," including assertion, severity, temporal pattern, laterality, spatial pattern, quadrant pattern, and body location, and (2) attributes for logic-based phenotypes, such as "WBC [white blood cell] 12.5×10^9/L," including specimen, analyte, and abnormality. The composition of the PhenoSSU model is shown in Figure S1 and Table S1 in Multimedia Appendix 1, as well as the definitions, typical values, and SNOMED CT codes of attributes included in the model.

The phenotype information in free text could be structurally represented by the PhenoSSU model. For example, the description "a sudden severe pain in the right-lower abdomen" could be represented as a PhenoSSU instance consisting of the phenotype concept "pain," the assertion attribute "present," the temporal pattern attribute "acute," the severity attribute "severe," the quadrant pattern attribute "right-lower," and the body location attribute "abdomen." Meanwhile, logic-based phenotypes (ie, qualitative and quantitative test results) were also included in the PhenoSSU model. For example, "WBC 12.5×10^9/L" could be represented as a PhenoSSU instance consisting of the analyte "WBC" and the abnormality attribute "abnormality: higher," which was combined and normalized as a concept of the "increased blood leukocyte number (414478003)" in SNOMED CT (Figure 2, A). The relevant knowledge came from our previous study, LATTE (transforming lab test results) [27], which was integrated into this work, including sample sources, analyte names, and reference ranges for 1098 laboratory tests. Detailed information about the knowledge base is shown in Figures S2 and S3 in Multimedia Appendix 1.

Based on the annotation guideline of the PhenoSSU model in our previous work, two Chinese authors with medical backgrounds (LC and SL) manually annotated these medical records independently. Annotations were made on the brat rapid annotation tool platform [28]. The initial annotating agreement measured with the Cohen κ statistic was 0.851. All inconsistent annotations were decided by the project supervisor (TJ).

During annotation, we found some linguistic patterns of PhenoSSU instances in the EHR text. For example, the description of a phrase-based phenotype, "右下腹部突发剧烈疼痛" (English translation: "a sudden severe pain in the right-lower abdomen"), could be summarized as "attribute (right-lower) + attribute (abdomen) + attribute (acute) + attribute (severe) + phenotype (pain)." Similarly, the description of logic-based phenotypes had common linguistic patterns in free text, such as "analyte (WBC) + number (12.5×10^9) + unit (cells/ L)" (Figure 2, B). If we can mine linguistic patterns of PhenoSSU instances from Chinese EHRs, it would be possible to develop pattern recognition–based deep phenotyping.

Figure 2. Free-text phenotype descriptions and linguistic patterns. A. Examples of structuring free text by the PhenoSSU model. B. Examples of linguistic patterns in free text. A: attribute; L: analyte; N: number; P: pain; PhenoSSU: Semantic Structured Unit of Phenotypes; U: unit; WBC: white blood cell.
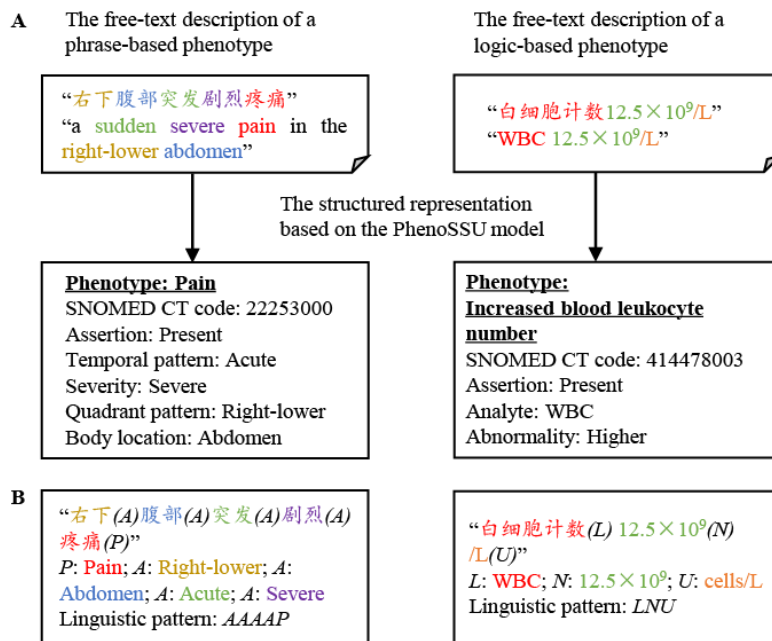

Learning Linguistic Patterns of PhenoSSU Instances From Chinese EHRs Using MEME: Workflow

Overview

In order to learn linguistic patterns of PhenoSSU instances from the Chinese EHR corpus, 1000 annotated Chinese EHRs in the study were divided into a training set (n=700, 70%) and test set (n=300, 30%). The workflow of linguistic pattern mining is shown in Figure 3, which includes two stages: pattern discovery and pattern enrichment. In the stage of linguistic pattern discovery, we used the MEME motif discovery tool, which solves the problem of motif mining with a maximum likelihood method [29] to obtain seed linguistic patterns of PhenoSSU instances. In the stage of linguistic pattern enrichment, a semiautomatic method was developed to check and fill linguistic pattern gaps. Through pattern discovery and enrichment, we built a linguistic pattern library of PhenoSSU instances.

Figure 3. The workflow of learning linguistic patterns of the PhenoSSU model from the corpus of Chinese electronic health records (EHRs). PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern.
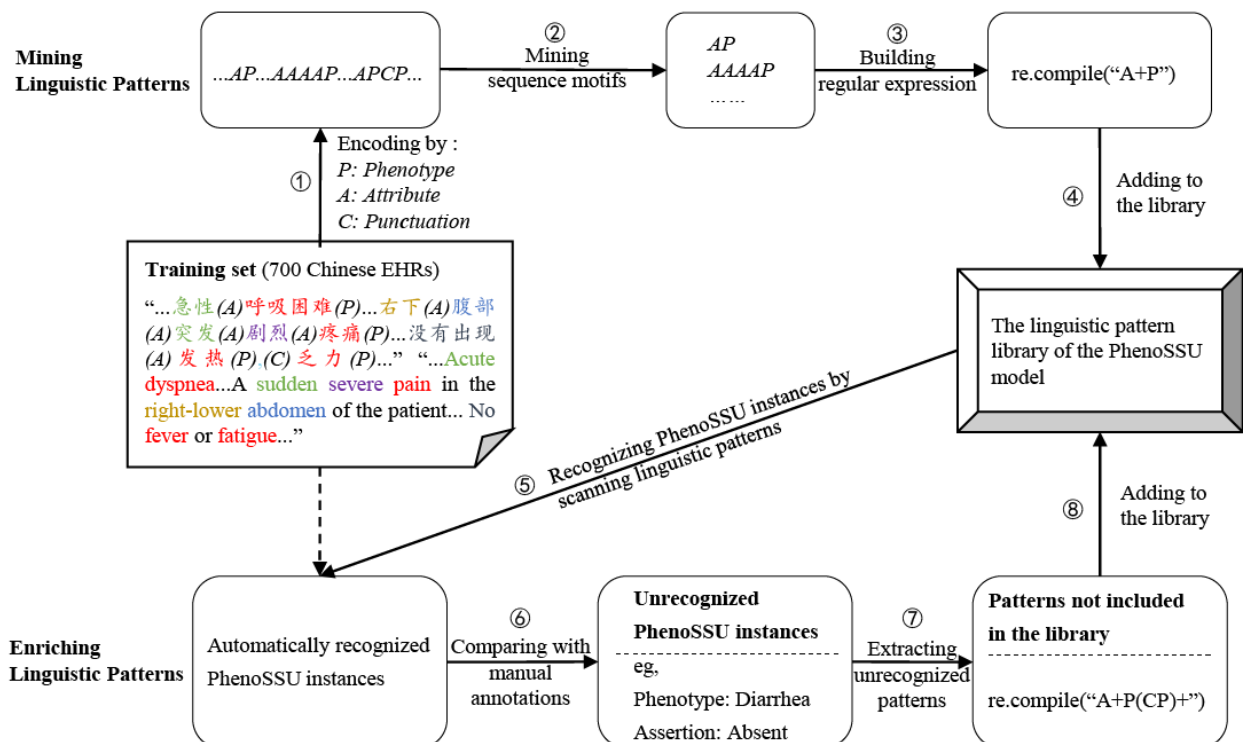
Stage 1: Linguistic Pattern Discovery

First, free text in the training set was encoded into single-letter sequences. To represent EHRs as the input of the MEME motif discovery tool, we encoded them as single-letter sequences with the following criteria: the phenotype (ie, "fever" and "cough") was encoded as "P" and the attribute (ie, "severe") was encoded as "A." In the description of phrase-based phenotypes, "P" and "A" could be directly recognized in the original text. However, to calculate the abnormality of a logic-based phenotype, we need to combine the specimen ("S"), analyte ("L"), number ("N"), and unit ("U"). Specifically, the source of laboratory examination (ie, "blood" and "urine") was encoded as "S," the analyte (ie, "leukocyte") was encoded as "L," the number was encoded as "N" (ie, "37"), and the unit (ie, "°C") was encoded as "U." Meanwhile, the punctuation (ie, a comma) was encoded as "C," and other information was encoded as "O." In this study, EHRs were encoded using the FlashText tool, a tool for string-based concept recognition and replacement [30]. FlashText can find and replace keywords based on the trie dictionary data structure, which is 82 times faster than regular expressions. Because of its efficiency in processing text, we chose the FlashText tool for encoding text as single-letter sequences. Note that FlashText can retain the index of the strings in the original text. For example, the free-text description "患者主诉(O)急性(A)呼吸困难(P)...右下(A)腹部(A)突发(A)剧烈(A)疼痛(P)...没有出现(A)发热(P)，(C)乏力(P)" (English translation: "Patient complained of acute dyspnea...A sudden severe pain in the right-lower abdomen...No fever and fatigue") could be encoded as "AP...AAAAP...APCP." During this stage, we finally obtained single-letter sequences from whole EHRs in the training set.

Second, the MEME motif discovery tool was used to mine motifs in the single-letter sequence. The pipeline of MEME motif discovery is composed of three steps: finding starting points, maximizing the likelihood expectation, and scoring the discovered motifs.

The input was a set of unaligned sequences, and the output was a list of probable motifs. The statistical significance of the motifs in MEME was evaluated by the *E* value, which is based on the log-likelihood ratio. The settings of the MEME motif discovery tool were optimized as follows:

1. Motif discovery mode: classic mode. In classic mode, only one sequence needs to be provided. The algorithm will find the repeated sequence fragments in the sequence by likelihood ranking.
2. Select the site distribution: any number of repetitions. This option means selecting motifs that occur repeatedly.
3. How wide can motifs be: from 2 to 30. This number is the width (ie, characters in the sequence pattern) of a single motif. MEME can choose an optimal width of each motif individually by using a heuristic function. In the process, there were some motifs containing "O" (ie, other information), which was irrelevant to phenotype descriptions. Therefore, we separated out the motifs with

the letter "O" to generate sequence segments that may represent linguistic patterns of PhenoSSU instances.

Third, we built regular expressions based on the discovered motifs. To make the motifs available in our algorithm, regular expressions were built. For example, we built a regular expression "A+P" based on motifs or sequence segments generated from motifs like "AP," "AAP," "AAAP," and "AAAAP."

## Stage 2: Linguistic Pattern Enrichment

In this stage, a linguistic pattern recognition–based method was first used to automatically recognize PhenoSSU instances from Chinese EHRs in the training set. The workflow of linguistic pattern recognition is shown in Figure 4, which includes the following steps:

1. Encoding text as single-letter sequences. For example, the description "右下腹部突发剧烈疼痛" (English translation: "a sudden severe pain in the right-lower abdomen") was encoded as the single-letter sequence "AAAAP." The FlashText tool could record the position index of Chinese characters in every single letter, making it possible to map single letters to the original text. An example of the position index recording is shown in Figure S4 in Multimedia Appendix 1.
2. Scanning the single-letter sequence with the linguistic patterns. In this case, "AAAAP" matched the linguistic pattern "A + P" perfectly, meaning that the four attributes were associated with the phenotype.
3. Mapping these letters to the original text by index. A: right-lower; A: abdomen; A: acute; A: severe; P: pain.
4. Filling phenotypes and associated attributes in the PhenoSSU model. Finally, the description "右下腹部突发剧烈疼痛" could be transformed into a PhenoSSU instance consisting of the phenotype "pain," the assertion attribute "present," the temporal pattern attribute "acute," the severity attribute "severe," the quadrant pattern attribute "right-lower," and the body location attribute "abdomen."

Based on the above steps, we discovered the unrecognized PhenoSSU instances by comparing the automatically recognized instances with manual annotation. For example, the description "没有出现(A)发热(P),(C)乏力(P)" (English translation: "No fever or fatigue") could be encoded as "APCP," in which "AP" matched the regular expression ("A + P") in our pattern library. By mapping to the original text, "没有出现发热，乏力" was transformed into a PhenoSSU instance consisting of the phenotype "fever" and the assertion attribute "absent." However, "absent" was also the attribute of the phenotypes "diarrhea" and "weight loss," which were not recognized by the algorithm.

Finally, to check why these PhenoSSU instances were not recognized, all of them were encoded as single-letter sequences, which could be scanned with linguistic patterns. If no pattern matched, we collected such sequences to build new regular expressions and add them to the linguistic pattern library. In this example, sequences such as "APCPCP" were enriched into a regular expression "(A + P (CP) +)."

Figure 4. The workflow of recognizing PhenoSSU instances from free text via linguistic pattern recognition. The numbers within the square brackets represent the position indexes of single letters in the original text. A: attribute; P: phenotype; PhenoSSU: Semantic Structured Unit of Phenotypes; re.compile: a Python method used to compile a regular expression pattern.
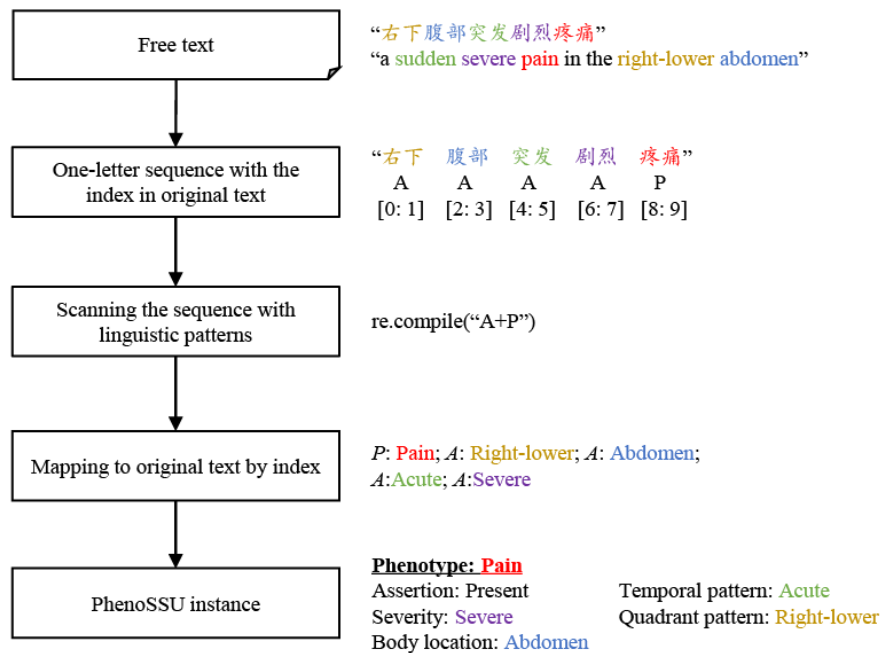
Recognizing PhenoSSU Instances From Chinese EHRs: Workflow

The recognition of PhenoSSU instances could be divided into two subtasks: entity recognition and attribute prediction. To find the best strategy for the two tasks, it was essential to compare our proposed method with current state-of-the-art methods.

The first subtask was entity recognition, which aimed to recognize the text spans corresponding to phenotype and attribute entities. For the subtask of named entity recognition from Chinese EHRs, the Bidirectional Encoder Representations from Transformers (BERT)–Bi-LSTM-CRF model has proven its effectiveness in the CCKS (China Conference on Knowledge Graph and Semantic Computing) 2018 Task 1: Named Entity Recognition in Chinese Electronic Medical Records, which achieved the best F1 score of 91.43 [31]. Therefore, we compared algorithm performances of the BERT-Bi-LSTM-CRF model and the classic dictionary-based method in this study. The parameters of the BERT model were trained with the Kashgari package in Python (version 3.6.1; Python Software Foundation). In the dictionary-based method, the knowledge base of phenotypes was derived from the Chinese translations of the International Classification of Diseases, 10th Revision and 11th Revision, and the Human Phenotype Ontology (details in Table S2 in Multimedia Appendix 1). Further, the knowledge base of attribute trigger words was from the annotation of the training set. Entity recognition, combined with the other coding rules, was applied to encode free text as single-letter sequences, which would be used in the subsequent attribute prediction subtask.

The phenotype's attribute recognition was the second subtask, which aimed to predict appropriate values for the 10 attributes in the PhenoSSU model. The encoded single-letter sequences

from the free text and the developed pattern recognition algorithm in the first subtask were used for attribute prediction. For the subtask of attribute prediction, we did not compare our pattern recognition algorithm with currently existing methods because the PhenoSSU model is a relatively new information model, and algorithms for deep-phenotyping Chinese EHRs based on the PhenoSSU model are very scarce. However, we have referred to state-of-the-art algorithms for deep-phenotyping English EHRs. For example, our previous work showed that the support vector machine (SVM)–based model performed best in the task of deep phenotyping of English clinical guidelines. That is why the SVM model was compared with the linguistic pattern–based method in this study. Three features were used in the SVM model: (1) the distance between phenotype and attribute words, (2) the number of pauses between phenotype and attribute words, and (3) the characteristics of attribute words (eg, some attribute words were only on the left side of phenotype words). The SVM model was built with the scikit-learn package (version 1.1.0) in Python. The parameter tuning of the SVM model was based on a hybrid search strategy. In this study, we did not use deep learning–based methods, because our previous work showed that they were not good at recognizing PhenoSSU instances, owing to the lack of training samples [21].

Evaluation of Algorithm Performance for Recognizing PhenoSSU Instances

To evaluate the algorithm's performance for recognizing PhenoSSU instances, we used the evaluation metrics in SemEval (Semantic Evaluation) 2015 Task 14: Analysis of Clinical Text [32].

In the subtask of entity recognition, the F1 score was taken as the evaluation metric. When a predicted entity word entirely coincided with a gold-standard text span, it was considered as a true positive. The precision metric was calculated as the

fraction of correctly predicted entities among all entities identified by the algorithm, and the recall metric was calculated as the fraction of correctly predicted entities among all entities identified by the annotators. The F1 score was calculated as the harmonic mean of precision and recall.

In the subtask of attribute prediction, the average accuracy and weighted average accuracy were taken as the evaluation metrics because the weighted average accuracy thoroughly considered the distribution of each attribute value in the corpus, which could better evaluate those attribute values with little distribution.

For the evaluation at the PhenoSSU-instance level, we used the combination of the F1 score for entity recognition and weighted average for attribute prediction. A PhenoSSU instance was considered correct when the phenotype and associated attribute values annotated by the algorithm were the same as the corresponding PhenoSSU instance annotated by experts.

Ethical Considerations

The 1000 Chinese EHRs of respiratory system diseases used in this study were obtained from the EHR database of the Iiyi website [26]. No ethics approval was needed because the data from downloaded EHRs, including patients' private information, were all masked by the Iiyi website.

Results

Linguistic Patterns of PhenoSSU Instances Learned From Chinese EHRs

A total of 51 sequence motifs were discovered from the Chinese EHRs in the training set (details are shown in Figure S5 in Multimedia Appendix 1). Based on the 51 motifs, we built six regular expressions (Table 1), namely linguistic patterns of the PhenoSSU instances in the Chinese EHRs. Among the regular expressions of phrase-based phenotypes, "AP +" appeared most

frequently. The most common description of this regular expression was "absent" plus phenotypes, which could be used for differential diagnosis in clinical practice. The second frequent regular expression was "A + P," which usually corresponded to a detailed description of phenotypes, such as "body location + severity + phenotype." There were also complex linguistic patterns to be generalized as "A × PC × A +," for example, "严重(A)咳嗽(P)，(C)呈持续性(A)"（ie, severe cough, consistently). Among the regular expressions of logic-based phenotypes, the most typical was "S × LNU," such as the description "WBC 12×10^9/L." There were also linguistic patterns that directly interpreted laboratory examination results: "S × LR [results of laboratory examination]," such as "血糖升高" (ie, high blood glucose). The above results suggest that there are inherent linguistic patterns in Chinese EHRs. The detailed frequency of linguistic patterns is shown in Table S3 in Multimedia Appendix 1.

In this study, six regular expressions were learned from 700 Chinese EHRs in the training set. However, the size of the training set could be smaller than 700 in order to build the six regular expressions. To explore the potential smallest size of the training set, we conducted an experiment to explore the minimum number of EHRs that could match all six regular expressions. In the experiment, we randomly selected EHRs from the training set with stepwise increased data size, which were scanned with the six regular expressions. When all six regular expressions could be matched, that data size was recorded. This process was repeated 1000 times to calculate the mean and SD of the EHR sums that covered the six regular expressions. Results showed that in a mean of 134 (SD 9.7) EHRs, the six regular expressions could be matched. We did not use the pattern discovery method illustrated in this study because there was a semiautomatic step in the method. Repeating the pattern discovery method 1000 times would be time-consuming. A line graph was plotted to show five examples among all 1000 tests (Figure S6 in Multimedia Appendix 1).

Table 1. Six regular expressions based on linguistic patterns of the Chinese electronic health record corpus in this study.

Phenotype category and regular expressions	Example in Chinese (English translation)
Phrase-based phenotypes	
re.compile[a]("A[b]+P[c](C[d]P)+")	"无/A咳嗽/P、/C发热/P" (no cough or fever)
re.compile("AP+")	"严重/A腹痛/P腹泻/P" (severe abdominal pain and diarrhea)
re.compile("A+P")	"右下腹/A严重/A疼痛/P" (severe right-lower abdominal pain)
re.compile("A×PC×A+")	"咳嗽/P，/C呈持续性/A" (cough, consistently)
Logic-based phenotypes	
re.compile("S[e]×L[f]N[g]U[h]")	"白细胞/L 12×10^9/N /L/U" (WBC[i] 12×10^9/L)
re.compile("S×LR[j]")	"血/S糖/L升高/R" (high blood glucose)

[a]re.compile: a Python method used to compile a regular expression pattern.

[b]A: attribute.

[c]P: phenotype.

[d]C: punctuation.

[e]S: specimen.

[f]L: analyte.

[g]N: number.

[h]U: unit.

[i]WBC: white blood cell.

[j]R: results of laboratory examination.

The Best Strategy for Recognizing PhenoSSU Instances

Based on the linguistic patterns of Chinese EHRs, we developed a pattern recognition–based method to identify PhenoSSU instances. To find the best strategy for recognizing PhenoSSU instances, we developed and compared different methods in the subtasks of entity recognition and attribute prediction. The results in Figure 5 show that the best strategy was to recognize entities using the deep learning–based method and then predict the attribute values using the pattern recognition–based method.

Specifically, in the entity recognition subtask, the method of deep learning (ie, BERT-Bi-LTSM-CRF) achieved the best performance, with an F1 score of 0.898 (Figure 5, B). As a comparison, the dictionary-based method achieved an F1 score of 0.804. In the subtask of attribute prediction, the pattern recognition–based method had the best performance, with an accuracy of 0.977 and a weighted average of 0.940 (Figure 5, C). The SVM-based method achieved an accuracy of 0.783 and a weighted average of 0.709. The deep-phenotyping algorithm for Chinese EHRs had an overall accuracy of 0.844 on the test set. The detailed performances of the two models for predicting attribute values are shown in Table S4 in Multimedia Appendix 1.

Figure 5. Determining the best strategy for recognizing PhenoSSU instances. A. The workflow of recognizing PhenoSSU instances from free text. B. The performance comparison between the dictionary-based method and the deep learning–based method in identifying phenotype concepts. C. The performance comparison between the SVM-based method and the pattern recognition–based method in recognizing a phenotype's attributes. PhenoSSU: Semantic Structured Unit of Phenotypes; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms; SVM: support vector machine.
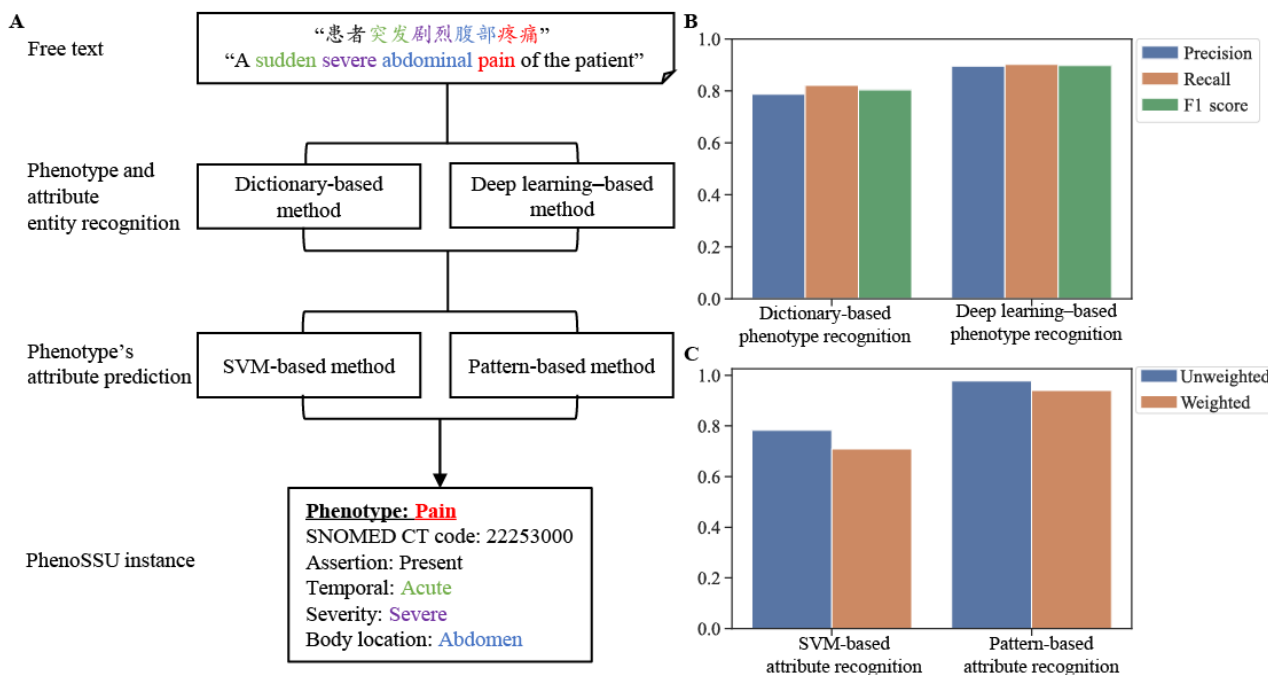
Case Study: Exploring the Real-World Evidence That Deep-Phenotyping EHRs Can Update Knowledge in Guidelines

With the pattern recognition algorithm, we could effectively structure phenotype information in Chinese EHRs. To demonstrate the potential application of deep phenotyping, a case study was conducted to update clinical guidelines by information retrieval of EHRs. In the case study, we selected the latest Chinese clinical guideline and 300 Chinese EHRs of chronic bronchitis. To recognize PhenoSSU instances from the guideline and the EHRs, we used the optimized hybrid strategy mentioned previously.

A total of 9 and 29 PhenoSSU instances were identified from the clinical guideline and the EHRs of chronic bronchitis, respectively (details are shown in Tables S5-S7 in Multimedia Appendix 1). The 9 PhenoSSU instances identified in the clinical guideline appeared in the EHRs, which meant another 20 PhenoSSU instances in the EHRs were not covered in the clinical guideline. For example, "cough: chronic" and "cough: recurrent" both appeared in the clinical guideline and the EHRs. However, the current guideline could not give suggestions to accurately diagnose patients with occasional cough or severe cough as having chronic bronchitis (Figure 6). This real-world evidence hints at the feasibility of updating knowledge in clinical guidelines through deep phenotyping of large-scale EHRs.

Figure 6. The comparison of PhenoSSU instances extracted from the clinical guidelines and electronic health records (EHRs) of chronic bronchitis. PhenoSSU: Semantic Structured Unit of Phenotypes.
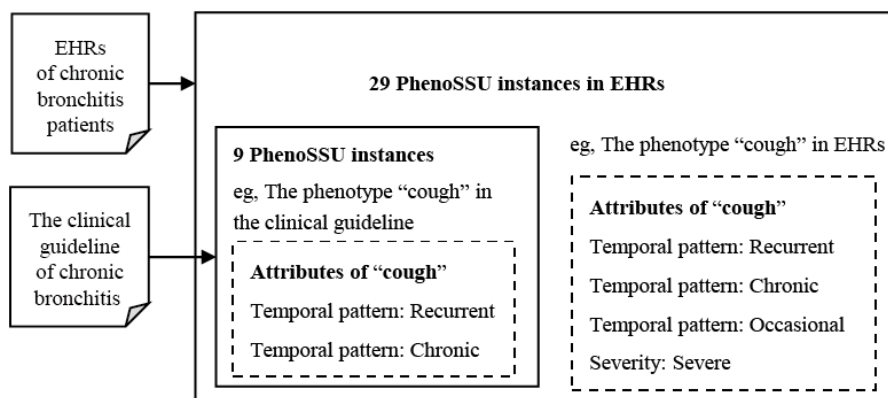
Discussion

Principal Findings

In this study, we developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs. The core of this strategy is learning linguistic patterns of PhenoSSU instances with a motif discovery tool from the field of bioinformatics. According to this research, biological sequence motif discovery tools could be used to effectively identify linguistic patterns of phenotype descriptions from medical texts after encoding them as DNA-like sequences. Meanwhile, the process of identifying linguistic patterns does not require too much annotation data; thus, our strategy is suitable for low-resource scenarios of deep-phenotyping Chinese EHRs.

This study was a preliminary attempt to use bioinformatics tools to tackle problems in medical informatics. By modeling natural language as single-letter sequences, it is possible that other advanced tools for analyzing biological sequences could also be used for processing natural language. For example, some researchers in the NLP field have applied a classic informatics algorithm, named the Basic Local Alignment Search Tool (BLAST), [33] to the text reuse detection task [34]. In Vesanto's work [35], the 23 most-used English letters in the data set were calculated to form a simple one-to-one mapping between English letters and arbitrary amino acids. In this way, text was encoded into single-letter sequences that BLAST could handle to calculate similarities between texts. It is believed that future communications between bioinformatics and medical informatics will become more frequent [36].

It can be concluded from this study that there exist some regular linguistic patterns for phenotype narratives in Chinese EHRs.

The origin of these linguistic patterns may be the common writing habits of clinicians who try to save time by recording clinical information faithfully in as few words as possible [37]. The reason our strategy does not require large annotation samples is that it uses the inner knowledge of linguistic patterns. As we know, data-hungry strategies, such as machine learning and deep learning, require many training samples to effectively identify patterns from data. However, there are many low-resource scenarios in practice that lack sufficient annotation samples for machine learning or deep learning. This is perhaps the reason why the majority (60%) of NLP studies in the medical domain have continued to use a knowledge-based approach rather than a machine learning–based approach [4]. In recent years, researchers have become increasingly focused on integrating machine learning with human knowledge [38], which is expected to become a new paradigm to deal with low-resource scenarios in medical informatics [39].

Limitations

One limitation of this study was that linguistic patterns were learned from EHRs of respiratory diseases, which may not be applicable to other diseases. In addition, limited by the data size, the linguistic patterns in our study might be incomplete. In the future, we will continue to improve the algorithm with more Chinese EHRs from different hospital departments.

Conclusions

We developed a simple but effective strategy to perform deep phenotyping of Chinese EHRs with limited fine-grained annotation data. Our work will promote the second use of Chinese EHRs and bring inspiration to other non–English-speaking countries.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.
[DOCX File , 902 KB-Multimedia Appendix 1]

References

1. Hong C, Rush E, Liu M, Zhou D, Sun J, Sonabend A, VA Million Veteran Program. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. NPJ Digit Med 2021 Oct 27;4(1):151 [FREE Full text] [doi: 10.1038/s41746-021-00519-z] [Medline: 34707226]
2. Katzan IL, Rudick RA. Time to integrate clinical and research informatics. Sci Transl Med 2012 Nov 28;4(162):162fs41. [doi: 10.1126/scitranslmed.3004583] [Medline: 23197569]
3. Elkin PL, Mullin S, Mardekian J, Crowner C, Sakilay S, Sinha S, et al. Using artificial intelligence with natural language processing to combine electronic health record's structured and free text data to identify nonvalvular atrial fibrillation to decrease strokes and death: Evaluation and case-control study. J Med Internet Res 2021 Nov 09;23(11):e28946 [FREE Full text] [doi: 10.2196/28946] [Medline: 34751659]

4. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]

5. Chiu P, Hripcsak G. EHR-based phenotyping: Bulk learning and evaluation. J Biomed Inform 2017 Jun;70:35-51 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.009] [Medline: 28410982]

6. 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 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]

7. Oniki TA, Zhuo N, Beebe CE, Liu H, Coyle JF, Parker CG, et al. Clinical element models in the SHARPn consortium. J Am Med Inform Assoc 2016 Mar;23(2):248-256 [FREE Full text] [doi: 10.1093/jamia/ocv134] [Medline: 26568604]

8. Shi X, Yi Y, Xiong Y, Tang B, Chen Q, Wang X, et al. Extracting entities with attributes in clinical text via joint deep learning. J Am Med Inform Assoc 2019 Dec 01;26(12):1584-1591 [FREE Full text] [doi: 10.1093/jamia/ocz158] [Medline: 31550346]

9. Hripcsak G, Albers DJ. High-fidelity phenotyping: Richness and freedom from bias. J Am Med Inform Assoc 2018 Mar 01;25(3):289-294 [FREE Full text] [doi: 10.1093/jamia/ocx110] [Medline: 29040596]

10. Zhang XA, Yates A, Vasilevsky N, Gourdine JP, Callahan TJ, Carmody LC, et al. Semantic integration of clinical laboratory tests from electronic health records for deep phenotyping and biomarker discovery. NPJ Digit Med 2019;2:32 [FREE Full text] [doi: 10.1038/s41746-019-0110-4] [Medline: 31119199]

11. Delude CM. Deep phenotyping: The details of disease. Nature 2015 Nov 05;527(7576):S14-S15. [doi: 10.1038/527S14a] [Medline: 26536218]

12. Weng C, Shah NH, Hripcsak G. Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability. J Biomed Inform 2020 May;105:103433 [FREE Full text] [doi: 10.1016/j.jbi.2020.103433] [Medline: 32335224]

13. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat 2012 May;33(5):777-780. [doi: 10.1002/humu.22080] [Medline: 22504886]

14. Peterson KJ, Jiang G, Liu H. A corpus-driven standardization framework for encoding clinical problems with HL7 FHIR. J Biomed Inform 2020 Oct;110:103541 [FREE Full text] [doi: 10.1016/j.jbi.2020.103541] [Medline: 32814201]

15. Aronson AR, Lang F. An overview of MetaMap: Historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]

16. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Inform 2021 Jul 30;9(7):e21929 [FREE Full text] [doi: 10.2196/21929] [Medline: 34328424]

17. Xu J, Li Z, Wei Q, Wu Y, Xiang Y, Lee H, et al. Applying a deep learning-based sequence labeling approach to detect attributes of medical concepts in clinical text. BMC Med Inform Decis Mak 2019 Dec 05;19(Suppl 5):236 [FREE Full text] [doi: 10.1186/s12911-019-0937-2] [Medline: 31801529]

18. Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc 2013 May 01;20(3):554-562 [FREE Full text] [doi: 10.1136/amiajnl-2012-001326] [Medline: 23268487]

19. Lu Z, Sim J, Wang JX, Forrest CB, Krull KR, Srivastava D, et al. Natural language processing and machine learning methods to characterize unstructured patient-reported outcomes: Validation study. J Med Internet Res 2021 Nov 03;23(11):e26777 [FREE Full text] [doi: 10.2196/26777] [Medline: 34730546]

20. Yeh MC, Wang Y, Yang H, Bai K, Wang H, Li YJ. Artificial intelligence-based prediction of lung cancer risk using nonimaging electronic medical records: Deep learning approach. J Med Internet Res 2021 Aug 03;23(8):e26256 [FREE Full text] [doi: 10.2196/26256] [Medline: 34342588]

21. Deng L, Chen L, Yang T, Liu M, Li S, Jiang T. Constructing high-fidelity phenotype knowledge graphs for infectious diseases with a fine-grained semantic information model: Development and usability study. J Med Internet Res 2021 Jun 15;23(6):e26892 [FREE Full text] [doi: 10.2196/26892] [Medline: 34128811]

22. Fard MJS, Namin PA. Review of Apriori based frequent itemset mining solutions on big data. In: Proceedings of the 6th International Conference on Web Research. 2020 Presented at: The 6th International Conference on Web Research; April 22-23, 2020; Tehran, Iran p. 157-164. [doi: 10.1109/icwr49608.2020.9122295]

23. Zainol Z, Wani S, Nohuddin PNE, Noormanshah WMU, Marzukhi S. Association analysis of cyberbullying on social media using Apriori algorithm. Int J Eng Technol 2018;7(4.29):72-75 [FREE Full text] [doi: 10.14419/ijet.v7i4.29.21847]

24. Ofer D, Brandes N, Linial M. Comput Struct Biotechnol J 2021;19:1750-1758 [FREE Full text] [doi: 10.1016/j.csbj.2021.03.022] [Medline: 33897979]

25. Castellana S, Biagini T, Parca L, Petrizzelli F, Bianco SD, Vescovi AL, et al. A comparative benchmark of classic DNA motif discovery tools on synthetic data. Brief Bioinform 2021 Nov 05;22(6):bbab303. [doi: 10.1093/bib/bbab303] [Medline: 34351399]

26. Iiyi. URL: https://bingli.iiyi.com/ [accessed 2022-05-25]

27. Jiang K, Yang T, Wu C, Chen L, Mao L, Wu Y, et al. LATTE: A knowledge-based method to normalize various expressions of laboratory test results in free text of Chinese electronic health records. J Biomed Inform 2020 Feb;102:103372 [FREE Full text] [doi: 10.1016/j.jbi.2019.103372] [Medline: 31901507]

28. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. brat: A web-based tool for NLP-assisted text annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. 2012 Presented at: The 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27, 2012; Avignon, France p. 102-107.

29. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res 2015 Jul 01;43(W1):W39-W49 [FREE Full text] [doi: 10.1093/nar/gkv416] [Medline: 25953851]

30. Singh V. Replace or retrieve keywords in documents at scale. ArXiv. Preprint posted online on November 9, 2017 [FREE Full text]

31. Zhang J, Li J, Jiao Z, Yan J. Overview of CCKS 2018 Task 1: Named Entity Recognition in Chinese Electronic Medical Records. In: Proceedings of the 4th China Conference on Knowledge Graph and Semantic Computing. 2019 Presented at: The 4th China Conference on Knowledge Graph and Semantic Computing; August 24-27, 2019; Hangzhou, China p. 158-164. [doi: 10.1007/978-981-15-1956-7_14]

32. Elhadad N, Pradhan S, Gorman S, Manandhar S, Chapman W, Savova G. SemEval-2015 Task 14: Analysis of Clinical Text. In: Proceedings of the 9th International Workshop on Semantic Evaluation. 2015 Presented at: The 9th International Workshop on Semantic Evaluation; June 4-5, 2015; Denver, CO p. 303-310 URL: https://aclanthology.org/S15-2051.pdf [doi: 10.18653/v1/s15-2051]

33. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: A better web interface. Nucleic Acids Res 2008 Jul 01;36(Web Server issue):W5-W9 [FREE Full text] [doi: 10.1093/nar/gkn201] [Medline: 18440982]

34. Vierthaler P, Gelein M. A BLAST-based, language-agnostic text reuse algorithm with a MARKUS implementation and sequence alignment optimized for large Chinese corpora. J Cult Anal 2019;4(2):1-25 [FREE Full text] [doi: 10.22148/16.034]

35. Vesanto A. Detecting and Analyzing Text Reuse With BLAST [master's thesis]. Turku, Finland: University in Turku; 2018. URL: https://www.utupub.fi/bitstream/handle/10024/146706/Vesanto_Aleksi_opinnayte.pdf?isAllowed=y&sequence=1 [accessed 2022-05-21]

36. Ouyang Z, Sargeant J, Thomas A, Wycherley K, Ma R, Esmaeilbeigi R, et al. A scoping review of 'big data', 'informatics', and 'bioinformatics' in the animal health and veterinary medical literature. Anim Health Res Rev 2019 Dec 18;20(1):1-18. [doi: 10.1017/s1466252319000136]

37. Han H, Lopp L. Writing and reading in the electronic health record: An entirely new world. Med Educ Online 2013 Feb 05;18:1-7 [FREE Full text] [doi: 10.3402/meo.v18i0.18634] [Medline: 23394976]

38. Deng C, Ji X, Rainey C, Zhang J, Lu W. Integrating machine learning with human knowledge. iScience 2020 Nov 20;23(11):101656 [FREE Full text] [doi: 10.1016/j.isci.2020.101656] [Medline: 33134890]

39. de Jong J, Cutcutache I, Page M, Elmoufti S, Dilley C, Fröhlich H, et al. Towards realizing the vision of precision medicine: AI based prediction of clinical drug response. Brain 2021 Jul 28;144(6):1738-1750 [FREE Full text] [doi: 10.1093/brain/awab108] [Medline: 33734308]

Abbreviations

A: attribute (in the context of single-letter sequences)
BERT: Bidirectional Encoder Representations from Transformers
Bi-LSTM-CRF: bidirectional long short-term memory and conditional random field
BLAST: Basic Local Alignment Search Tool
C: punctuation (in the context of single-letter sequences)
CCKS: China Conference on Knowledge Graph and Semantic Computing
CEM: clinical element model
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
L: analyte (in the context of single-letter sequences)
LATTE: transforming lab test results
MEME: Multiple Expectation Maximums for Motif Elicitation
N: number (in the context of single-letter sequences)
NLP: natural language processing
O: other information (in the context of single-letter sequences)
P: phenotype (in the context of single-letter sequences)
PhenoSSU: Semantic Structured Unit of Phenotypes
R: results of laboratory examination (in the context of single-letter sequences)
S: specimen (in the context of single-letter sequences)

XSL•FO
RenderX

SemEval: Semantic Evaluation
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
SVM: support vector machine
U: unit (in the context of single-letter sequences)
WBC: white blood cell

Review

Contexts and Outcomes of Proxy Online Health Information Seeking: Mixed Studies Review With Framework Synthesis

Reem El Sherif[1], MBChB, MSc; Pierre Pluye[1], MD, PhD; Fidelia Ibekwe[2], PhD

[1]Department of Family Medicine, McGill University, Montreal, QC, Canada

[2]School of Journalism & Communication, Aix-Marseille University, Aix-en-Provence, France

Corresponding Author:
Reem El Sherif, MBChB, MSc
Department of Family Medicine
McGill University
5858 Chemin de la Côte-des-Neiges
Montreal, QC, H3S 1Z1
Canada
Phone: 1 (514) 398 7375
Email: reem.elsherif@mail.mcgill.ca

Abstract

Background: High-quality online health information (OHI) can reduce unnecessary visits to health professionals and improve health. One of the ways that people use OHI is to support others with health conditions through proxy OHI seeking. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges. There are several models on proxy information seeking. Yet, we know little about the use and outcomes of OHI on behalf of someone else.

Objective: The objectives of this paper are to explore and revise a framework on the context and outcomes of proxy OHI seeking

Methods: We conducted a mixed studies literature review integrating qualitative and quantitative evidence with thematic analysis of the findings of 28 studies, followed by framework synthesis incorporating the derived themes.

Results: We explored 4 main themes: (1) characteristics of proxy seekers, (2) context of proxy OHI seeking, (3) use of OHI to provide social support, and (4) outcomes of proxy OHI seeking. Our conceptual framework incorporates these themes and builds on previous work.

Conclusions: By better understanding how people use information together, information providers can adapt the information to meet all users' needs.

Introduction

Background

Two-thirds (67%) of respondents to the American Health Information National Trends Survey between 2008 and 2017 reported turning to the internet first for health information [1]. Similarly, 69% of Canadians reported using the internet to search for health information in 2020 [2], and the proportion of adults seeking online health information (OHI) in other Organisation for Economic Cooperation and Development (OECD) countries more than doubled between 2008 and 2017 [3]. The use of OHI can improve quality of life and is generally associated with positive outcomes, such as increased empowerment of seekers and their families and improved health outcomes [4-7].

Based on the results of a recent systematic review on the outcomes of OHI seeking (hereafter, OHI outcomes), several contextual factors associated with these outcomes were identified, such as age, education, income, and eHealth literacy [8]. Another contextual factor is social support, defined broadly as "support accessible to an individual through social ties to other individuals, groups, and the larger community"[9]. Social support is an important factor because one of the ways people use OHI is to support family members or friends with health

XSL•FO
RenderX

conditions [10]. In fact, recent studies report that 61%-66% of OHI seekers are proxy seekers, meaning they seek OHI on behalf of someone else [11,12]. Moreover, findings from a study exploring internet use trends between 2008 and 2013 showed a significant increase in the use of family and friends to obtain health information [13].

However, while proxy information seeking has been explored in the literature, especially as it relates to health information, little is known about its relationship with the outcomes of OHI. This is a critical knowledge gap; previous research examining how to reduce negative outcomes of OHI suggests that OHI seekers may be able to overcome low eHealth literacy by discussing the information they find with others [14]. People are sometimes more likely to turn to their social circle to make sense of information they find rather than discuss it with a health professional [11,15]. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges (eg, if they are too physically weak or mentally incapacitated to search themselves) [15].

By better understanding how people and their social circles use information together, information providers can better adapt the information to meet both their needs, and public health interventions can target patients' friends and family with information for dissemination and use [16]. Accordingly, the purpose of this paper is to contribute to our understanding of the role of social support in online health information outcomes by focusing on the outcomes of proxy OHI seeking.

This review will focus on the intersection of 3 main constructs: proxy information seeking, social support, and OHI outcomes.

Proxy Information Seeking

Information seeking encompasses "all the information that comes to a human being during a lifetime, not just in those moments when a person actively seeks information" [17]. In active information seeking mode, monitoring and directed searching are ways to answer known information needs (that are recognized and articulated). There are intervening variables that may be related to personal characteristics, social or interpersonal issues, or environmental considerations [18]. They can be defined as "those who seek information in a nonprofessional or informal capacity on behalf (or because) of others without necessarily being asked to do so" [15]. Proxy seekers may also be "experts," such as health librarians or health care professionals, with the specialized knowledge or skills to use the information with the person with whom they share a personal relationship [19].

The role of proxy information seeking has been explored in the literature and has also been referred to as surrogate seeking or lay information mediation [12,20]. In one of the earliest models on information seeking behavior, Wilson [21] used pathways to explain different patterns of information seeking. In his model, the user encounters "information systems" that can be technology (eg, the internet) and mediators, and these systems connect the user to "information resources" or actual information. Of 10 pathways proposed in this model, 2 indicate seeking that is "conducted by a mediator to fulfill an information request" [21]. This phenomenon is also described in McKenzie's

[22] 2-dimensional model of information practices of women pregnant with twins. In her model, one of the modes of information practice is "by proxy," where the person interacts with information through another agent, including "intermediaries or gatekeepers" such as friends or family members.

Social Support

Social support is one of the positive products of "social relationships" that may have short- and long-term effects on health, for better and for worse, depending on their quality and quantity [23]. A 2004 model by Uchino [24] describes 2 broad dimension of support: structure and function. Structural aspects of support are the extent or composition of one's social network (size, contact, type, density, and strength) and the interconnections among them. Functions have 4 aspects that are highly related to each other: emotional, informational, tangible, and belonging. Most relevant to this review is informational support, which includes the provision of advice or guidance and may provide direction and carry an emotional message when received from a close source. Informational support could be construed as supportive, unsupportive, or mixed depending on the context [25-27].

Social support has consistently been linked to better health [24,28,29]. Several theories have been proposed to explain why this occurs; for example, social support can act as a mediator of stress that reduces its impact, thereby improving mental health [23]. Several studies have reported that those who perceive low social support experience increased stress and report a greater number of stressful events, while those who feel more satisfaction with their received social support report fewer emotional problems [30-33]. Another theory to explain the link between social support and better health is the provision of informational support, which encourages the receivers to manage their health. If we use pregnant women as an example, those who were more satisfied with perceived and received social support initiated prenatal care earlier than those who were less satisfied [34]. Pregnant women who received more informational support from people in their social network delivered babies with higher Apgar scores and higher birth weights [34,35].

Online Health Information Outcomes

A theoretical framework on OHI outcomes and the factors associated with these outcomes was developed by Pluye and colleagues [8] based on a systematic review with a framework synthesis. This framework was derived from previous research by the authors and combines the information theory and psychosocial theory of behavior. It includes 4 types of contextual factors that influence OHI outcomes: (1) individual factors (eg, health literacy); (2) social and technical factors (eg, access to the internet); (3) relationships with professionals (eg, satisfaction with health care provider); and (4) education, health, and social services (eg, access to a family doctor). It also includes 4 levels of individual outcomes of OHI seeking: (1) situational relevance, (2) cognitive/affective impact (eg, being able to understand the information or not liking the information found), (3) use (eg, in discussions with a health care provider or to make a medical decision), and (4) subsequent health/well-being outcomes of use (eg, improved health or reduced worrying) of information.

These levels are presented in Figure 1. For each level, different types of outcomes were identified and validated using systematic mixed studies reviews and qualitative, quantitative, and mixed methods primary research studies [10,36,37].

However, this framework is focused exclusively on an individual perspective: it is the same person that starts the OHI seeking process and experiences the outcomes of this process. Studies that tested this framework therefore focused on people who used the OHI for their own health care and reported the health outcomes they themselves experienced. Little is known about what happens when the information need is to answer a question about someone else's health or what is involved when the information is used with someone else (for providing social support) [14]. Therefore, to adapt this framework to the context of proxy OHI seeking, we are interested in 4 sections of this framework: (1) influencing factors of OHI seeking, (2) OHI seeking behavior including information needs, (3) OHI use, and (4) outcomes of OHI use.

Figure 1. Online health information (OHI) outcomes conceptual framework.
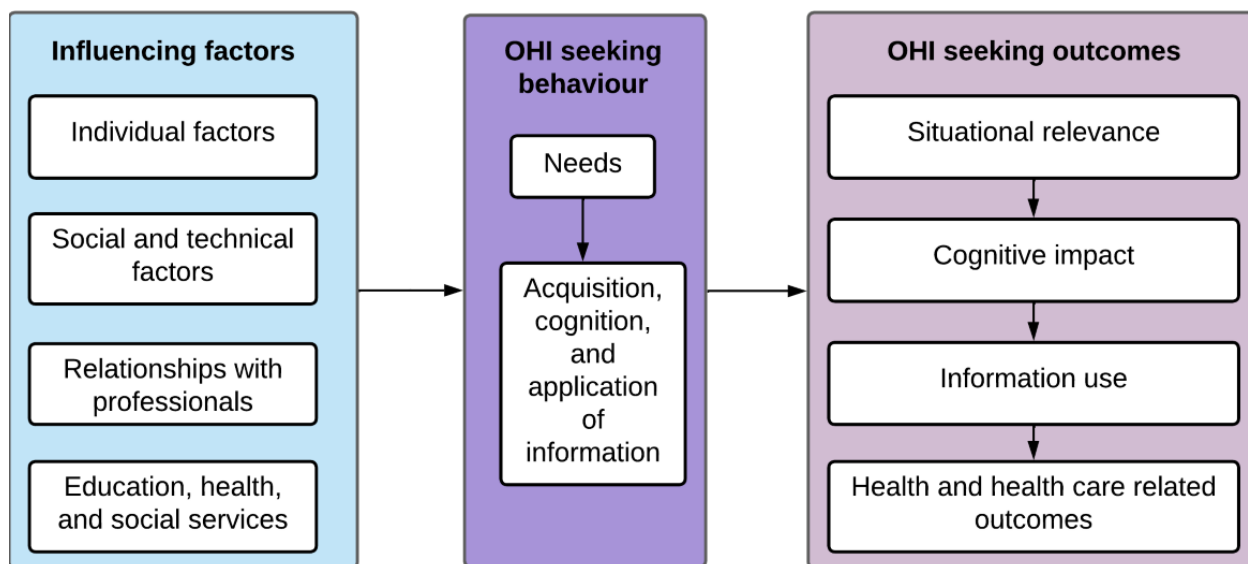
Intersection of the 3 Concepts

There appears to be no comprehensive conceptual model on the outcomes of proxy OHI seekers using OHI to provide social support. Reifegerste et al [38] modified and extended the existing Comprehensive Model of Information Seeking (CMIS) with concepts of social network ties to predict proxy information seeking and the resulting social support intentions. They developed hypothetical scenarios (N=607) of people with varying severity in depression and with varying relationship closeness. Structural equation modeling was used to test the associations between the health-related factors (including demographics), proxy health information seeking intentions, and social support intentions. They hypothesized that support is the resulting action of proxy OHIS. This is an important study that modifies an existing information seeking model to proxy seeking; however, seeking and support were measured only as intentions. Moreover, the demographic characteristics were not found to be relevant, potentially due to the low variance of the study sample. Therefore, our review aims to build on this work by further exploring the context of proxy OHI seeking and the outcomes of using OHI to provide social support.

Methods

Design

A mixed studies review was conducted using a data-based convergent synthesis design in which qualitative and quantitative data were analyzed together using a qualitative thematic analysis [39,40]. A mixed studies review is ideal in this context because the evidence is from diverse fields of inquiry, and it uses diverse methods to provide a rich and highly practical understanding of complex health interventions [41,42]. Framework synthesis was then conducted to produce a revised conceptual framework.

Eligibility Criteria

Table 1 lists the inclusion and exclusion criteria that were deemed appropriate for identifying relevant studies.

Table 1. Inclusion and exclusion criteria.

	Inclusion criteria	Exclusion criteria
Research methods	Primary and secondary research (ie, qualitative, quantitative, and mixed methods empirical studies and literature reviews)	Not empirical research or a literature review (eg, commentary, editorials, reports)
OHI[a]	- Focus on online health information seeking - Online resource about health and medical topics	- No mention of OHI - Offline health information resources (eg, books or pamphlets) - Studies that tested specific online interventions (eg, testing the use of an e-kiosk or e-mental health services)
Proxy OHI seeking	Explore the phenomenon of proxy OHI seeking: - Characteristics of proxy seekers - Context of proxy OHI seeking - Use of OHI - Outcomes of OHI	- No mention of proxy OHI seeking - No mention of seekers that are physical members of the social circle that the person knows and is in contact with on a regular or semi-regular basis (eg, anonymous social media or online forum members) - Exclude parents of young children or surrogate decision-makers of incapacitated adults (eg, unconscious patients in an ICU[b])

[a]OHI: online health information.

[b]ICU: intensive care unit.

Sources and Search Strategy

Papers were searched in 5 databases (Medline, PsycInfo, CINAHL, LISA, and Scopus) from inception to May 25, 2021. A search strategy was compiled with the help of a health librarian and included 2 main concepts: OHI and proxy OHI seeking or social support. The term "surrogate seeking" was discovered after reviewing articles from the first 4 databases and was thus added to the Scopus search strategy. The sets were combined using Boolean operators depending on the database being searched, as presented in Table 2. The search was limited to English and French languages, with no limit on years. All the records were transferred to a reference manager software (EndNote x8) and duplicates were removed using the Bramer method [43]. After the selection stage, additional potentially relevant records were retrieved by tracking the citations (snowballing) of the selected documents.

Table 2. Search strategy.

Database	Date of latest search	Search terms	Records, n
Medline	May 20, 2021	*social support/ AND online.mp. AND "Health Information".af.	82
		"informational support".mp. AND online.mp. AND "Health Information".af.	14
CINAHL	May 20, 2021	"online health information" AND "social support"	16
		"online health information" AND "informational support"	5
PsycInfo	May 20, 2021	*social support/ AND online.mp. AND "Health Information".af.	141
		"informational support".mp. AND online.mp. AND "Health Information".af.	36
LISA	May 20, 2021	"proxy" AND "information seeking" AND "online health"	54
		"social support" AND "online health" AND Information	294
Scopus	May 20, 2021	"surrogate" or "proxy" AND "information seeking" AND "online health"	25
		mediator AND "online health information"	118

Selection of Relevant Studies

The 775 records were then imported into DistillerSR, a web-based application for conducting systematic reviews for selection [44]. For each record, eligibility codes were assigned according to the criteria described in Table 1. For every included record, the corresponding full-text publications were retrieved. Subsequently, full texts were imported into DistillerSR again and coded using the same eligibility criteria. Included studies were then exported into NVivo (Version 12).

Data Extraction and Synthesis of Included Studies

Characteristics of the included studies and results related to the role of social support in OHI seeking and outcomes were coded in NVivo. A deductive-inductive analytical approach was adopted for thematic analysis of the extracted evidence [45]. A coding manual was developed following the framework proposed by Pluye et al [8] that included (1) characteristics of proxy-OHI seekers, (2) context of proxy-OHI seeking, (3) use of OHI by proxy seekers, and (4) outcomes of OHI use for the seeker and recipient. The codes were then progressively clustered into major themes and subthemes.

Framework Synthesis

The initial framework in Figure 1 was revised following the qualitative synthesis stage. An iterative collaborative process was adopted over a series of meetings. All major themes were placed into textboxes and added to the figure representing the initial framework. Alternative figures were proposed until consensus was reached among the authors. The framework was then reviewed by 2 peer reviewers and presented at 2 research meetings (1 local and 1 international), and the feedback received was used to produce the final framework.

Results

Characteristics of Included Studies

Of 775 unique records identified in our search, 28 were deemed relevant and included in our review (Figure 2). Those referred to 15 (53.6%) quantitative studies (including 1 experimental study), 10 (35.7%) qualitative studies, 1 (3.6%) mixed methods study, and 2 (7.1%) systematic reviews. Over half (n=16, 57.1%) of the empirical studies were conducted in North America. The corresponding 28 full-text articles were divided into 3 groups depending on who the focus of the study was: OHI proxy seekers (n=9, 32.1%), OHI recipients (n=2, 7.1%), or both (n=17, 60.7%). Full details of the study characteristics are in presented in Multimedia Appendix 1.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.
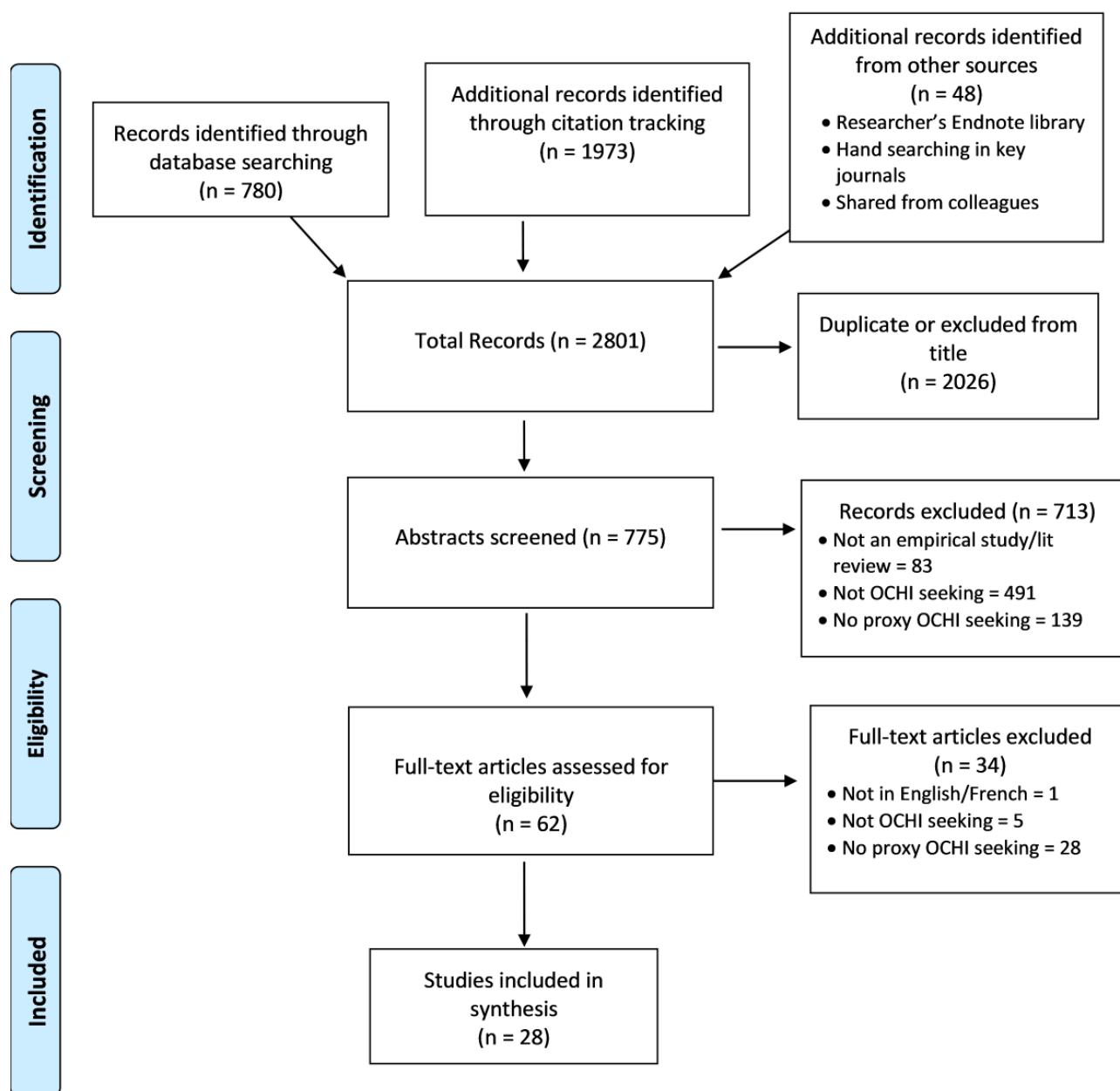
Characteristics of Proxy Seekers

The results of a telephone survey of 18,750 European citizens show that 61% of those seeking OHI searched on behalf of someone else, and of those, 26.6% exclusively searched on behalf of someone else. These surrogate OHI seekers were more likely to live with others and more likely to search on behalf of their partners, children, or other family members rather than for friends or colleagues [11]. This finding was echoed in several studies that reported that the proxy seeker was most often a member of the same household or with whom the person had close ties [12,15,46-51].

This was especially highlighted in relationships where the proxy seeker considered themselves responsible for someone else's health. We found 5 studies that focused on informal (unpaid) caregivers who reported higher and more constant proxy seeking behavior than noncaregivers [46,52,53]. A study exploring information seeking in families affected by multiple sclerosis describes the disease as a shared concern or responsibility that necessitates sharing information about it [54]. Dutta et al [55] described 3G households (3 generations of family members residing together) in Singapore, where the children and grandchildren play vital roles as sources of health information for grandparents.

Several other proxy seeker characteristics influenced OHI seeking behavior. One important factor is gender; 7 studies reported that most people who searched OHI on behalf of others were female [11,15,48,50,54,56,57]. Proxy seekers were generally younger and more educated [11,15,47,48,53,56,58] although 1 study reported that age, education, and income were not significant factors that influenced proxy OHI seeking behavior [59]. Another factor is related to the proxy seeker's experience with OHI: respondents in several of the included studies were reported as having higher health literacy [12,54] and engaging in frequent OHI seeking behavior [11].

Information Needs and Triggers of Proxy Seeking

OHI seeking was triggered by different reasons and at different times in the included studies (Table 3). The proxy seeker may be asked explicitly to search for OHI on behalf of someone who is unable to search for themselves, who has a complex health situation, or who needs to confirm information they had found online themselves [51,55,60,61]. On the other hand, more studies report that the proxy seeker initiates the search unsolicited out of interest [15,61], when they do not have enough information to support a person living with a health condition [47,54], immediately following a diagnosis [62-65], or following a visit with a health care provider [62,66]. Finally, the proxy seeker may also initiate the search themselves as a coping mechanism to help deal them with their emotions following the diagnosis of a loved one [53,61].

Table 3. Information needs and triggers of proxy seeking.

Code	Excerpt
Explicit request	"The carer may be asked to search for information on behalf of the person with cancer. This mostly occurs in situations where the patient does not have access to the internet or is not internet savvy or the person with cancer finds they are too ill to search." [61]
To make a decision	"Both patients and caregivers also mentioned that they surfed the internet again at specific moments later during the lung cancer treatment trajectory, such as during chemotherapy, at the appearance of new symptoms or disease progression, or when having to make a choice between 2 treatment options." [63]
To support someone with a health condition	"A high percentage of the 795 caregivers (87%) had used [the] internet to search for information about the disease of the patient they were taking care for in the last year prior to the survey." [47]
Out of interest or obligation	"For Gina, a 26-year-old Chinese participant, her role as a granddaughter constitutes her interpretation of HIS[a] as she mostly seeks out information for her grandparents. Jamila, a 37-year-old Malay woman, seeks out health information from the internet when one of her family members is not feeling well." [55]
Following a health care practitioner visit	"Patients and caregivers mentioned that their need to seek information often arose once they had time to rest and think about what they had been told, often at a time when their questions could not directly be answered by the treating specialist anymore: 'Once you have come home, you have forgotten half of what you have been told, which is exactly the moment you would want to ask something.'" [63]
Coping mechanism	"Carers also tended to act as 'gatekeepers' of information, and constantly sought new information as a means of coping." [53]

[a]HIS: health information seeker.

How Proxy Seekers Use OHI

Proxy seekers used OHI to better understand someone else's illness or to help themselves feel more empowered in their role as caregivers [49,64,65,67]. Several studies reported the sharing of information between caregiver and patient either directly by sending them a link or printout or indirectly by discussing the information found [49,50,57,60,64,68]. One study describes sharing and resharing the information among a social network so that it reaches a larger number of people [55] or so that a larger number of people are involved in making sense of the information [57].

One aspect of providing informational support involves acting as gatekeeper and controlling incoming information flow for the person [15]. An included literature review exploring the role of caregivers of cancer patients identified this role in 3 included studies, potentially as a way to manage the cancer experience of the patient [61]. Families developed strategies for controlling information sharing, either explicitly with the patient or

XSL•FO

RenderX

implicitly, especially if the information was potentially distressing or could lead to conflict [54,63].

Proxy seekers used the information in discussion with health care providers at a clinical visit [49,55,61,62,64]. This led to asking more questions and feeling more empowered during the visit, as well as involving the provider in the interpretation of the information [49,61,67]. In some cases, it led to requesting more testing or to trying a new treatment plan [62,69]. On the other hand, especially if the provider was not receptive to discussing the information, it also led to confronting or challenging the provider's decision [62].

Proxy seekers also used the information to provide emotional [51,62] and material support, especially as informal caregivers [46,61] to the person. They used the information to change that person's lifestyle; for example, mothers in 1 study cooked healthier food and encourage their families to walk together as a form of exercise [55]. In another study, the authors report that family members used the information to exert control on the patient, using techniques such as pushing or guilting [68].

Outcomes of OHI Use

The outcomes reported by the included studies were overwhelmingly positive. Empowered by the informational they received, proxy seekers and effected individuals felt better informed and more confident, were able to discuss the information with their health care providers, and request different management options [61,62,69]. Information helped people make a health behavior change like quitting smoking [15]. It also helped lessen worries about their own health [15,66]. One study described a 87-year-old participant who reported she feels calmer when her grandchildren print out information and explain treatment options for her [55]. People described how having proxy seekers "care so much" about their health made them feel supported [51] and allowed them to have someone to talk to about their health [64].

Negative outcomes were rarely reported. A literature review found limited reports of patients' anxiety or decisions to refuse cancer treatment [61]. There were 2 studies that reported that the proxy seekers themselves experienced more anxiety, sometimes because of information overload [65,66]. The proxy seeker and the person did not always have the same approach to OHI: in situations where the person did not want to "know" or ignored the information, this led to tensions and conflict [54,68].

Revised Conceptual Framework

Figure 3 shows the revised conceptual framework after the review. The following paragraphs describe proxy seekers, their motivations, how they seek information, and their outcomes.

Figure 3. Outcomes of proxy online health information (OHI) seeking framework.
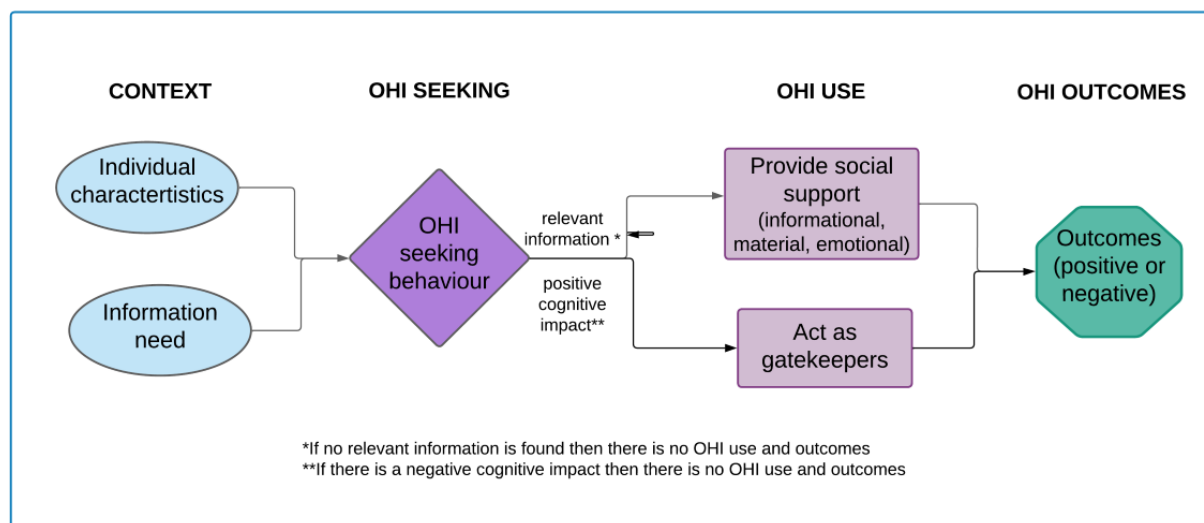
Who Proxy Seekers Are

Proxy seekers are more likely to be female and are also more likely to share health information with others, as they are considered the "central nodes" of health information within a community [70,71]. Moreover, they are more likely to be more educated, with higher eHealth literacy, and frequent internet users in general. Proxy seekers are likely to be in frequent contact with the people for whom they are seeking OHI and to report strong social ties with these people (eg, family members of the same household).

Why And When Does Proxy Seeking Occurs

The OHI seeking process is triggered by an explicit or implicit information need. Explicit information needs may be communicated to the proxy seeker with or without a request for informational support. Proxy seekers who are also informal caregivers may initiate OHI seeking as part of their caregiving responsibilities. The proxy seeker may also initiate the search themselves out of curiosity, for reassurance, or as a coping mechanism to help deal with their emotions following a diagnosis of their loved ones.

How Proxy Seekers Use Information

When proxy seekers find a situationally relevant information object that they understand or agree with (examples of positive cognitive impacts on the seeker), they can use it to provide social support for someone else. This support is most commonly informational: either by sharing the OHI found directly or discussing it with the person to help them make sense of it. Support may also be emotional or material, such as offering to cook meals. The proxy seeker also acts as an information gatekeeper by filtering the information for the person to reduce stress due to information overload.

Outcomes of OHI Use by Proxy Seekers

Using the information will lead to separate outcomes experienced by the person and the proxy seeker, which are generally positive; for example, feeling more confident discussing the information at a clinical visit. In situations where the information is conflicting or unsolicited, it may lead to negative outcomes such as increased worrying or worsening of an interpersonal relationship.

Discussion

Principal Results

To our knowledge, this is the first review to explore the outcomes of proxy OHI seeking and use of OHI to provide social support to others. We adapted a framework on individual OHI outcomes to proxy seekers and described and explained the context, use, and outcomes. Although there are 2 included reviews that reported interesting results, they did not fully address our question: the first explored the role of the internet in supporting and informing caregivers of people with cancer [61], and the second explored how informal caregivers of children with health care needs used internet-based health care services and resources [72]. Another relevant review that explored the proxy OHI seeking behavior of parents for their children and describing a conceptual model was not included in our review because parents are also proxy decision-makers for their children [73]. Another recent study adapted the existing Comprehensive Model of Information Seeking to surrogate health information seeking but did not explore the outcomes of social support [38].

Comparison to Existing Models on OHI Seeking Outcomes

In his revised 1996 model, Wilson [74] added "information processing and use." Our conceptual framework goes further and, in addition to describing the context of information seeking behavior by the proxy seeker, also explores OHI use and outcomes. Similar to the OHI outcomes framework by Pluye [8], our framework includes factors that influence information seeking behavior and leads to 4 levels of outcomes. The use of OHI in our framework revolves around types of social support, and the health and health care–related outcomes are reported by both the proxy seeker and the affected person. Moreover, we identified 2 additional consequences of informational support: sharing misleading information and acting as a gatekeeper to the information.

Our findings echo those of other studies exploring offline proxy health information seeking. In situations where the information need is explicit and the proxy seeker has high health literacy, informational support is associated with positive emotional support, and other outcomes are generally positive. First, people who can discuss the information they found with others are more likely to better understand the information, use that information to make decisions about their health care, and experience better health outcomes such as reduced worries [75-78]. Other potential outcomes include improvement in the receiver's health, buffering of potential negative outcomes, and increase in perceived social support [9,32,79]. This is especially true if the provider has higher health literacy than the receiver, in that they are better able to explain, contextualize, or validate the information [80,81]. Some people may prefer information avoidance, defined as "any behavior designed to prevent or delay the acquisition of available but potentially unwanted information" [82], which may lead to tensions between the proxy seeker and the affected person.

Second, for the seekers themselves, these outcomes include a change in their relationship with the person (improved or worsened) and feeling more involved in the health care of others [83]. Moreover, social support providers who reported feeling more satisfied with their interaction with the person and who felt better about themselves after providing informational support were more likely to continue doing so and more likely to seek information from other sources [83]. Negative outcomes for the seekers reported include increased anxiety due to information overload. This is defined as "when the information processing demands on time…exceed the supply or capacity of time available for such processing" [84].

In situations where the informational support is unsolicited and the person does not feel that the information is relevant to their situation, interpersonal tensions may develop [14]. This may also occur in relation to sharing sensitive or intimate information with family members; for example, a study examining the effects of discussing information on sexuality and contraception on mother-daughter relationships reported that a strain in the relationship may develop [85]. In addition, sharing misleading health information from unreliable sources may also lead to negative health outcomes, as described in 2 recent systematic reviews [86,87]. More specifically, in this context, the seekers do not intend to cause harm and are in fact spreading misinformation that may lead to delayed care, decreased quality of life, and increased risk of mortality.

Limitations

There are some limitations to our review. Unlike in a systematic review, only 1 reviewer carried out the selection phase, so some relevant studies may have been missed. However, our goal was to revise a framework and not necessarily to be exhaustive (in contrast to the needs of a systematic that aims to measure effectiveness of an intervention). Similar to other reviews, there may have been underreporting of negative outcomes due to publication bias. Finally, systematically reviewing all the models on information seeking behavior was beyond the scope of this review, but we reviewed and discussed the most common models with a specialized expert librarian.

Directions For Future Research

Most studies on information seeking behavior do not explore how the information is used by proxy seekers, and what happens next [88]. While this review explores the outcomes of OHI proxy seeking, few studies report outcomes for the seekers themselves. As such, future empirical studies can focus on these outcomes from the seekers' perspectives. Furthermore, little is known about which contextual factors or seeker characteristics are associated with positive and negative OHI outcomes. Future studies can test our framework in different contexts, revise it, and propose research-based solutions to help the proxy seekers use OHI with others.

Conclusion

The outcomes of proxy OHI seeking constitute an important topic for both information specialists and health care practitioners. Members of a person's social circle may help them overcome information-seeking barriers and illness challenges (eg, when they are too physically weak or mentally incapacitated to search themselves) [15]. People are sometimes more likely to turn to members of their social circle to make sense of OHI they find rather than discuss it with a health care professional [11]. By better understanding how affected people and their social circle use OHI together, OHI providers can better adapt their platforms and information to meet both their needs, and health care practitioners can target patients' social circles with information for dissemination and use [16]. Potential public health intervention strategies can focus improving proxy OHI seeking and OHI use to promote positive outcomes for proxy seekers and the people they seek for through strategies that help proxy OHI seekers find relevant OHI, evaluate it, and use it appropriately. Strategies can also include extending social support networks for people without an effective social circle by identifying social support interventions from previous work that may be applicable in the context of proxy OHI seeking.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of Included Studies.
[DOCX File , 42 KB-Multimedia Appendix 1]

References

1. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. Public Health Rep 2019;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]

2. Table 22-10-0137-01 Selected online activities by gender, age group and highest certificate, diploma or degree completed. Statistics Canada. 2020. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210013701 [accessed 2021-05-31]

3. ICT Access and Usage by Households and Individuals database on the Internet. OECD Telecommunications and Internet Statistics. 2017. URL: https://www.oecd-ilibrary.org/content/data/b9823565-en [accessed 2021-05-31]

4. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]

5. Prescott J, Mackie L. "You sort of go down a rabbit hole...you're just going to keep on searching": a qualitative study of searching online for pregnancy-related information during pregnancy. J Med Internet Res 2017 Jun 05;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302]

6. Case D, Given L. In: Mai JE, editor. Looking For Information: A Survey of Research on Information Seeking, Needs, and Behavior 4th ed. London, UK: Emerald Group Publishing; Dec 21, 2016:2284-2286.

7. Pluye P, El Sherif R, Gonzalez-Reyes A, Turcotte E, Schuster T, Bartlett G, et al. Outcomes of equity-oriented, web-based parenting information in mothers of low socioeconomic status compared to other mothers: participatory mixed methods study. J Med Internet Res 2020 Nov 10;22(11):e22440 [FREE Full text] [doi: 10.2196/22440] [Medline: 33170125]

8. Pluye P, El Sherif R, Granikov V, Hong QN, Vedel I, Galvao MCB, et al. Health outcomes of online consumer health information: A systematic mixed studies review with framework synthesis. J Assoc Inf Sci Technol 2019 Jul 30;70(7):643-659. [doi: 10.1002/asi.24178] [Medline: 31423458]

9. Lin N, Simeone RS, Ensel WM, Kuo W. Social support, stressful life events, and illness: a model and an empirical test. J Health Soc Behav 1979 Jun;20(2):108-119. [Medline: 479524]

10. Pluye P, Grad R, Repchinsky C, Jovaisas B, Johnson-Lafleur J, Carrier M, et al. Four levels of outcomes of information-seeking: A mixed methods study in primary health care. J Am Soc Inf Sci Tec 2012 Dec 10;64(1):108-125 [FREE Full text] [doi: 10.1002/asi.22793]

XSL•FO

RenderX

11. Reifegerste D, Bachl M, Baumann E. Surrogate health information seeking in Europe: influence of source type and social network variables. Int J Med Inform 2017 Jul 01:7-14 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.04.006]

12. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]

13. Massey PM. Where do U.S. adults who do not use the internet get health information? examining digital health information disparities from 2008 to 2013. J Health Commun 2016 Nov 23;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]

14. El Sherif R, Pluye P, Thoër C, Rodriguez C. Reducing negative outcomes of online consumer health information: qualitative interpretive study with clinicians, librarians, and consumers. J Med Internet Res 2018 May 04;20(5):e169 [FREE Full text] [doi: 10.2196/jmir.9326] [Medline: 29728350]

15. Abrahamson JA, Fisher KE, Turner AG, Durrance JC, Turner TC. Lay information mediary behavior uncovered: exploring how nonprofessionals seek health information for themselves and others online. J Med Libr Assoc 2008 Oct;96(4):310-323 [FREE Full text] [doi: 10.3163/1536-5050.96.4.006] [Medline: 18974809]

16. Kim W, Kreps GL, Shin C. The role of social support and social networks in health information-seeking behavior among Korean Americans: a qualitative study. Int J Equity Health 2015 Apr 28;14(1):40 [FREE Full text] [doi: 10.1186/s12939-015-0169-8] [Medline: 25927546]

17. Bates M. Toward an integrated model of information seeking and searching. In: New Review of Information Behaviour Research. 2002 Presented at: Fourth international Conference on Information Needs, Seeking and Use in Different Contexts; September 11, 2002; Lisbon, Portugal p. 1-15.

18. Wilson T. Models in information behaviour research. J Doc 1999 Aug;55(3):249-270. [doi: 10.1108/EUM0000000007145]

19. Wyatt S, Henwood F, Hart A, Smith J. The digital divide, health information and everyday life. New Media Soc 2016 Jun 30;7(2):199-218. [doi: 10.1177/1461444805050747]

20. Wathen N, Wyatt S, Harris R. Mediating Health Information: The Go-Betweens in a Changing Socio-Technical Landscape. London, UK: Palgrave Macmillan; 2008.

21. Wilson TD. On user studies and information needs. J Doc 1981 Jan;37(1):3-15. [doi: 10.1108/eb026702]

22. McKenzie P. A model of information practices in accounts of everyday life information seeking. J Doc 2003;59:19-40. [doi: 10.1108/00220410310457993]

23. Umberson D, Montez JK. Social relationships and health: a flashpoint for health policy. J Health Soc Behav 2010 Oct 08;51 Suppl(1_suppl):S54-S66. [doi: 10.1177/0022146510383501] [Medline: 20943583]

24. Uchino B. Social Support and Physical Health: Understanding the Health Consequences of Relationships. London, UK: Yale University Press; 2004:0300127987.

25. Dubois S, Loiselle C. Cancer informational support and health care service use among individuals newly diagnosed: a mixed methods approach. J Eval Clin Pract 2009 Apr;15(2):346-359. [doi: 10.1111/j.1365-2753.2008.01013.x] [Medline: 19335496]

26. Loiselle CG, Lambert SD, Dubois S. Beyond the mere dichotomy of active search versus avoidance of information about the self. J Med Libr Assoc 2006 Oct;94(4):375. [Medline: 17082826]

27. McKinley C, Wright P. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Comput Hum Behav 2014 Aug;37:107-116 [FREE Full text] [doi: 10.1016/j.chb.2014.04.023]

28. Drentea P, Moren-Cross JL. Social capital and social support on the web: the case of an internet mother site. Sociol Health Illn 2005 Nov;27(7):920-943 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00464.x] [Medline: 16313523]

29. House JS. Social isolation kills, but how and why? Psychosom Med 2001;63(2):273-274. [doi: 10.1097/00006842-200103000-00011] [Medline: 11292275]

30. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Managing uncertainty in the context of clubfoot care: exploring the value of uncertainty management theory and the sense of virtual community. Iowa Orthop J 2013;33:142-148. [Medline: 24027474]

31. Hamlett KW, Pellegrini DS, Katz KS. Childhood chronic illness as a family stressor. J Pediatr Psychol 1992 Feb;17(1):33-47. [doi: 10.1093/jpepsy/17.1.33] [Medline: 1545320]

32. Dunst C, Trivette C, Cross A. Mediating influences of social support: personal, family, and child outcomes. Am J Ment Defic 1986 Jan;90(4):403-417. [Medline: 2418680]

33. Kiecolt-Glaser JK, Fisher LD, Ogrocki P, Stout JC, Speicher CE, Glaser R. Marital quality, marital disruption, and immune function. Psychosom Med 1987;49(1):13-34. [doi: 10.1097/00006842-198701000-00002] [Medline: 3029796]

34. Cutrona CE, Suhr JA. Controllability of stressful events and satisfaction with spouse support behaviors. Commun Res 2016 Jun 30;19(2):154-174. [doi: 10.1177/009365092019002002]

35. Guillory J, Niederdeppe J, Kim H, Pollak JP, Graham M, Olson C, et al. Does social support predict pregnant mothers' information seeking behaviors on an educational website? Matern Child Health J 2014 Nov 27;18(9):2218-2225. [doi: 10.1007/s10995-014-1471-6] [Medline: 24671467]

36. Pluye P, Granikov V, Bartlett G, Grad RM, Tang DL, Johnson-Lafleur J, et al. Development and content validation of the information assessment method for patients and consumers. JMIR Res Protoc 2014 Feb 18;3(1):e7 [FREE Full text] [doi: 10.2196/resprot.2908] [Medline: 24550180]

37. Bujold M, El Sherif R, Bush PL, Johnson-Lafleur J, Doray G, Pluye P. Ecological content validation of the Information Assessment Method for parents (IAM-parent): A mixed methods study. Eval Program Plann 2018 Feb;66:79-88 [FREE Full text] [doi: 10.1016/j.evalprogplan.2017.09.011] [Medline: 29053984]

38. Reifegerste D, Blech S, Dechant P. Understanding information seeking about the health of others: applying the comprehensive model of information seeking to proxy online health information seeking. J Health Commun 2020 Feb 01;25(2):126-135. [doi: 10.1080/10810730.2020.1716280] [Medline: 32009552]

39. Pluye P, Hong QN, Bush PL, Vedel I. Opening-up the definition of systematic literature review: the plurality of worldviews, methodologies and methods for reviews and syntheses. J Clin Epidemiol 2016 May;73:2-5. [doi: 10.1016/j.jclinepi.2015.08.033] [Medline: 26898706]

40. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]

41. Pluye P, Hong QN. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45. [doi: 10.1146/annurev-publhealth-032013-182440] [Medline: 24188053]

42. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]

43. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243 [FREE Full text] [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]

44. DistillerSR.Version 2.35.: Evidence Partners; 2021. URL: https://www.evidencepartners.com/ [accessed 2021-02-28]

45. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]

46. Sadasivam RS, Kinney RL, Lemon SC, Shimada SL, Allison JJ, Houston TK. Internet health information seeking is a team sport: analysis of the Pew Internet Survey. Int J Med Inform 2013 Mar;82(3):193-200. [doi: 10.1016/j.ijmedinf.2012.09.008] [Medline: 23149121]

47. Chua GP, Ng QS, Tan HK, Ong WS. Caregivers of cancer patients: what are their information-seeking behaviours and resource preferences? Ecancermedicalscience 2020 Jul 10;14:1068. [doi: 10.3332/ecancer.2020.1068] [Medline: 32728384]

48. Cutrona SL, Mazor KM, Agunwamba AA, Valluri S, Wilson PM, Sadasivam RS, et al. Health information brokers in the general population: an analysis of the Health Information National Trends Survey 2013-2014. J Med Internet Res 2016 Jun 03;18(6):e123 [FREE Full text] [doi: 10.2196/jmir.5447] [Medline: 27260952]

49. Kirschning S, von Kardorff E, Merai K. Internet use by the families of cancer patients—help for disease management? J Public Health 2006 Oct 3;15(1):23-28. [doi: 10.1007/s10389-006-0070-4]

50. Nicholas D, Huntington P, Gunter B, Russell C, Withey R. The British and their use of the web for health information and advice: a survey. Aslib Proceedings: New Information Perspectives 2003;55(5/6):261-276.

51. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]

52. Bangerter LR, Griffin J, Harden K, Rutten LJ. Health information-seeking behaviors of family caregivers: analysis of the Health Information National Trends Survey. JMIR Aging 2019 Jan 14;2(1):e11237 [FREE Full text] [doi: 10.2196/11237] [Medline: 31518309]

53. James N, Daniels H, Rahman R, McConkey C, Derry J, Young A. A study of information seeking by cancer patients and their carers. Clin Oncol (R Coll Radiol) 2007 Jun;19(5):356-362. [doi: 10.1016/j.clon.2007.02.005] [Medline: 17399963]

54. Mazanderani F, Hughes N, Hardy C, Sillence E, Powell J. Health information work and the enactment of care in couples and families affected by multiple sclerosis. Sociol Health Illn 2019 Feb 24;41(2):395-410. [doi: 10.1111/1467-9566.12842] [Medline: 30677163]

55. Dutta MJ, Kaur S, Luk P, Lin J, Lee ST. Health information seeking among Singaporeans: roles and collective contexts. Health Commun 2018 Apr 02;33(4):433-442. [doi: 10.1080/10410236.2016.1278493] [Medline: 28151015]

56. Oh YS. Predictors of self and surrogate online health information seeking in family caregivers to cancer survivors. Soc Work Health Care 2015 Dec;54(10):939-953. [doi: 10.1080/00981389.2015.1070780] [Medline: 26671245]

57. Turner A, Osterhage K, Taylor J, Hartzler A, Demiris G. A closer look at health information seeking by older adults and involved family and friends: design considerations for health information technologies. AMIA Annu Symp Proc 2018;2018:1036-1045 [FREE Full text] [Medline: 30815147]

58. Li H. Informal caregivers' use of the internet for caregiving information. Soc Work Health Care 2015 Jul 17;54(6):532-546. [doi: 10.1080/00981389.2015.1045577] [Medline: 26186424]

59. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney RLJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]

60. Carpenter DM, Elstad EA, Sage AJ, Geryk LL, DeVellis RF, Blalock SJ. The relationship between partner information-seeking, information-sharing, and patient medication adherence. Patient Educ Couns 2015 Jan;98(1):120-124. [doi: 10.1016/j.pec.2014.10.001] [Medline: 25455797]

61. Kinnane NA, Milne DJ. The role of the Internet in supporting and informing carers of people with cancer: a literature review. Support Care Cancer 2010 Sep;18(9):1123-1136. [doi: 10.1007/s00520-010-0863-4] [Medline: 20336326]

62. Dolce MC. The internet as a source of health information: experiences of cancer survivors and caregivers with healthcare providers. Oncol Nurs Forum 2011 May;38(3):353-359. [doi: 10.1188/11.ONF.353-359] [Medline: 21531685]

63. Schook RM, Linssen C, Schramel FM, Festen J, Lammers E, Smit EF, et al. Why do patients and caregivers seek answers from the internet and online lung specialists? A qualitative study. J Med Internet Res 2014 Feb;16(2):e37 [FREE Full text] [doi: 10.2196/jmir.2842] [Medline: 24496139]

64. Simon C, Schramm S. Cancer and the computerized family: towards a clinical ethics of "indirect" internet use. Med Health Care Philos 2008 Sep 19;11(3):337-341. [doi: 10.1007/s11019-008-9127-1] [Medline: 18283559]

65. Coder M. Information needs and information seeking by family members and friends of terminal cancer patients: an exploratory study. J Hosp Librariansh 2020 Feb 04;20(1):1-26. [doi: 10.1080/15323269.2020.1702838]

66. Bouju P, Tadié JM, Uhel F, Letheulle J, Fillatre P, Lavoué S, et al. Internet use by family members of intensive care unit patients: a pilot study. Intensive Care Med 2014 Aug 28;40(8):1175-1176. [doi: 10.1007/s00134-014-3371-z] [Medline: 24972885]

67. Tonsaker T, Law S, Ormel I, Nease C, Bartlett G. Engaging caregivers: exploring perspectives on web-based health information. Fam Pract 2017 Aug 01;34(4):479-484. [doi: 10.1093/fampra/cmw084] [Medline: 27543794]

68. Brown LK, Veinot TC. Information behavior and social control: toward an understanding of conflictual information behavior in families managing chronic illness. J Assoc Inf Sci Technol 2020 Apr 28;72(1):66-82. [doi: 10.1002/asi.24362]

69. Coffey NT, Cassese J, Cai X, Garfinkel S, Patel D, Jones R, et al. Identifying and understanding the health information experiences and preferences of caregivers of individuals with either traumatic brain injury, spinal cord injury, or burn injury: a qualitative investigation. J Med Internet Res 2017 May 10;19(5):e159 [FREE Full text] [doi: 10.2196/jmir.7027] [Medline: 28490418]

70. Altizer KP, Grzywacz JG, Quandt SA, Bell R, Arcury TA. A qualitative analysis of how elders seek and disseminate health information. Gerontol Geriatr Educ 2014 Nov 22;35(4):337-353. [doi: 10.1080/02701960.2013.844693] [Medline: 24188253]

71. Colon-Ramos U, Atienza AA, Weber D, Taylor M, Uy C, Yaroch A. Practicing what they preach: health behaviors of those who provide health advice to extensive social networks. J Health Commun 2009 Mar 16;14(2):119-130. [doi: 10.1080/10810730802659111] [Medline: 19283537]

72. Park E, Kim H, Steinhoff A. Health-related internet use by informal caregivers of children and adolescents: an integrative literature review. J Med Internet Res 2016 Mar 03;18(3):e57 [FREE Full text] [doi: 10.2196/jmir.4124] [Medline: 26940750]

73. Kubb C, Foran HM. Online health information seeking by parents for their children: systematic review and agenda for further research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]

74. Wilson T. Information behaviour: An interdisciplinary perspective. Inf Process Manag 1997 Jul;33(4):551-572. [doi: 10.1016/s0306-4573(97)00028-9]

75. Tanis M, Hartmann T, Te Poel F. Online health anxiety and consultation satisfaction: A quantitative exploratory study on their relations. Patient Educ Couns 2016 Jul;99(7):1227-1232. [doi: 10.1016/j.pec.2016.01.021] [Medline: 26873545]

76. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]

77. Iverson S, Howard K, Penney B. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008;108(12):699-711.

78. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: A systematic review. Patient Educ Couns 2021 Apr;104(4):770-784. [doi: 10.1016/j.pec.2020.11.016] [Medline: 33358253]

79. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]

80. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]

81. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-05-31]

82. Sweeny K, Melnyk D, Miller W, Shepperd JA. Information avoidance: who, what, when, and why. Rev Gen Psychol 2010 Dec 01;14(4):340-353. [doi: 10.1037/a0021288]

83. Hether HJ, Murphy ST, Valente TW. It's better to give than to receive: the role of social support, trust, and participation on health-related social networking sites. J Health Commun 2014 Dec 25;19(12):1424-1439. [doi: 10.1080/10810730.2014.894596] [Medline: 24766297]

84. Schick A, Gordon L, Haka S. Information overload: A temporal approach. Accounting, Organizations and Society 1990 Jan;15(3):199-220. [doi: 10.1016/0361-3682(90)90005-F]

XSL•FO

RenderX

85. Amsellem-Mainguy Y. Prescrire et proscrire des conduites, véhiculer des normes: les mères comme actrices privilégiées de prévention en matière de sexualité et de contraception. Recherches familiales 2006;3(1):49-59 [FREE Full text] [doi: 10.3917/rf.003.0049]

86. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]

87. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41(1):433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]

88. Case DO, O'Connor LG. What's the use? Measuring the frequency of studies of information outcomes. J Assn Inf Sci Tec 2015 Jan 30;67(3):649-661. [doi: 10.1002/asi.23411]

Abbreviations

CMIS: Comprehensive Model of Information Seeking
OECD: Organisation for Economic Cooperation and Development
OHI: online health information

XSL•FO
RenderX

<u>Original Paper</u>

Medication Use and Clinical Outcomes by the Dutch Institute for Clinical Auditing Medicines Program: Quantitative Analysis

Rawa Kamaran Ismail[1*], PhD, PharmD; Jesper van Breeschoten[2*], PharmD; Silvia van der Flier[3], PhD; Caspar van Loosen[4], MSc; Anna Maria Gerdina Pasmooij[5], PhD; Maaike van Dartel[5], PhD; Alfons van den Eertwegh[6], PhD, MD; Anthonius de Boer[1], MD, PhD; Michel Wouters[7], MD, PhD; Doranne Hilarius[8], PhD, PharmD

[1]Division of Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Utrecht, Netherlands

[2]Dutch Institute for Clinical Auditing, Leiden, Netherlands

[3]Stichting Volksgezondheidszorg (VGZ), Arnhem, Netherlands

[4]Dutch Association of Hospital Pharmacists, De Meern, Netherlands

[5]Medicines Evaluation Board, Utrecht, Netherlands

[6]Department of Medical Oncology, Amsterdam University Medical Center, location VUmc, Amsterdam, Netherlands

[7]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, Netherlands

[8]Rode Kruis Ziekenhuis, Beverwijk, Netherlands

[*]these authors contributed equally

Corresponding Author:
Jesper van Breeschoten, PharmD
Dutch Institute for Clinical Auditing
Rijnsburgerweg 10
Leiden, 2333AA
Netherlands
Phone: 31 613382358
Email: j.vanbreeschoten@dica.nl

Abstract

Background: The Dutch Institute for Clinical Auditing (DICA) Medicines Program was set up in September 2018 to evaluate expensive medicine use in daily practice in terms of real-world effectiveness using only existing data sources.

Objective: The aim of this study is to describe the potential of the addition of declaration data to quality registries to provide participating centers with benchmark information about the use of medicines and outcomes among patients.

Methods: A total of 3 national population-based registries were linked to clinical and financial data from the hospital pharmacy, the Dutch diagnosis treatment combinations information system including in-hospital activities, and survival data from health care insurers. The first results of the real-world data (RWD) linkage are presented using descriptive statistics to assess patient, tumor, and treatment characteristics. Time-to-next-treatment (TTNT) and overall survival (OS) were estimated using the Kaplan-Meier method.

Results: A total of 21 Dutch hospitals participated in the DICA Medicines Program, which included 7412 patients with colorectal cancer, 1981 patients with metastasized colon cancer, 3860 patients with lung cancer, 1253 patients with metastasized breast cancer, and 7564 patients with rheumatic disease. The data were used for hospital benchmarking to gain insights into medication use in specific patient populations, treatment information, clinical outcomes, and costs. Detailed treatment information (duration and treatment steps) led to insights into differences between hospitals in daily clinical practices. Furthermore, exploratory analyses on clinical outcomes (TTNT and OS) were possible.

Conclusions: The DICA Medicines Program shows that it is possible to gather and link RWD about medicines to 4 disease-specific population-based registries. Since these RWD became available with minimal registration burden and effort for hospitals, this method can be explored in other population-based registries to evaluate real-world efficacy.

XSL•FO
RenderX

Introduction

Regulatory authorities approve the majority (76%) of new cancer drugs based on evidence provided by randomized controlled trials (RCTs) [1]. These RCTs have high internal validity and are widely considered the gold standard for establishing the efficacy of new drugs [2]. Many new cancer drugs have been recently approved based on very specific patient groups, surrogate outcomes, and lower patient numbers; these drugs are increasingly approved in accelerated tracks [3,4]. The selected patient groups and well-controlled setting of these RCTs has led to criticisms of their external validity [5]. In addition, recent research has shown that almost one-half of RCTs that applied for marketing authorization for new cancer drugs in Europe had a high risk of bias. This increased risk of bias was caused by their design, conducted analyses, and conduct deficits [1]. Further, due to the increase in newly approved cancer and rheumatic disease drugs, health care costs have increased. The total expenditures by hospitals on expensive medicines in the Netherlands reached €2.1 billion (US $2.2 billion) in 2019 [6].

Following market entry, new cancer drugs are prescribed to a broader group of patients with different characteristics. This leads to a gap in clinical outcomes evidence between RCTs and the real world [7,8]. During routine clinical practice, real-world data (RWD) are generated and registered in validated population-based cancer registries. Clinical quality registries are an important tool for quality assessment and improvement in hospitals, consequently leading to demonstrable improvements in patient outcomes [9]. Comparing the quality of care across hospitals results in insights into differences in outcomes, which can lead to improvements in care [9,10]. Furthermore, data from quality registries are used for outcomes research and to study practice variation between centers using quality indicators [11]. Besides clinical quality registries, detailed administrative and declaration data are available specifically on the use of (expensive) drug treatments. The combination of these data in clinical quality registries, hospital administrative data, and declaration data of drugs used in these indications could be valuable to bridge the efficacy-effectiveness gap.

Previous initiatives linked various databases on drugs to clinical data. This linkage made it feasible to study drug use, health resource use, costs, effectiveness, and the safety of medicines [12]. However, a gap remains for recently approved expensive cancer drugs.

To better understand the effectiveness of expensive cancer medicines in a real-world population, the Dutch Institute for Clinical Auditing (DICA) initiated the Medicines Program in 2018. The program aims to identify variation in use and clinical outcomes of expensive medicines, provide postmarketing authorization data, provide a tool for clinicians to benchmark their practice on the use of expensive medicines, and stimulate interactions between clinicians to share best practices. In this program existing data sources were used. This study aims to describe the potential for the addition of declaration data to quality registries to provide participating centers with benchmark information about the use of medicines and associated outcomes.

Methods

Ethics Approval

In compliance with Dutch regulations, the DICA quality registries were approved by the medical ethical committee of the Leiden University Medical Center and was not subject to the Medical Research Involving Human Subjects Act.

Data Sources

Different existing data sources were used in the DICA Medicines Program; these data sources were linked. The first data sources were national population-based registries that are managed by the DICA. The DICA is a nonprofit organization that facilitates 23 population-based registries on different disciplines and diseases. These registries include information on clinical characteristics but contain limited data on the use of medicines. The DICA Medicines Program uses the Dutch Colorectal Audit (DCRA) [10], the Dutch Lung Cancer Audit [13], and the National Breast Cancer Organization Breast Cancer Audit (NBCA) [14]. These quality registries include information on patient, tumor, and treatment characteristics, and are used to compare hospitals on structure, processes, and clinical outcomes [15,16]. A previous study has shown that the data entered in the DICA registries are accurate and complete [17].

The second data source was financial and administrative data, including hospital pharmacies' declarations of expensive medicines for health insurers. These expensive medicines are listed as expensive (>€1000 per patient per year, equivalent to >US $1058.39) by the Dutch Healthcare Authority [18]. This data source includes precise and valid information about the diagnosis, date of prescription, dose, and quantity of a prescribed drug. Administrative data from hospitals include declarations for the reimbursement of expensive medicines. Only expensive medicines that were relevant and related to the diagnosis were linked to the clinical data.

The third data source includes the Dutch diagnosis treatment combinations (DBC) information system, which contains information on in-hospital activities, such as computerized tomography (CT) scans, infusions, hospital admissions, day treatments, and radiology treatments. The DBC information system is used for the registration and reimbursement of hospital and medical specialist care. This system was introduced in the Netherlands to increase the transparency of care. Furthermore, DBC information systems were initiated to create a supply-led system, increase efficiency, and facilitate competition between health care providers [19]. Because the DCRA and NBCA quality registries only include patients undergoing surgical operations, patients with metastasized cancers who do not undergo surgical operations are missing. To include patients with metastasized colorectal and metastasized breast cancer, the DBC data were used and linked to the fourth data source.

The fourth data source was survival data from the national claims database (VEKTIS) from health insurers [20]. VEKTIS is the national insurance database, which contains administrative data from Dutch national health care insurers, covering approximately 17 million individuals. By adding this data source, we could assess overall survival (OS) from diagnosis

XSL·FO

RenderX

and the start of systemic therapy. Data were retrospectively collected from patients treated from 2017 to 2020. Although the DICA Medicines Program was established in 2018, data from 2017 were available from the hospitals and were therefore linked.

Data Linkage and Privacy

The first step in data linkage was identifying patients diagnosed with colorectal cancer, lung cancer, breast cancer, and rheumatic disease using the DBC information system. The DBC information system is used for the registration and reimbursement of health care in the Netherlands. The second step was to identify whether these patients used relevant expensive drugs, and the third step was to determine whether these patients are registered in the national quality registry. Information on the date of death from the VEKTIS database was added for deceased patients (Figure 1). Data were linked based on hospital patients' ID. A third party pseudonymized patient IDs. The results were visualized in dynamic web-based dashboards in which (systemic) treatments were linked to clinical parameters. Filters on patient and tumor characteristics, clinical outcomes, and therapy varied for the different diagnoses, depending on relevance. Furthermore, participating hospitals were compared, and practice variation was visualized and discussed to share knowledge on medical treatment differences.

Figure 1. Visualization of the patients included in our study and the different data sources used.
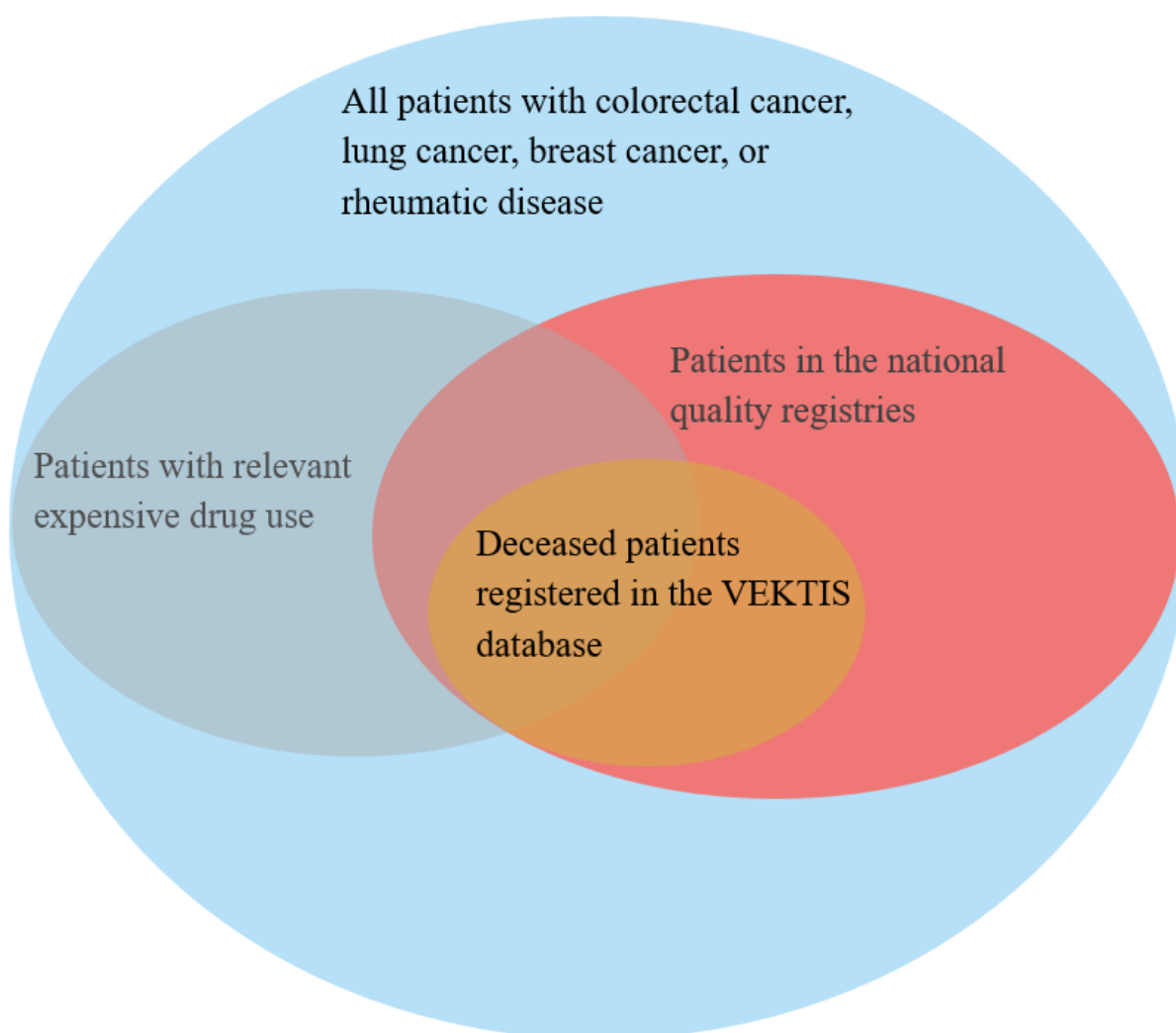
Statistical Analysis

The analyses in this manuscript are exploratory. Descriptive statistics were used to assess patient, tumor, and treatment characteristics. Time-to-next-treatment (TTNT) and OS were estimated with the Kaplan-Meier method. Survival times were calculated from the start of a systemic therapy until subsequent treatment (TTNT) or until death from any cause (OS). Patients who were alive or lost to follow-up were right censored at the time of their last registered expensive medicine use. All the statistical data were analyzed using R (version 4.0.2; R

Foundation for Statistical Computing) within the RStudio environment (version 3.5.2; RStudio PB; packages tidyverse [21], TableOne [22], Survminer [23]).

Results

Database

A total of 21 Dutch hospitals participated in the DICA Medicines Program and were included in this study. Of these hospitals, 9 were top clinical hospitals, 11 were peripheral

hospitals, and 1 was an academic hospital. The geographic location of these hospitals is shown in Figure 2, which indicates they are spread across the country. The DICA Medicines database included a total of 7412 patients with colorectal cancer,

1981 patients with metastasized colon cancer, 3860 patients with lung cancer, 1253 patients with metastasized breast cancer, and 7564 patients with rheumatic disease.

Figure 2. A map of the Netherlands including the geographic location of the participating hospitals in the Dutch Institute for Clinical Auditing Medicines Program (red dots).

Benchmarking

The DICA Medicines Program provides the ability to compare results between hospitals to improve the quality of care provided. Hospitals were provided with web-based dynamic dashboards (Multimedia Appendix 1), continuously comparing their data to the benchmark. The benchmark consisted of all other participating hospitals. An example of benchmarking is the use of systemic therapies at the end of life in patients with metastatic colorectal cancer. This varied between hospitals from 4.2% (5/119) to 27.8% (5/18), with a median of 13.4%. The dashboards also provide information on the type of systemic therapy used at the end of life. A signaling function is included in the dashboard if hospitals deviate from the benchmark (Multimedia Appendix 2). Deviation from the benchmark was defined as a ranked average calculated as follows: (Percentage

of cases within hospital X – Percentage of cases within the benchmark)^2 + Total number of patients in the benchmark.

Use of Medicines and Patient Characteristics

The linkage of different data sources led to new insights into hospitals' use of medicines and patient populations. The patient and tumor characteristics are listed for each medicine in the dashboard as a table that hospitals can compile with available variables. One of the participating hospitals discovered a deviation from the benchmark in the percentage of mesothelioma using the dashboard (Multimedia Appendix 3). This was 9.2% (14/153) for the specific hospital, compared to only 18% (105/3480) in the benchmark.

Treatment Information

Linking clinical data to systemic treatment information also led to detailed information for each medicine, such as treatment duration in months and the number of cycles per patient. An example is the number of courses of capecitabine and oxaliplatin for the adjuvant treatment of colorectal cancer (Figure 3). Furthermore, administrative data were used to visualize treatment steps in Sankey diagrams in the dashboard, which can be adjusted for specific filters on patient, tumor, and treatment characteristics. Figure 4 shows the Sankey diagram for patients with metastasized colon cancer who were treated between 2017 and 2020. The dashboards also contain detailed information on diagnostic imaging (CT and magnetic resonance imaging scans), the number of consults (or teleconsults), clinical admissions, and emergency room visits pre- and post treatment for each medicine (Multimedia Appendix 4).

Figure 3. Number of courses of capecitabine + oxaliplatine for the adjuvant treatment of colorectal cancer patients per hospital.
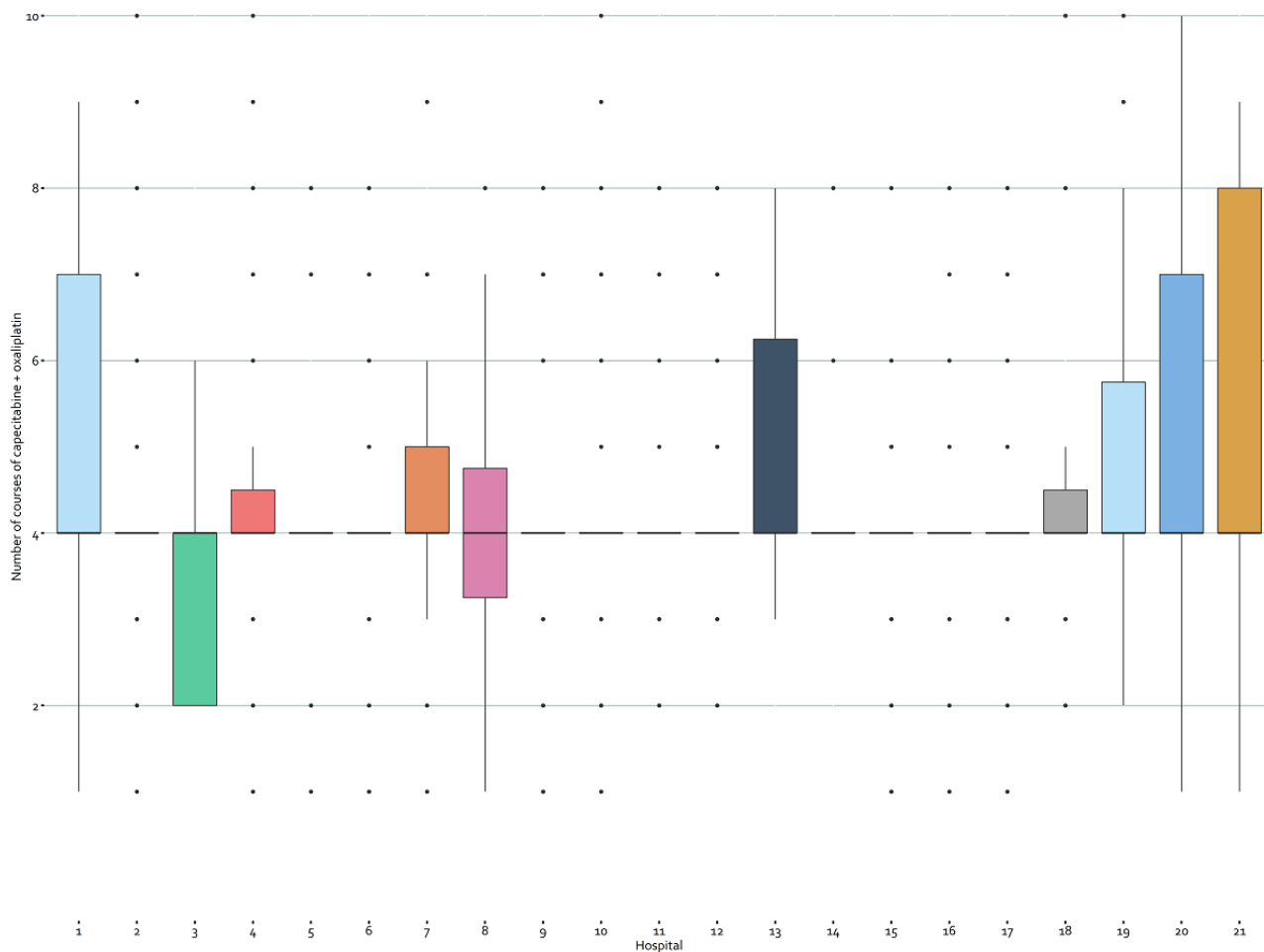
Figure 4. Treatment patterns in patients with stage III colon cancer treated between 2017 and 2020 (N=1668). The Sankey diagram shows the flow of patients from the first treatment step to the second treatment step and from the second treatment step to the third treatment step. The width of the lines corresponds with the number of patients. Systemic therapies with less than 5 patients are not displayed in this graph. Cap-B: Capecitabine plus bevacizumab; Capox: Capecitabine plus oxaliplatin; Capox-B: Capecitabine plus oxaliplatin plus bevacizumab; Folfiri-B: Fluorouracil plus irinotecan plus bevacizumab; Folfox: Fluorouracil plus oxaliplatin.
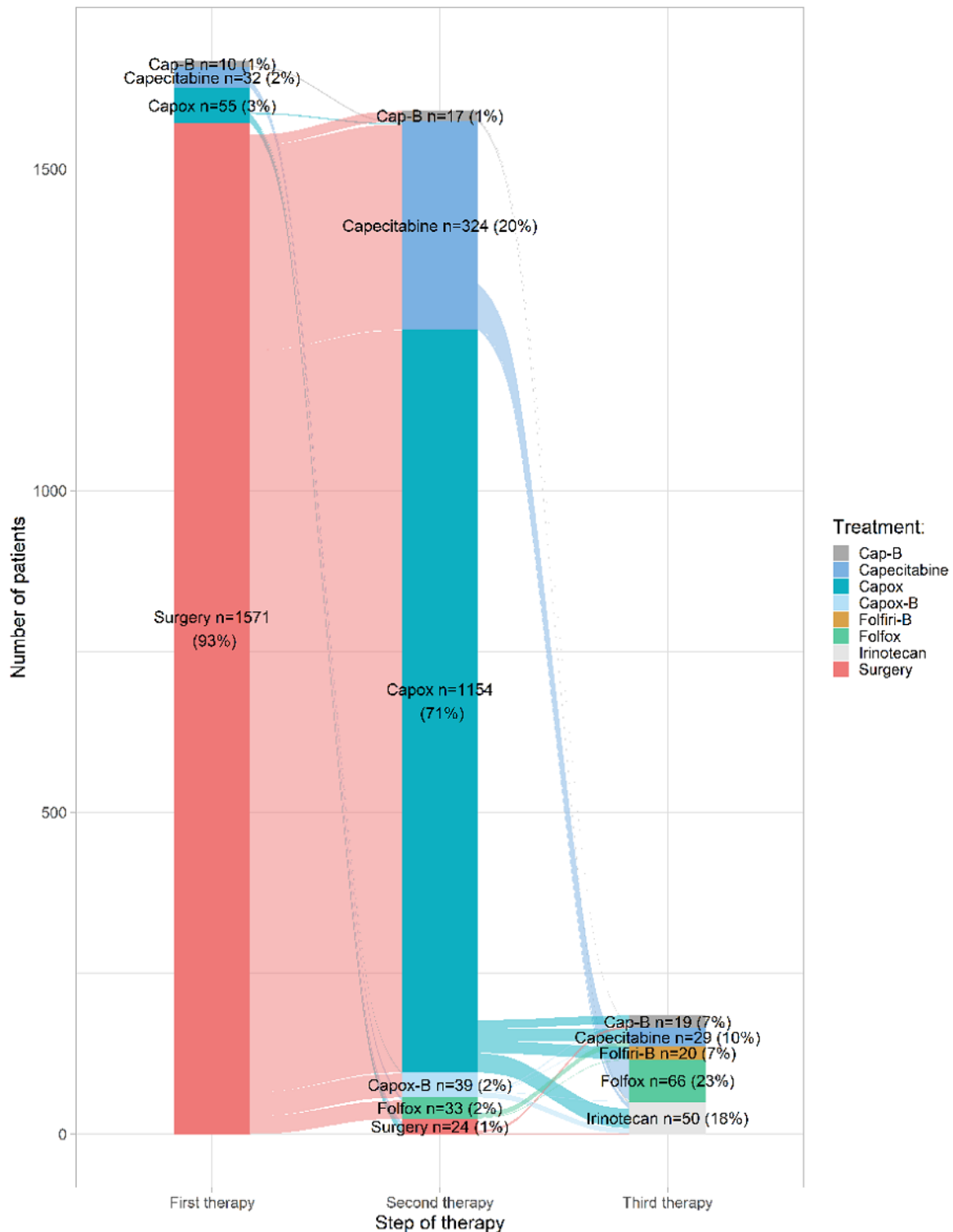
Clinical Outcomes

The DICA Medicines Program also provides hospitals with data on clinical outcomes, such as TTNT and OS. Figure 5 shows the TTNT of patients with metastasized lung cancer treated with first-line pembrolizumab or pembrolizumab and pemetrexed combination therapy. The median TTNT was 22.5 (95% CI 17.0, upper range not available) months and 14.9 (95% CI 12.4-21.6) months for pembrolizumab monotherapy and the combination of pembrolizumab and pemetrexed, respectively. The OS of these treatments is presented in Figure 5. Each hospital's outcomes are compared to the benchmark. It is also

possible to compare clinical outcomes between treatments in exploratory head-to-head comparisons for hospitals specifically or the benchmark in specific patient populations.

Figure 5. (A) Time-to-next-treatment of lung cancer patients treated with first-line pembrolizumab or pembrolizumab + pemetrexed between 2017 and 2020 and (B) overall survival of lung cancer patients treated with first-line pembrolizumab or pembrolizumab + pemetrexed between 2017 and 2020.
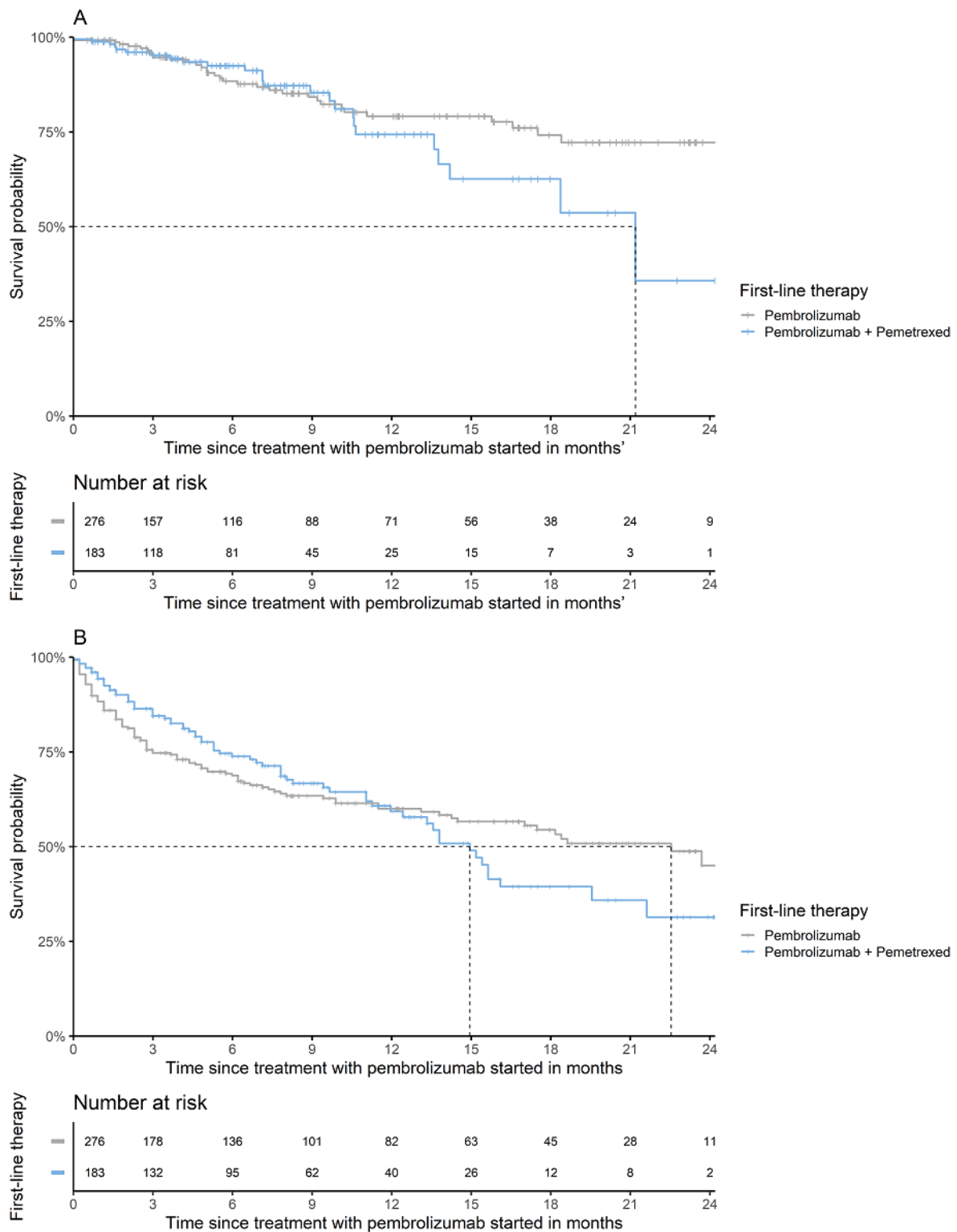
Costs

Use of the financial and administrative database from hospital pharmacies provided us with access to detailed information on the costs of systemic therapies (total costs per treatment and costs per patient) in certain subgroups. Hospitals can upload their paid prices to the dashboard, which is then connected to the medicine and patient information (Multimedia Appendix 5). Prices paid by other hospitals are not shown due to

confidential agreements between pharmaceutical companies and hospitals.

Discussion

Principal Findings

This paper reports on the initial results on the potential applications of data from the DICA Medicines Program; in this program, RWD are generated by linking 4 data sources, including data from quality registries, financial pharmacy data, in-hospital activities systems data, and reimbursement data from 21 Dutch hospitals. In this paper, we reported on the potential of this program in terms of benchmarking, treatment information, clinical outcomes, and costs. To be able to use the data as benchmark information, the data were visualized in web-based dashboards available to clinicians, insurers, and researchers; this led to insights on medication use, clinical outcomes, and costs without any additional registration burden for hospitals. Benchmarking hospital performance is relatively uncommon in the field of medical oncology in contrast to surgical oncology, where many quality registries exist that monitor the quality of care in every hospital [10]. Benchmarking information can support hospital pharmacists, oncologists, and other medical professions involved in the systemic treatment of patients to reach a certain level of care. RWD on the use and efficacy of systemic therapies are needed in daily clinical practice. As the real-world setting differs from the RCT setting, these data are needed after marketing authorization. This project provides real-world evidence, for which there is growing interest. One should be cautious when making definitive conclusions based on observational data. Minor observed differences could be the result of unknown confounding factors [24]. Other initiatives on the linkage of administrative data are similar and link patient-centered health data such as patient-reported outcome measures and clinical laboratory measurements but involve small patient groups [25] or limited patient and tumor characteristics [12].

Strengths

First, data are validated at the time of delivery from the hospitals with the clinicians. A lot of effort is put into the validation of the algorithms that are used in the dashboards, for example, in building Sankey diagrams for treatment sequences in specific patient populations. The second strength of the DICA Medicines Program is the use of existing data sources, thereby minimizing the extra registration burden for medical specialists. This strategy could also be used by other parties to minimize registration burden and maximize the value of available RWD sources. Variables that could easily be derived from the declaration data were the number of expenses, start dates of medications, and the total dosages. Third, the program consists of many participating hospitals within a widespread geographic location, resulting in the inclusion of many patients, who are representative of the Dutch population. Another strength is the linkage of survival data to the other data sources. The database from the national health insurers is a valid source as health care insurance coverage stops when a patient dies. The final strength is that the data are up-to-date and representative of the current situation. This is especially valuable in situations such as the

COVID-19 pandemic, where the systemic treatment of some patients with cancer was adjusted. Since the data are updated quarterly, it was possible to monitor the impact of COVID-19 in certain subgroups of patients in the dashboard. The DICA Medicines Program led to various insights into medication use. Questions related to the use of (expensive) medicines can be answered using the dashboards, in which users can select patient populations or treatments of interest.

Limitations and Future Perspectives

In the clinical registries used for this study, some indications had incomplete data. The DCRA only includes patients undergoing surgical operations, which leads to incomplete clinical data on patients with metastasized cancers and colorectal cancers. This was also the case for patients with metastasized breast cancer. In this subpopulation of patients with breast cancer, essential tumor information, such as receptor status, is lacking. In addition, we are unable to extract information about weight, response status, date of progression, or toxicities from the declaration data. These are mostly data registered in unstructured text in electronic medical records. However, our intention is to complement the clinical data of these patient groups with other techniques that do not lead to further registration burden, such as text mining. Second, due to privacy regulations in the Netherlands [26], it is not permitted to follow-up on patients when they are referred to other hospitals for treatment. An individual patient may seek a second opinion from another hospital. This may have led to incomplete treatment information and individual patients being included twice in the database. Especially for university hospitals, where many patients are referred, it is necessary to have the complete treatment information. Previous analyses on the entire population of patients with lung cancer showed this was the case in <5% of all patients in the Netherlands. In this study, there may be an overestimate in the number of patients but not the number of prescriptions as these are validated declarations made by the hospitals. Third, more information on patient and tumor characteristics is needed to allow for head-to-head comparisons of medicines. Registries should therefore include information on response status and detailed treatment-related toxicity within each line of treatment. At this moment, emergency room visits and hospital admissions are linked to the use of medications and presented in the dashboards. However, these are only surrogate outcomes and do not give insight into the exact response or toxicity. Adding more outcomes of systemic therapies will also be an opportunity for surgical quality registries to become multidisciplinary, where both surgeons and medical oncologists register specific patients' characteristics and outcomes. We are currently exploring text mining opportunities to add information on toxicities and response statuses to the quality registries.

Presently, hospitals use dashboards to benchmark their results against those of other hospitals and gain insights into the use of medications and patient populations, as we showed in this study. The dashboards can also be used for multiple other purposes and by different stakeholders in the future. First, dashboards and RWD can serve as communication tools between physicians and their patients. Based on specific patient and tumor characteristics, clinical outcomes can help patients better

understand their disease course and improve shared decision-making. Second, registration authorities can also benefit from data as presented in this study. Data on newly approved medicines used in clinical practice are included in financial pharmacy data and can be linked to population-based registries. Especially for postapproval measurements, this information is valuable in monitoring the safety and effectiveness of medicines [27]. This can, in certain cases, eventually lead to the replacement of postapproval clinical studies, which will save time and financial resources. The European Medicines Agency and US Food and Drug Agency are increasingly interested in RWD for the evaluation of medicines [28,29]. Furthermore, health care insurers are interested in these data for reimbursement and effective use of expensive medications in the real-world setting [30].

In the future, accurate data from DBC's and financial information could automatically prefill quality registry items. The DICA quality registry items are now entered manually, which is time-consuming and prone to registration errors. Reusing these data sources will lower the registration burden, reduce missing data, and validate data. These data can be used to complete registries and reduce hospital differences. Furthermore, RWD can also be used in health technology assessment decisions. This will be explored in the near future within European Union programs [31]. However, other data sources, such as pathology databases, must be linked to enrich the data. This additional data on histopathology and mutation status are essential as certain medications targeting specific mutations can influence outcomes. To improve shared decision-making, additional data sources, including patient-reported outcome measurements, must be linked to existing data sources.

Conclusions

The DICA Medicines Program has shown that it is possible to gather and link RWD sources pertaining to medicines. In addition, these data became available with minimal registration burden and effort for hospitals. This method of providing RWD can be used in other population-based registries. The DICA Medicines Program provided participating centers with benchmark information and tools to evaluate the effectiveness of expensive medicines in real-world settings.

Conflicts of Interest

AE has advisory relationships with Amgen Inc, Bristol Myers Squibb, F. Hoffmann-La Roche AG, Novartis International AG, MSD, Laboratoires Pierre Fabre, Sanofi SA, Pfizer Inc, Ipsen SA, and Merck & Co Inc, and has received research grants not related to this paper from Sanofi SA, F. Hoffmann-La Roche AG, Bristol Myers Squibb, Idera Inc, and Teva Pharmaceutical Industries Ltd. AE has received travel expenses from MSD Oncology, F. Hoffmann-La Roche AG, Pfizer Inc, and Sanofi SA, and has received speaker honoraria from Bristol Myers Squibb and Novartis International AG. All other authors declare no conflicts of interest.

Multimedia Appendix 1

The medication page of the dashboard shows the active ingredients in a selected patient population and the percentage of patients treated with that active ingredient versus the benchmark. This page also shows the trend over the years. This overview can be adjusted with the filters to show the use of medicines in a specific population or year. Furthermore, it is also possible to receive an overview of the medicines with the number of courses instead of usage. The patients list contains the data of all patients that are selected for the shown results.
[PNG File , 96 KB-Multimedia Appendix 1]

Multimedia Appendix 2

The signals page of the dashboard shows the insights of the dashboard in which a hospital deviates from the benchmark. There are also some specific signals to stimulate improving patient care.
[PNG File , 105 KB-Multimedia Appendix 2]

Multimedia Appendix 3

There is a deep-dive function for every medicine in the dashboard that shows detailed information about the use of these medicines. The deep-dive also shows baseline patient and tumor characteristics of patients treated with this specific medicine, compared to the benchmark.
[PNG File , 131 KB-Multimedia Appendix 3]

XSL•FO
RenderX

Multimedia Appendix 4

Details on the diagnostics are included per medicine in the dashboard for the hospital versus the results in the benchmark. These are related to the moment of medicine use (pre- or post treatment).

[PNG File , 143 KB-Multimedia Appendix 4]

Multimedia Appendix 5

The costs page of the dashboard shows the total costs and the costs per patient for the hospital and the benchmark. This page can be adjusted for specific patient populations by using the filters.

[PNG File , 116 KB-Multimedia Appendix 5]

References

1. Naci H, Davis C, Savović J, Higgins JPT, Sterne JAC, Gyawali B, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. BMJ 2019 Sep 18;366:l5221. [doi: 10.1136/bmj.l5221] [Medline: 31533922]
2. Jones DS, Podolsky SH. The history and fate of the gold standard. Lancet 2015 May 18;385(9977):1502-1503. [doi: 10.1016/S0140-6736(15)60742-5] [Medline: 25933270]
3. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ Open 2016 Jun 30;6(6):e011666 [FREE Full text] [doi: 10.1136/bmjopen-2016-011666] [Medline: 27363818]
4. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? J Clin Oncol 2012 May 01;30(10):1030-1033. [doi: 10.1200/JCO.2011.38.7571] [Medline: 22370321]
5. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
6. Uitgaven aan dure geneesmiddelen stijgen naar 2,1 miljard. Phamaceutisch Weekblad. 2019. URL: https://www.pw.nl/nieuws/2019/uitgave-dure-medicatie-stijgt-naar-20ac-2-1-miljard [accessed 2022-05-27]
7. Cramer-van der Welle CM, Peters BJ, Schramel FM, Klungel OH, Groen HJ, van de Garde EM, Santeon NSCLC Study Group, Santeon NSCLC study group. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur Respir J 2018 Dec;52(6):1801100 [FREE Full text] [doi: 10.1183/13993003.01100-2018] [Medline: 30487206]
8. van Zeijl MCT, Ismail RK, de Wreede LC, van den Eertwegh AJM, de Boer A, van Dartel M, et al. Real-world outcomes of advanced melanoma patients not represented in phase III trials. Int J Cancer 2020 Dec 15;147(12):3461-3470. [doi: 10.1002/ijc.33162] [Medline: 32559817]
9. Beck N, van Bommel AC, Eddes E, van Leersum NJ, Tollenaar R, Wouters M, Dutch Clinical Auditing Group. The Dutch Institute for Clinical Auditing: achieving Codman's dream on a nationwide basis. Ann Surg 2020 Apr;271(4):627-631. [doi: 10.1097/SLA.0000000000003665] [Medline: 31972639]
10. Van Leersum N, Snijders H, Henneman D, Kolfschoten N, Gooiker G, ten Berge M, Dutch Surgical Colorectal Cancer Audit Group, et al. The Dutch surgical colorectal audit. Eur J Surg Oncol 2013 Oct;39(10):1063-1070. [doi: 10.1016/j.ejso.2013.05.008] [Medline: 23871573]
11. Hoeijmakers F, Beck N, Wouters MWJM, Prins HA, Steup WH. National quality registries: how to improve the quality of data? J Thorac Dis 2018 Oct;10(Suppl 29):S3490-S3499. [doi: 10.21037/jtd.2018.04.146] [Medline: 30510784]
12. van Herk-Sukel MP, van de Poll-Franse LV, Lemmens VE, Vreugdenhil G, Pruijt JF, Coebergh JWW, et al. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System. Eur J Cancer 2010 Jan;46(2):395-404. [doi: 10.1016/j.ejca.2009.09.010] [Medline: 19811904]
13. Ismail R, Schramel F, van Dartel M, Hilarius D, de Boer A, Wouters M, Dutch Lung Cancer Audit Scientific Committee. The Dutch Lung Cancer Audit: nationwide quality of care evaluation of lung cancer patients. Lung Cancer 2020 Nov;149:68-77 [FREE Full text] [doi: 10.1016/j.lungcan.2020.08.011] [Medline: 32979634]
14. van Bommel AC, Spronk PE, Vrancken Peeters MT, Jager A, Lobbes M, Maduro JH, NABON Breast Cancer Audit. Clinical auditing as an instrument for quality improvement in breast cancer care in the Netherlands: the national NABON Breast Cancer Audit. J Surg Oncol 2017 Mar;115(3):243-249. [doi: 10.1002/jso.24516] [Medline: 27885679]
15. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med 2005 May 30;24(8):1185-1202. [doi: 10.1002/sim.1970] [Medline: 15568194]
16. Rakow T, Wright RJ, Spiegelhalter DJ, Bull C. The pros and cons of funnel plots as an aid to risk communication and patient decision making. Br J Psychol 2015 May;106(2):327-348. [doi: 10.1111/bjop.12081] [Medline: 25123852]
17. van der Werf LR, Voeten SC, van Loe CMM, Karthaus EG, Wouters MWJM, Prins HA. Data verification of nationwide clinical quality registries. BJS Open 2019 Dec;3(6):857-864 [FREE Full text] [doi: 10.1002/bjs5.50209] [Medline: 31832593]
18. NZA. Nederlandse Zorgautoriteit. URL: https://www.nza.nl/english [accessed 2022-05-31]
19. Oostenbrink JB, Rutten FFH. Cost assessment and price setting of inpatient care in The Netherlands. The DBC case-mix system. Health Care Manag Sci 2006 Aug;9(3):287-294. [doi: 10.1007/s10729-006-9096-y] [Medline: 17016935]

XSL•FO

RenderX

20. Vektis. 2021. URL: https://www.vektis.nl/ [accessed 2022-05-27]
21. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Software 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
22. Yoshida K, Bartel A. tableone: create 'Table 1' to describe baseline characteristics with or without propensity score weights. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/package=tableone [accessed 2022-05-27]
23. Kassambra A, Kosinski M, Biecek P. survminer: drawing survival curves using 'ggplot2'. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/package=survminer [accessed 2022-05-27]
24. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
25. Dhruva SS, Ross JS, Akar JG, Caldwell B, Childers K, Chow W, et al. Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. NPJ Digit Med 2020;3:60. [doi: 10.1038/s41746-020-0265-z] [Medline: 32352038]
26. Quintel T. Article 29 Data Protection Working Party Opinion on the Law Enforcement Directive. European Commission. 2018. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp131_en.pdf [accessed 2022-05-27]
27. Ismail R, Sikkes N, Wouters M, Hilarius D, Pasmooij M, van den Eertwegh F, et al. 1120P Post-approval trials versus patient registries: Comparability of advanced melanoma patients with brain metastases. Ann Oncol 2020 Sep 01;31(1):S754-S766 [FREE Full text] [doi: 10.1016/j.annonc.2020.08.1243]
28. Patient registries. European Medicines Agency. 2015 Sep. URL: https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries [accessed 2022-05-27]
29. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and food and drug administration staff. US Food and Drug Administration. 2017. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices [accessed 2022-05-27]
30. Roberts MH, Ferguson GT. Real-world evidence: bridging gaps in evidence to guide payer decisions. Pharmacoecon Open 2021 Mar;5(1):3-11. [doi: 10.1007/s41669-020-00221-y] [Medline: 32557235]
31. New methods for the effective use of real-world data and/or synthetic data in regulatory decision-making and/or in health technology assessment. European Commission. URL: https://www.ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-hlth-2022-tool-11-02 [accessed 2022-05-27]

Abbreviations

CT: computerized tomography
DBC: diagnosis treatment combinations
DCRA: Dutch Colorectal Audit
DICA: Dutch Institute for Clinical Auditing
NBCA: National Breast Cancer Organization Breast Cancer Audit
NVZA: Dutch Association for Hospital Pharmacists
OS: overall survival
RCT: randomized controlled trial
RWD: real-world data
TTNT: time-to-next-treatment
VGZ: Institute of Public Healthcare

XSL•FO

RenderX

Review

Video Relay Interpretation and Overcoming Barriers in Health Care for Deaf Users: Scoping Review

Minerva Rivas Velarde[1], PhD; Caroline Jagoe[2], PhD; Jessica Cuculick[3], EdD

[1]Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva, Geneve, Switzerland

[2]Department of Clinical Speech & Language Studies, Trinity College Dublin, University of Dublin, Dublin, Ireland

[3]Department of Liberal Studies, National Technical Institute for the Deaf, Rochester Institute of Technology, Rochester, NY, United States

Corresponding Author:
Minerva Rivas Velarde, PhD
Department of Radiology and Medical Informatics
Faculty of Medicine
University of Geneva
CMU 1 Rue Michel Servet
Geneve, 1211
Switzerland
Phone: 41 223794600
Fax: 41 22 379 4600
Email: minerva.rivas@unige.ch

Abstract

Background: Persons who are deaf are more likely to avoid health care providers than those who can hear, partially because of the lack of means of communication with these providers and the dearth of available interpreters. The use of video remote interpretation, namely the video camera on an electronic device, to connect deaf patients and health providers has rapidly expanded owing to its flexibility and advantageous cost compared with in-person sign language interpretation. Thus, we need to learn more about how this technology could effectively engage with and respond to the priorities of its users.

Objective: We aimed to identify existing evidence regarding the use of video remote interpretation (VRI) in health care settings and to assess whether VRI technology can enable deaf users to overcome barriers to interpretation and improve communication outcomes between them and health care personnel.

Methods: We conducted a search in 7 medical research databases (including MEDLINE, Web of Science, Embase, and Google Scholar) from 2006 including bibliographies and citations of relevant papers. The searches included articles in English, Spanish, and French. The eligibility criteria for study selection included original articles on the use of VRI for deaf or hard of hearing (DHH) sign language users for, or within, health care.

Results: From the original 176 articles identified, 120 were eliminated after reading the article title and abstract, and 41 articles were excluded after they were fully read. In total, 15 articles were included in this study: 4 studies were literature reviews, 4 were surveys, 3 were qualitative studies, and 1 was a mixed methods study that combined qualitative and quantitative data, 1 brief communication, 1 quality improvement report, and 1 secondary analysis. In this scoping review, we identified a knowledge gap regarding the quality of interpretation and training in sign language interpretation for health care. It also shows that this area is underresearched, and evidence is scant. All evidence came from high-income countries, which is particularly problematic given that most DHH persons live in low- and middle-income countries.

Conclusions: Furthering our understanding of the use of VRI technology is pertinent and relevant. The available literature shows that VRI may enable deaf users to overcome interpretation barriers and can potentially improve communication outcomes between them and health personnel within health care services. For VRI to be acceptable, sign language users require a VRI system supported by devices with large screens and a reliable internet connection, as well as qualified interpreters trained on medical interpretation.

XSL•FO
RenderX

KEYWORDS

deafness; disability; accessibility; communication; video; remote interpretation; health care; system; deaf users; sign language; interpreter; medical interpretation; mobile phone

Introduction

Background

Accessible information and communications technology (ICT), mobile phones, and tools such as video remote interpretation (VRI) aim to enable effective communication between persons who are D/deaf ("Deaf" refers to the linguistic minority while "deaf" refers to persons with hearing impairment) and hard of hearing and those who use sign language as their first language (hereafter, deaf or hard of hearing [DHH] sign language users) and health care personnel. VRI refers to a video camera on an electronic device, either a computer or tablet, that is used to connect patients and health providers with a sign language interpreter via video call. Its use has rapidly expanded owing to its flexibility and advantageous cost compared with in-person sign language interpretation [1]. The cost-efficiency of such technology is a serious concern given that 80% of the DHH population live in low- and middle-income countries (LMICs), where resource constraints tend to limit the availability of qualified sign language interpreters [2]. VRI aims to overcome communication barriers in health care. DHH persons are more likely to avoid health care providers than those who can hear, partially because of the lack of means of communication with these providers and the dearth of available interpreters [3,4]. Even if interpreters are available, the pool of sign language interpreters tends to be relatively narrow, even in high-income contexts [5]. Forthcoming research suggests that general sign language training does not cover skills to work effectively within the health care context; therefore, issues arise from the limited number of interpreters and their lack of skills [6-8]. Furthermore, health care personnel tend to lack awareness about working with sign language interpreters, alongside limited awareness of deaf communities in general. This results in poor communication, and ultimately, patients do not obtain the information they need to decide on their health or treatment [5].

DHH populations tend to be particularly disadvantaged compared with other persons with a disability. They tend to occupy poorer socioeconomic positions, hold lower health literacy, have insufficient knowledge of health-related vocabulary, and are often unaware of their family medical histories, all of which prevent them from outlining risk factors for their health [9]. DHH individuals have a greater prevalence of obesity, higher levels of hypertension, and higher levels of self-reported depression compared with hearing persons [6,9,10]. There is also a particular concern of underdiagnoses of raised blood pressure and undertreatment of hypertension, diabetes, hyperlipidemia, and cardiovascular disease, among others, due to the lack of effective means of communication between health personnel and deaf patients [6,9-11]. Recent studies claim that by improving communication between deaf persons and nondeaf persons hearing health personnel would have a positive impact on preventive care [12-14].

Objective

The rapid adoption of VRI technology in health care opens up opportunities to set up more accessible health care. Thus, we need to learn more about how this technology could effectively engage with and respond to the priorities of its users. Emerging literature shows that DHH users tend to prefer in-person to VRI interpretation [15-17]. Furthermore, satisfaction with VRI interpretation tends to be low [15]. We do not have evidence on whether users are comparing interpreters with the same level of skills one via VRI and one in-person, so they are comparing the sentiment of indeed like with like or not. Thus, we need more clarity on the elements of VRI systems that have been examined, such as procedures, available protocols, challenges, and successes. Having detailed data, all elements regarding in-person and VRI interpretation protocols would allow determining the technology that holds some constraints more clearly or the protocol could be improved and made more efficient. It is also necessary to identify the essential elements of VRI as a precondition to encourage rigorous studies and ensure fidelity when implemented. The scoping review approach chosen for this study will allow us to determine the state of available evidence, which is needed before rigorous empirical studies are conducted. Therefore, for the purpose of this study, we used the guidance for conducting systematic scoping reviews by Peter et al [18] to determine the following with respect to the use of VRI in the health care context: does the existing literature provide sufficient evidence on how VRI can enable deaf users to overcome interpretation barriers and improve communication outcomes between them and health care personnel within health care settings?

Methods

Overview

In this review, we identified relevant studies in English, Spanish, and French published between 2006 when the first relevant publication in the area was identified and March 2021 in PubMed, Web of Science, Embase, MEDLINE, and Google Scholar. The key search terms used were as follows: Sign language user*s, Deaf, Hard of Hearing, Deafblind and VRI, video remote sign language interpretation, video interpreting service, video conference interpreting and community health, health system, and health personnel. The search also covered all types of health-related activities that are often linked to community health. See search strategies in Multimedia Appendix 1.

Study Selection
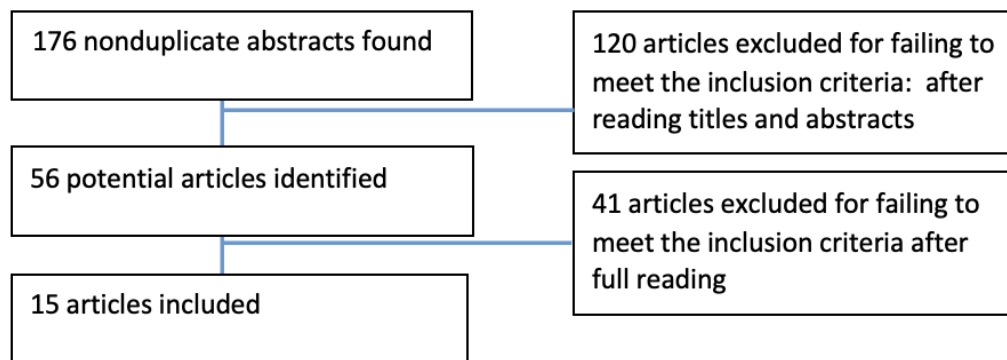
Articles were included for full-text reviews if they were about the use of VRI for DHH users for, or within, health care. Titles and abstracts were screened, and if an article was considered representative of the inclusion criteria, the full text was reviewed. Data extraction was conducted by 2 reviewers, independently, on 20% of the papers. The discrepancies were minimal.

XSL·FO

RenderX

If the paper was selected for full review, data related to the use of VRI for sign language users within the health care context were extracted. Data extracted from the articles that reported on the analysis, use, or implementation of VRI within the health care context were entered into an Excel (Microsoft Inc) form.

Key findings were extracted in a summary format. Information on authorship, publication year, article type, methodology, population, lessons learned, and recommendations regarding the use of VRI were recorded in this form (Figure 1).

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


```
176 nonduplicate abstracts found  ──────  120 articles excluded for failing to
                                           meet the inclusion criteria: after
                                           reading titles and abstracts

56 potential articles identified  ───────  41 articles excluded for failing to
                                           meet the inclusion criteria after
                                           full reading

15 articles included
```

Analysis

We conducted an inductive content analysis of the selected records following the steps outlined by Elo [19]. The extracted findings from each study were subjected to open coding, and similar codes across articles were then identified as concepts coded inductively into the key concepts. Finally, in line with the aims of the study, the concepts were grouped into either the advantages of VRI or the challenges or limitations of VRI.

Patient and Public Involvement

This study was performed without the involvement of DHH patients. However, it does involve organizations for DHH individuals as well as persons with a disability. The National Deaf Federation of Colombia (FENASCOL) advised MRV on the pertinence of this research. JC, a DHH scientist, has coauthored this paper, contributing to its conceptualization, interpretation of the results, and attainment of clarity and accuracy of the writing.

Ethics Approval

The research protocol of this study was approved by the ethics committee of the University of Geneva (CUREG_2021-05-50).

Results

Overview

From the original 176 articles identified, 120 were eliminated after reading the article title and abstract, and 41 articles were excluded after they were fully read. In total, 15 articles were included in this study: 4 studies were literature reviews, 4 were surveys, 3 were qualitative studies, and 1 was a mixed methods study that combined qualitative and quantitative data, 1 brief communication, 1 quality improvement report, and 1 secondary analysis. Table 1 includes summaries of the articles that met our inclusion criteria.

There is limited research on the use and efficiency of VRI to improve communication between DHH individuals and health personnel within health settings. The current published scientific literature does not allow us to understand either the use of this technology or its impact on quality of care, patient satisfaction, or health outcomes. Nearly half (n=7, 46%) of the articles included empirical evidence on adult DHH VRI users, 1 (6%) on DHH children, 1 (6%) on sign language interpreters, and 1 (6%) on subject matter experts working with older DHH adults. Less than half (n=6, 40%) of the articles explicitly addressed the role of DHH persons as coauthors of the articles and steps followed to fulfill ethical and moral obligations of putting the voice of the DHH population at the center of their research, promoting well-being and the human rights of this population.

A limitation of the available literature is the lack of representation of the DHH population as a whole, given that all the articles are from high-income countries, namely 12 from the United States, 1 from Denmark, 1 from Norway, and 1 from Canada. This is a significant gap, given that 80% of persons with disabling hearing loss live in LMICs [19]. Currently, resource constraints and other social and political barriers in LMICs that could affect the availability, use, and efficiency of sign language interpretation via VRI within health care are not included in the published literature.

The current literature shows the key advantages of pursuing improvements in this technology as well as some recurring challenges and limitations (Textbox 1).

Table 1. Summaries of studies included in this review.

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Berry and Stewart, 2006 [20]	United States	To outline challenges that D/deaf[b] people face within health care. It outlines recommendations to ensure a successful medical visit.	D/deaf	Literature review	No information	• Suggest capacity building for medical staff regarding communication needs of D/deaf patients. It provides a protocol to identify interpreters, as well as a list of tips for working with an interpreter, such as speaking to patients when using an interpreter.
Steinberg et al, 2006 [21]	United States	To better understand the health care experiences of deaf people who communicate in ASL[c]	Participants were deaf, communication preference for ASL, and willingness to share health care experiences	Qualitative studies (semistructured focus group meetings)	No information	• It points out that fear, mistrust, and frustration were prominent in participants' descriptions of health care encounters, as well as a list of inadequate common practices such as writing notes and using family members as interpreters.
Masland et al, 2010 [1]	United States	This study reviews published literature and unpublished data, documenting the use of telephonic and video interpretation methodologies to improve health care communication.	Published and unpublished literature on the interpretation in health care	Brief communication	No information	• This study looks at the cost-effectiveness of VRI for all language translation including sign language. • VRI advantages outlined in the study are flexibility, convenience, quality of interpretation, and cost. Some arguments are made that the savings in hiring an ASL interpreter can pay for the installation of video interpretation networks in some hospitals. • The results linked the use of VRI to fewer tests, less visits to the hospital, and better treatment adherence. However, evidence represented is in spoken leagues not sign language.
Hommes et al, 2018 [22]	United States	This research aimed to identify ASL interpreters' perceptions of barriers to effective communication between deaf and HOH[d] patients and health care providers.	ASL interpreters	A cross-sectional survey	June 15	• The results indicated that VRI technology in the absence of an ASL interpreter is considered a better option by many deaf and HOH patients than note-writing or lip-reading; however, the occasional technology malfunctions limit it as a consistently reliable tool.

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Dammeyer et al, 2017 [23]	Denmark	This study examined the prevalence of technology use and interpreting services use among people with hearing loss as they relate to demographic characteristics of this population.	269 children (0-15 years of age) and 839 adults (16-65 years of age)	National surveys of children and adults with hearing loss	2014	• This study found that sign language users, both children and adults, prefer VRI over other communication technology. Adults with a bachelor's degree or higher reported more frequent use of mobile video interpretation and texting devices. • This study underlines the need for a user-centered approach and user involvements to address environmental and personal factors affecting assistive technology use. It recommends that deaf people may benefit from accessing well-trained personnel who understand the individual's needs and facilitate technology-person match.
Myers et al, 2021 [16]	United States	To examine the extent to which communication aids and services used by ASL users and their health care providers aligns with preferences, satisfaction, and unmet needs and to elicit from stakeholders' strategies to address disparities	ASL users in North Carolina	Web-based survey (cross-sectional study)	May 2018 until March 2019	• The study found that accessible communication was associated with 81% lower odds of dissatisfaction with communication. • Better communication was linked to better relationships with the health providers. The study claims that improving communication would have a positive impact on preventive care. • The study identifies several issues with the use of VRI. One of the most common barriers to accessible communication via VRI were technical problems, as well as quality of sign language interpreting services. • Communication via VRI was considered not user-friendly, creating frustrations for both deaf individuals and their professional health care providers. Health providers attempted to adapt to VRI issues by lipreading or speech or writing notes back-and-forth, both methods were inadequate and did not lead to improved communication. • The study made specific technical recommendations on when and how to use VRI in clinical settings.
Kushalnagar et al, 2019 [15]	United States		Persons that use ASL as a primary language, age of 18 years or above, and presence of bilateral hearing loss	Secondary Analysis of National health trends Survey in ASL	Between 2016 and 2018	

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
		This study aimed to investigate the national trends of deaf patients' satisfaction with the quality of VRI in health settings and recommend actions to improve VR quality and deaf patients' satisfaction with VRI in health care settings.				• The study shows that almost half of the people reached by the survey did not have access to VRI over the last 12 months. • It also shows that those who have access were largely dissatisfied with the quality of the service. About 41% (n=228) of the deaf patient sample rated the quality of VRI as satisfactory. The rest (n=327, 59%) rated their VRI experience as unsatisfactory. • VRI tends to be cost-effective and its flexibility is of great advantage to service providers, users, and interpreters. • The study claims that if D/deaf ASL health care users are provided with a fully functioning VRI system with qualified interpreters, this system can potentially reduce the number of emergency visits and unnecessary diagnostic tests, all of which are associated with cost burden.
Yabe, 2020 [17]	United States	This study identifies health care providers' and DHH[e] patients' interpreting preferences for VRI and in-person interpretation during critical care and noncritical care	1. Health care providers who had used VRI in clinical settings in the past 10 years were 18 years or older and spoke English. 2. DHH patients who had used VRI in clinical settings in the past 10 years were 18 years or older and used ASL	Mixed methods design incorporating both an online survey and qualitative interviews	No information	• This study provides the views of both health workers and sign language users—the findings pointed out that VRI is the preferred way of communication of patients and health providers for noncritical care. • VRI offers preparedness unattainable with in-person interpretation. Furthermore, in-person interpretation is limited in its availability and represents at times economic loss. • It outlines technical limitations regarding VRI and recommendation for its use. • It points out that patient's acceptance of VRI was linked to time constraints and type of care. Thus, acceptance was limited as it was described as waste of money as it did not prove effective for communication. • For providers, its convenience and flexibility were very important.
Kushalnagar et al, 2017 [24]	United States		Deaf adults (ages 18-90 years and above) who use ASL	Qualitative studies (cognitive interviews)	N/A[g]	• This article outlines the protocol of cultural adaptation national survey items exploring VRI. • Linguistic adaptation of items related to time, explanation of illness and use of diagrams, captions and videos is very useful for validation studies using sign language.

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
		The objectives of this study are (1) to culturally adapt and linguistically translate the HINTS[f] items to ASL (HINTS-ASL) and (2) to gather information about deaf people's health information–seeking behaviors across technology-mediated platforms.				
Singleton et al, 2019 [25]	United States	This study explored technology use among older deaf adults with regard to attitudes, adoption style, and frequency of use for a wide range of technologies, including ATs[h] for persons with hearing loss and general everyday technologies.	Participants had to be 50 years of age or older and self-identify as DHH	Online or paper copy questionnaire	__i	• Older adults are moving away from TTY[j] and TDD[k] to embrace VP[s][l] and VRS[m]; 51% of respondents use VRI. They noticed that consumer service and support such as free delivery and personnel to set technology up had a very positive impact on the consumer experience. • Participants reported difficulty keeping up with software updates and other technology maintenance activities that require a higher level of computer literacy. Thus, many older adults in the deaf community appear to be comfortable with daily technologies and ATs and especially video-based internet technologies that support communication accessibility such as VP and VRS.
Kasales et al, 2020 [26]	United States	The goal of this review is to help members of the breast center team better understand (1) the mandates of the ADA[n] and the challenges faced by patients with select communication disabilities.	Descriptive review	Literature review (descriptive review)	N/A	• This article reviews some relevant literature and points out recommendations to use VRI. However, it does not include any empirical evidence. • They recommend using VRI when an in-person interpreter is not available and only in agreement with the patient. • It lays out the recommendation of the National Association of the Deaf Seniors of America for the use of VRI for ASL communication.
Meulder and Haualand, 2019 [27]	Norway	To critically assess the impact and role of SLIS[o] in those countries where SLIS have been institutionalized	VRI deaf users	Literature review (conceptual analysis)	N/A	

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
						• This article presents an analysis of the role that sign language interpretation has in social services including health care. • The paper makes a strong argument for the importance of language-concordant services. • It does refer broadly to sign language interpretation including VRI. It highlights that access and communication in the health care setting are mainly conceptualized and arranged with a hearing person's perspective. Little has been done to allow health settings or personnel to be bilingual and therefore more accommodating to the sign language users, cultural gaps, discriminatory set up, and other issues might not be apparent to the interpreter and shall be considered.

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
Preusse et al, 2016 [28]	United States	The goal of this study was to identify the range of challenges in everyday activities that might be experienced by older adults aging with preexisting impairments in vision, hearing, or mobility.	Interviews with subject matter experts working with older deaf adults	Qualitative study (interviews)	—	• Findings of the study revealed challenges faced by deaf persons as they age. These challenges include access to social services, adequate housing, and technology. • The findings state that access to interpreters is an issue in most health settings. Experts interviewed pointed out that this shortage of qualified sign language interpreters can be overcome by using VRI. Thus, they also pointed out that VRI may be inappropriate when people are dealing with high levels of stress such as a medical emergency. In these cases, in-person interpretation may be more appropriate, if available. • The findings show that device maintenance and software updates are difficult for this population. The study recommends one-to-one training for uptake of new technologies, as well as mixed available technologies such as haptic devices as medication reminders.
McKee et al, 2015 [29]	United States	The aim of this paper is to summarize evidence and good practices on how to enable better communication between DHH and health personnel, particularly physicians.	—	Literature review	N/A	• This paper offers an overview of good practices and questions regarding health service provision for DHH patients. • It lays out that DHH patients are more likely to experience poverty and less likely to access ICTp including smartphones. • VRI is mentioned as a tool to overcome communication barriers and improve satisfaction, quality of care, and health outcomes. However, it also mentioned that evidence on the impact of interpretation and VRI is lacking. These recommendations assume that interpretation availability either via VRI or in person is an efficient way forward.
Kwok et al, 2021 [30]	Canada	This report documents the experience in using web-based technology in an emergency department to meet communication needs of our patients who have LEP[q] including deaf sign language users during the COVID-19 pandemic.	—	Quality improvement report	March 30 and May 31, 2020	

Study	Country	Aims	Study population and sample size	Design	Duration of the intervention	Main findings that related to the use of VRI[a] within health care
						• This study focuses on the use of VRI more generally for patients of linguistic minorities including sign language. • It reports on the cost-efficiency of the intervention, laying out prices of VRI inclusive of sign language and claiming that such a cost is not problematic to absorb by the hospital. • VRI technical issues were easily overcome and personnel became acquainted to its use relatively easily. Furthermore, the study claims that the use of VRI also complies with security protocols in place during the COVID-19 pandemic and allows the protection of interpreters and others from exposure. • The authors of the paper judged that this intervention was successful for both hearing patients and DHH patients. Thus, there is no evidence that it was the case.

[a]VRI: video remote interpretation.

[b]D/deaf: "Deaf" refers to the linguistic minority while "deaf" refers to persons with hearing impairment.

[c]ASL: American Sign Language.

[d]HOH: hard of hearing.

[e]DHH: deaf or hard of hearing.

[f]HINTS: Health Information National Trends Survey.

[g]N/A: not applicable.

[h]AT: assistive technologies.

[i]Data not available.

[j]TTY: (teletypewriter) is a communication device used by people who are deaf, hard-of-hearing, or have severe speech impairment.

[k]TDD: test-driven development.

[l]VPs: videophones.

[m]VRS: video relay service.

[n]ADA: Americans with Disabilities Act.

[o]SLIS: Nottinghamshire Sign Language Interpreting Service.

[p]ICT: information and communications technology.

[q]LEP: limited English proficiency.

Textbox 1. Summary of advantages and disadvantages.

Advantages

- Convince

- Preparedness unattainable with in-person interpretation

- Access to qualified interpreters

- Possibility to work remotely for interpreters

- Safety, limiting social contact in health care environment

- Cost

- Flexibility

Disadvantages

- Technology malfunctions

- Inaccessible to deaf patients in certain physical positions and those with vision impairment

- Requires higher level of computer literacy

- Not user-friendly

- For some it might limit patient-provider relationship

- Relays on the availability of reliable internet access and adequate devices

---

Advantages of Using VRI Interpretation

Early literature [1,20,21] described sign language interpretation using VRI in health care settings as equally efficient as in-person interpretation. Advantages attributed to the technology, such as flexibility and affordability, encourage the idea that this technology could help overcome the shortage of qualified sign language interpretation in health care settings. It also pointed out that the use of VRI could help to override the use of inadequate techniques such as lipreading and note-reading, which are often used in health consultations with DHH patients. DHH sign language users prefer to use VRI over these techniques primarily because it allows them to communicate in their preferred language, sign language [22,26]. Lipreading and note-reading often assume that sign language users are proficient in reading and writing in a spoken language, which is often not the case. The literacy rates of DHH communities are at a sixth grade reading level or lower [29,31,32].

Articles exploring technology preferences highlight that sign language users (both children and adults) prefer VRI to other communication technology over texting devices (sign language, text, and speech interpretation linked by a call center or voice recognition technology) [16,23]. As the proliferation of VRI technology increases, consumer choices increase. With this technology, deaf patients have the possibility to choose communication tools and assistance that they deem more appropriate for their medical consultation [15-17,27]. For some noncritical medical services, VRI is preferred over in-person interpretation [16,17,24].

Sign language interpreters saw a significant advantage to this technology as it allowed them to eliminate time for transportation, given that most of their time assisting in a medical consultation is consumed by traveling to the location [22]. Saving in traveling time often translates to saving in the total cost of the interpretation. This is a key advantage often mentioned in the literature and an underlying motivation to continue expanding the use of VRI in health care settings [1,15,17,20-22,25,28,30]. VRI has also proven advantageous during the COVID-19 pandemic, allowing qualified interpreters to be available at emergency services while protecting both parties from risking potential exposure at the emergency room and complying with access restrictions [30].

The current literature suggests that the use of VRI to use qualified sign language interpreters, despite where they are located, has the potential to help overcome the scarcity of sign language interpretation and enable better communication between deaf patients and health care personnel. The advantages offered by VRI are likely to be enhanced as technology devices such as tablets, laptop computers, and smartphones become more affordable and reliable internet bandwidth becomes more available [15,17,25].

Challenges and Limitations of the Technology

As evidence grows, we are learning more about VRI technology because of its shortcomings, particularly with regard to the specificities of health care settings. A national survey conducted in the United States showed that only almost half of the representative sample did not have access to VRI during health care appointments over the last 12 months [15]. It is not clear whether the technology was needed but not available, suggesting that even in a high-income context, the availability of this technology remains limited or if participants chose not to use VRI because they had access to in-person interpretation or preferred to use other communication techniques.

Several articles in the hospital context in the United States showed that VRI was not user-friendly and led to frustration for both DHH individuals and their professional health care providers. The most common barriers noted were technical problems and poor quality of sign language interpreting services [16]. Although VRI is preferred for noncritical care, it is

considered inappropriate for critical care or stressful situations [16,25]. A second article reiterated the issues found in the first study and laid out other technical issues, such as limited placement and positioning of devices, negatively impacting the experience of using this technology [17]. They also found that VRI was seen as inconducive to enriching patient-health provider relationships and that providing VRI without previously notifying, seeking, and obtaining the agreement of the patient first was bothersome [17].

A national survey from the United States also looked at preferences of the DHH population between VRI or in-person interpretation within health care settings and found that 59% of their respondents rated their VRI experience as unsatisfactory and preferred in-person interpretation. Sign language interpreters have also reported concerns regarding technology. According to interpreters' views, in-person interpretation is more efficient at identifying when users do not understand a diagnosis, medical instructions, or other information compared with VRI. Interpreters also pointed out that the extra time before and after the appointment is useful for reviewing information available in the lobby and preparing for consultation, which enables them to provide better interpretation services [22]. The VRI does not allow interpreters to prepare or debrief DHH patients before and after consultations [31]. In turn, VRI could be more prone to incomplete communication between DHH and hearing health personnel. Capacity building among health personnel was noted as a significant communication barrier for DHH patients but also as a hindrance to technology development [20].

The efficiency of this technology is partially determined by the appropriateness of the video device used. The recommended screen of a minimum of 49.5 cm (19.5 in) is often not available [26]. Keeping up with software updates among other technologies, maintenance was considered burdensome among older DHH adults [25]. Other reported limitations of the technology included constraints due to the physical position of the patient. VRI is not accessible for patients undergoing clinical examination that requires them to be face down; VRI is also not accessible for DHH persons who are blind or have low vision [16]. The use of electronic means of communication for health information also raises security and privacy concerns. We found no information on whether the video feeds were encrypted.

The literature also shows methodological shortcomings of using health research instruments, such as surveys that explore VRI on DHH individuals, which have been developed and tested only with hearing participants. Given the cultural and linguistic differences between DHH and hearing populations, some concepts, questions, and wording may be inappropriate or incomprehensive for DHH individuals [24].

Adding to the technical and methodological issues, a more troubling challenge was assuming that an efficient VRI technology would be sufficient to overcome barriers to health care for DHH individuals (or communities). Research has shown that the use of VRI services alone is not fully accessible to DHH communities. Little research has been conducted to promote bilingualism or language-concordant practices across health settings or personnel and accessibility in broader health-related communication practices [27]. Furthermore, there is a risk that

the VRI could be conceptualized and put in place from a hearing person's perspective. This limited, 1-sided view ignores issues related to cultural differences and discrepancies, discriminatory practices, intrinsic bias, and intersectionality issues related to hearing status, ethnicity, race, or multiple disabilities.

Discussion

Principal Findings

This scoping review provides an overview of the current evidence on the efficiency of the use of VRI with deaf users within health care settings. It shows that this area is under research, and the evidence is scant. It is particularly concerning that all articles found were from high-income countries, given that most DHH people live in LMICs. There is a dearth of evidence on the use of VRI and its efficiency and potential across LMICs. This reflects the long-lasting absence of voices of persons with disabilities from non-Western nations on both disability scholarship and technology innovation [33-35]. The lack of knowledge regarding the needs and realities of DHH individuals in LMICs extends beyond VRI technology. Technological progress has often overlooked the experience of disability and the everyday needs and constraints of DHH persons from the Global South. Nearly all research on assistive technology and ICT accessibility for DHH individuals and for persons with disabilities, whether from the legal, technical, or development fields, has focused on high-income countries and very little to no attention has been paid to LMICs [36]. Technological progress has often overlooked the experience of disability and the everyday needs and constraints of persons with disabilities from the Global South, among other reasons, because it is perceived as nonprofitable [34]. Failing to address this gap will cause persons with disabilities in LMICs to continue to be left behind in relation to universal health coverage.

At present, 164 countries are signatories to the Convention on the Rights of Persons with Disabilities (CRPD). CRPD Article 25 on health and Article 9 on accessibility provided the legal basis for ensuring the right to the highest attainable standard. Thus, the implementation of the CRPD remains limited, particularly in LMICs. The dominant presence of the literature from the United States may be linked to the Americans with Disabilities Act of 1990 [37], which lays the legal grounds for accessibility and nondiscrimination, as well as for the adoption of reasonable accommodation. However, similar legal frameworks have been adopted in other high-income countries with sufficient infrastructure to provide VRI services, such as the Disability Discrimination Act 2005 [38] in the United Kingdom, and we did not observe the same level of engagement on behalf of either public health or disability scholars. Nevertheless, the implementation of such CRPD rights to health and accessibility in health care settings will require robust evidence regarding the priorities, needs, and constraints of persons with disabilities in LMICs.

A major strength of this review is the use of a comprehensive search in 3 languages in a rapidly expanding technology and a focus on highlighting available evidence and gaps. A key issue highlighted by the available literature is that the availability of

VRI technology has the *potential* to address communication barriers within the health care setting, in addition to other available services and tools aside from, inter alia, in-person interpretation, telephone typewriters, and telecommunications relay services. The views, needs, and rights of the DHH community should be at the core of the development of these technologies. However, the VRI is not a quick fix to overcome accessibility issues [15,27,39]. It is important that its expansion and convenience do not undermine the possibility for DHH communities to choose whichever means of communication they prefer or which is more appropriate for the type of care they seek.

This review also pointed out a knowledge gap regarding the quality of interpretation and training in sign language interpretation for health care. It is not clear if poor-quality interpretation is a recurring issue when using in-person interpretation or if it is only an issue when using VRI [15-17]. There are no data on whether in-person interpretation, as requested in advance, the assigned interpreter is likely to use time before the consultation to undergo a prescreening for interpretation competencies, allowing better preparation for their job. Perhaps interpreters are better matched at the time of assigning the task; thus, we do not know whether this could improve the quality of interpretation. Nor do we know if such prescreening for qualification takes place for VRI interpreters or if such practice would lead to better outcomes and positive experiences across DHH users. There is a gap in the evidence on this issue, although most articles mentioned the pertinence of training for sign language interpreters on health interpretation for better communication outcomes.

The challenges documented in the literature highlight recurring technical issues regarding internet reliability, availability, and adequacy of devices in hospital settings. Although the internet is growing globally [40], it is clear that internet reliability has imposed utmost complex infrastructural challenges that could hamper VRI development in LMICs. The literature is not clear on whether, when VRI is used, users use their own devices or if they have to personally purchase internet data (and devices). This raises questions and concerns, as persons with disabilities are more likely to experience poverty in both high- and low-income countries. The financial challenges of DHH communities will have an impact on access to devices and the internet, and in turn, these challenges will impose further barriers to communication and health care. This is perhaps more acute in the Global South.

For future research, there is a need to raise awareness and build capabilities across health systems to improve accessibility for DHH individuals. The literature suggests that having more bilingual health workers, language-concordant services, better technologies, and raising awareness will contribute to better communication between DHH communities and health personnel [41-47]. New developments include technologies such as intelligent personal assistants such as Alexa, which can be used with sign language to improve communication [48]. Thus, we need to learn more about how to make health systems more accessible to DHH individuals. Accessible communication in health settings has been linked to fewer hospital visits, better treatment adherence, more cancer screening, and better oral health [10,12,14,41,42,49,50].

Comparison With Previous Literature

There have been no similar publications in this area. This study provides a well-needed analysis regarding knowledge gaps and the need for future research on the efficiency of VRI technology for sign language users in the health care context.

Limitations

Our study has a few limitations. We looked at articles examining VRI in health care settings, including hospitals, preventive care, and community health. Few rigorous articles have studied VRI for sign language users in the health care context. The protocols used and examined regarding the use of VRI for sign language are not generalizable at a national level or international level. We attempted to map and assess the available scientific literature.

Conclusions

The available literature shows that VRI may enable deaf users to overcome interpretation barriers and can potentially improve communication outcomes between them and health personnel within health care services. Communication between DHH health care users and personnel shall improve if sign language users are provided with a VRI system supported by devices with large screens and a reliable internet connection, as well as qualified interpreters trained on medical interpretation. Perhaps issues regarding lack of preparation for interpreters could be overcome by providing VRI interpreters with a brief summary of the purpose of the visit, as well as the background of the consultation before the discussion. Such preparation may allow interpreters and users to develop a rapport during health visits, and research is needed in this area.

Furthermore, our understanding of the use of VRI technology is pertinent and relevant. All articles mentioned that sign language interpretation is a scarce resource within health care systems, even in high-income countries. Thus, learning more about the possibilities and limitations of VRI is even more urgent in LMICs, because the dearth of data and in-person interpretation are largely unavailable and perhaps unfeasible in the near future in resource-constrained contexts.

Authors' Contributions

MRV developed the review question. MRV was the first reviewer, and CJ was the second reviewer. MRV and CJ conducted the study and the analysis. MRV drafted the manuscript. MRV, CJ, and JC reviewed and edited the manuscript. MRV is the senior author and acts as guarantor. All authors have reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.
[DOCX File , 28 KB-Multimedia Appendix 1]

References

1. Masland MC, Lou C, Snowden L. Use of communication technologies to cost-effectively increase the availability of interpretation services in healthcare settings. Telemed J E Health 2010;16(6):739-745 [FREE Full text] [doi: 10.1089/tmj.2009.0186] [Medline: 20626299]

2. Zulu T, Heap M, Sinanovic E. The cost and utilisation patterns of a pilot sign language interpreter service for primary health care services in South Africa. PLoS One 2017 Dec 22;12(12):e0189983 [FREE Full text] [doi: 10.1371/journal.pone.0189983] [Medline: 29272272]

3. Kritzinger J, Schneider M, Swartz L, Braathen SH. "I just answer 'yes' to everything they say": access to health care for deaf people in Worcester, South Africa and the politics of exclusion. Patient Educ Couns 2014 Mar;94(3):379-383. [doi: 10.1016/j.pec.2013.12.006] [Medline: 24388666]

4. Bachman SS, Vedrani M, Drainoni ML, Tobias C, Andrew J. Variations in provider capacity to offer accessible health care for people with disabilities. J Soc Work Disabil Rehabil 2007;6(3):47-63. [doi: 10.1300/J198v06n03_03] [Medline: 17989034]

5. Leeson L, Saeed JI. Irish Sign Language: A Cognitive Linguistic Approach. Edinburgh, UK: Edinburgh University Press; 2012.

6. Kuenburg A, Fellinger P, Fellinger J. Health care access among deaf people. J Deaf Stud Deaf Educ 2016 Jan;21(1):1-10. [doi: 10.1093/deafed/env042] [Medline: 26405210]

7. Napier J, Skinner R, Turner G. "It's good for them but not so for me": inside the sign language interpreting call centre. Transl Interpreting 2017 Jul 24;9(2):1-23. [doi: 10.12807/ti.109202.2017.a01]

8. Martín MC, Phelan M. Interpreters and cultural mediators – different but complementary roles. Translocations 2010;6(1):1-19 [FREE Full text]

9. Barnett S, McKee M, Smith SR, Pearson TA. Deaf sign language users, health inequities, and public health: opportunity for social justice. Prev Chronic Dis 2011 Mar;8(2):A45 [FREE Full text] [Medline: 21324259]

10. Berman BA, Jo A, Cumberland WG, Booth H, Britt J, Stern C, et al. Breast cancer knowledge and practices among D/deaf women. Disabil Health J 2013 Oct;6(4):303-316 [FREE Full text] [doi: 10.1016/j.dhjo.2013.05.001] [Medline: 24060253]

11. Emond A, Ridd M, Sutherland H, Allsop L, Alexander A, Kyle J. The current health of the signing deaf community in the UK compared with the general population: a cross-sectional study. BMJ Open 2015 Jan 25;5(1):e006668 [FREE Full text] [doi: 10.1136/bmjopen-2014-006668] [Medline: 25619200]

12. Kushalnagar P, Engelman A, Sadler G. Deaf patient-provider communication and lung cancer screening: Health Information National Trends survey in American Sign Language (HINTS-ASL). Patient Educ Couns 2018 Jul;101(7):1232-1239 [FREE Full text] [doi: 10.1016/j.pec.2018.03.003] [Medline: 29548598]

13. Kushalnagar P, Hill C, Carrizales S, Sadler GR. Prostate-Specimen Antigen (PSA) screening and shared decision making among deaf and hearing male patients. J Cancer Educ 2020 Feb;35(1):28-35 [FREE Full text] [doi: 10.1007/s13187-018-1436-3] [Medline: 30353474]

14. McKee MM, Barnett SL, Block RC, Pearson TA. Impact of communication on preventive services among deaf American Sign Language users. Am J Prev Med 2011 Jul;41(1):75-79 [FREE Full text] [doi: 10.1016/j.amepre.2011.03.004] [Medline: 21665066]

15. Kushalnagar P, Paludneviciene R, Kushalnagar R. Video remote interpreting technology in health care: cross-sectional study of deaf patients' experiences. JMIR Rehabil Assist Technol 2019 Mar 11;6(1):e13233 [FREE Full text] [doi: 10.2196/13233] [Medline: 30855233]

16. Myers MJ, Annis IE, Withers J, Williamson L, Thomas KC. Access to effective communication aids and services among American sign language users across North Carolina: disparities and strategies to address them. Health Commun 2022 Jul;37(8):962-971. [doi: 10.1080/10410236.2021.1878594] [Medline: 33541113]

17. Yabe M. Healthcare providers' and deaf patients' interpreting preferences for critical care and non-critical care: video remote interpreting. Disabil Health J 2020 Apr;13(2):100870. [doi: 10.1016/j.dhjo.2019.100870] [Medline: 31791822]

18. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

19. 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]

20. Berry JA, Stewart AJ. Communicating with the deaf during the health examination visit. J Nurs Pract 2006 Sep 1;2(8):509-515. [doi: 10.1016/j.nurpra.2006.06.013]

21. Steinberg AG, Barnett S, Meador HE, Wiggins EA, Zazove P. Health care system accessibility. Experiences and perceptions of deaf people. J Gen Intern Med 2006 Mar;21(3):260-266 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00340.x] [Medline: 16499543]

22. Hommes RE, Borash AI, Hartwig K, DeGracia D. American sign language interpreters perceptions of barriers to healthcare communication in deaf and hard of hearing patients. J Community Health 2018 Oct;43(5):956-961. [doi: 10.1007/s10900-018-0511-3] [Medline: 29696596]

23. Dammeyer J, Lehane C, Marschark M. Use of technological aids and interpretation services among children and adults with hearing loss. Int J Audiol 2017 Oct;56(10):740-748. [doi: 10.1080/14992027.2017.1325970] [Medline: 28509597]

24. Kushalnagar P, Harris R, Paludneviciene R, Hoglind T. Health Information National Trends Survey in American Sign Language (HINTS-ASL): protocol for the cultural adaptation and linguistic validation of a national survey. JMIR Res Protoc 2017 Sep 13;6(9):e172 [FREE Full text] [doi: 10.2196/resprot.8067] [Medline: 28903891]

25. Singleton JL, Remillard ET, Mitzner TL, Rogers WA. Everyday technology use among older deaf adults. Disabil Rehabil Assist Technol 2019 May;14(4):325-332. [doi: 10.1080/17483107.2018.1447609] [Medline: 29522377]

26. Kasales CJ, Alkebsi ZA, Tong NT, Stephens AW. Caring for the deaf, hard-of-hearing, blind, and low-vision patients in the breast center. J Breast Imaging 2020;2(6):598-602. [doi: 10.1093/jbi/wbaa069]

27. De Meulder M, Haualand H. Sign language interpreting services: a quick fix for inclusion? Transl Interpreting Stud 2019 Sep 6;16(1):19-40. [doi: 10.1075/tis.18008.dem]

28. Preusse KC, Gonzalez ET, Singleton JL, Mitzner TL, Rogers WA. Understanding the needs of individuals ageing with impairment. Int J Human Factors Ergon 2016;4(2):144-168.

29. McKee MM, Moreland C, Atcherson SR, Zazove P. Hearing loss: communicating with the patient who is deaf or hard of hearing. FP Essent 2015 Jul;434:24-28. [Medline: 26161525]

30. Kwok MM, Chan RK, Hansen C, Thibault K, Wong HY. BMJ Open Qual 2021 Feb;10(1):e001062 [FREE Full text] [doi: 10.1136/bmjoq-2020-001062] [Medline: 33547156]

31. Napier J, Kidd MR. English literacy as a barrier to health care information for deaf people who use Auslan. Aust Fam Physician 2013 Dec;42(12):896-899 [FREE Full text] [Medline: 24324995]

32. Henning MA, Krägeloh CU, Sameshima S, Shepherd D, Shepherd G, Billington R. Access to New Zealand Sign Language interpreters and quality of life for the deaf: a pilot study. Disabil Rehabil 2011;33(25-26):2559-2566. [doi: 10.3109/09638288.2011.579225] [Medline: 21591984]

33. Rivas Velarde M. Indigenous perspectives of disability. Disabil Stud Q 2018 Dec 21;38(4):6114. [doi: 10.18061/dsq.v38i4.6114]

34. Chavarria MA, Schönenberger K, Mugeere A, Hurst S, Rivas Velarde M. Design approaches for creating person-centered, context sensitive, and sustainable assistive technology with the global south. In: Stein MA, Lazar J, editors. Accessible Technology and the Developing World. Oxford, UK: Oxford University Press; 2021.

35. Rivas Velarde MC. Addressing double layers of discrimination as barriers to health care: indigenous peoples with disabilities. AbOrig 2017 Oct 1;1(2):269-278. [doi: 10.5325/aboriginal.1.2.0269]

36. Stein AM, Lazar J. Accessible Technology and the Developing World. Oxford, UK: Oxford University Press; 2021.

37. Americans With Disabilities Act of 1990. Public Law 101-336 – 108th Congress. 1990. URL: https://www.govinfo.gov/content/pkg/STATUTE-104/pdf/STATUTE-104-Pg327.pdf [accessed 2022-03-03]

38. Disability Discrimination Act 2005. Act of the Parliament of the United Kingdom. 2005. URL: https://www.legislation.gov.uk/ukpga/2005/13/pdfs/ukpga_20050013_en.pdf [accessed 2005-12-31]

39. Kushalnagar RS, Tart JA. A survey on video relay service application interface preferences. J Technol Persons Disabil 2016;4:223-233.

40. Gould M, Montenegro V. 2016 CRPD ICT Accessibility Progress Report — A global analysis of the progress made by states parties to the convention on the rights of persons with disabilities to implement its dispositions on the accessibility of information and communication technologies and assistive technologies. G3ict. 2017. URL: https://g3ict.org/publication/2016-crpd-ict-accessibility-progress-report [accessed 2021-02-18]

41. The Lancet. The health of deaf people: communication breakdown. Lancet 2012 Mar 17;379(9820):977. [doi: 10.1016/S0140-6736(12)60411-5] [Medline: 22423869]

42. Laplante-Lévesque A, Hickson L, Worrall L. What makes adults with hearing impairment take up hearing AIDS or communication programs and achieve successful outcomes? Ear Hear 2012;33(1):79-93. [doi: 10.1097/AUD.0b013e31822c26dc] [Medline: 21841487]

43. McKee M, Schlehofer D, Cuculick J, Starr M, Smith S, Chin NP. Perceptions of cardiovascular health in an underserved community of deaf adults using American Sign Language. Disabil Health J 2011 Jul;4(3):192-197 [FREE Full text] [doi: 10.1016/j.dhjo.2011.04.001] [Medline: 21723526]

44. Thew D, Smith SR, Chang C, Starr M. The deaf strong hospital program: a model of diversity and inclusion training for first-year medical students. Acad Med 2012 Nov;87(11):1496-1500 [FREE Full text] [doi: 10.1097/ACM.0b013e31826d322d] [Medline: 23018327]

45. Hatakeyama T, Watanabe T, Takahashi K, Doi K, Fukuda A. Development of communication assistive technology for persons with deaf-blindness and physical limitation. Stud Health Technol Inform 2015;217:974-979. [Medline: 26294595]

46. Laubreton J, Morvan R, Roblot P. [Hospital consultations for deaf people]. Presse Med 2013 Nov;42(11):1427-1429. [doi: 10.1016/j.lpm.2013.05.002] [Medline: 23809425]

47. Barnett DD, Koul R, Coppola NM. Satisfaction with health care among people with hearing impairment: a survey of Medicare beneficiaries. Disabil Rehabil 2014;36(1):39-48. [doi: 10.3109/09638288.2013.777803] [Medline: 23594058]

48. Shahin N, Watfa M. Deaf and hard of hearing in the United Arab Emirates interacting with Alexa, an intelligent personal assistant. Technol Disabil 2020 Nov 20;32(4):255-269. [doi: 10.3233/tad-200286]

49. Cumberbatch K, Jones T. Use of Jamaican Sign language in the provision of dental health care. Community Dent Health 2017 Jun;34(2):72-76. [doi: 10.1922/CDH_3913Cumberbatch05] [Medline: 28573834]

50. Bown S, Aldersson R, Dekesel K. Supporting patients who are deaf who use a signed language in general practice. Br J Gen Pract 2020 Jan;70(690):10-11 [FREE Full text] [doi: 10.3399/bjgp20X707285] [Medline: 31879290]

Abbreviations

CRPD: Convention on the Rights of Persons with Disabilities
DHH: deaf or hard of hearing
FENASCOL: National Deaf Federation of Colombia
LMIC: low- and middle-income country
VRI: video remote interpretation

Original Paper

Evaluation of Web-Based Health Information From the Perspective of Women With Eating Disorders: Thematic Analysis

Hana Drtilova, MA; Hana Machackova, PhD; Martina Smahelova, PhD

Interdisciplinary Research Team on Internet and Society, Faculty of Social Studies, Masaryk University, Brno, Czech Republic

Corresponding Author:
Hana Drtilova, MA
Interdisciplinary Research Team on Internet and Society
Faculty of Social Studies
Masaryk University
Jostova 10
Brno, 60200
Czech Republic
Phone: 420 549 49 5870
Email: drtilova@apps.fss.muni.cz

Abstract

Background: Users with experience of eating disorders use the internet as a source of information, whether for prorecovery activities (such as web-based treatment, looking for information, support, and sharing) or activities that promote eating disorder behavior as a desirable lifestyle choice (such as pro–eating disorder communities and reading and creating pro–eating disorder posts). Their assessment of web-based eating disorder–related information is crucial for understanding the context of the illness and for health professionals and their web-based interventions.

Objective: This study aimed to understand the criteria young women with the experience of eating disorders use in evaluating eating disorder–related web-based information and what eating disorder–related characteristics of these women are involved in their evaluation.

Methods: We analyzed 30 semistructured individual interviews with Czech women aged 16 to 28 years with past or present eating disorder experience using a qualitative approach. Thematic analysis was adopted as an analytical tool.

Results: The specifics of eating disorder phases (the *disorder stage* and the *treatment process*) emerged as important aspects in the process of information assessment. Other specific characteristics of respondents (eg, motivation, abilities, and resources) addressed how the respondents arrived at certain web-based information and how they evaluated it. In addition, the respondents described some content cues as features of information (eg, novelty and social information pooling). Another finding is that other users' attitudes, experiences, activities, and personal features are involved in the information evaluation of these users and the information presented by them. Finally, the respondents evaluated the websites' visual look and graphic components.

Conclusions: This study shows that web-based information evaluation reported by women with experience of eating disorders is a complex process. The assessment is influenced by current personal characteristics related to the illness (mainly the motivation for maintaining or curing the eating disorder) using cues associated with information content, other users, and website look. The study findings have important implications for health professionals, who should ask their clients questions about web-based communities and their needs to understand what information and sources they choose.

Introduction

Background

During the past 2 decades, the internet has been integrated into our lives as an everyday tool that opens the gate to an unlimited amount of web-based information. In this content-rich environment with almost no quality control [1], the burden of information assessment shifts toward the information seeker. The internet provides user-generated content where nonspecialists offer health tips or information based on personal experiences [2], which can be highly relevant, especially in

health-related web-based searches. Wilson et al [3] revealed that among people with eating disorders (EDs) experience, 75% used the internet as a source of information, 40.8% reported visiting prorecovery sites, and 35.5% reported visiting sites that promote ED behavior. Health providers often use the internet as a tool for prevention and as an environment for ED treatment programs [4]. Consequently, for people with ED experience, web-based information can be potentially beneficial and helpful (eg, web-based treatment [5]) but also dangerous and harmful (eg, anorexia-related misinformation on YouTube [6]). For these reasons, this study aimed to deepen the knowledge of users' web-based activities with ED experience by examining their evaluation of web-based information relevant to ED topics. Specifically, the aim was to investigate the individual characteristics that shape this process and the cues used for information assessment.

EDs and Internet Use

EDs are part of the spectrum of pathological eating patterns and are perceived as either a medical illness with psychiatric features or a psychiatric illness with medical indications [7]. Regardless of the subtypes (the most known are anorexia nervosa [AN], bulimia nervosa [BN], and binge ED), EDs are burdensome in terms of significantly impaired health-related quality of patients' lives [8], not only because of the impact that the disorder can have on all body systems. The lifetime prevalence of any threshold ED among adolescents and young adults is 2.9% among women and 0.1% among men [9]. In 2020, there were 5167 patients with EDs treated in outpatient departments and hospital admissions, and the number has increased by approximately 15% in 10 years. Most patients (87%) were women and girls [10]. However, these statistics do not include people with EDs who seek help without a psychiatric context (eg, clients of nonprofit organizations) or, for instance, people who do not seek help at all. Both male and female patients with ED have high rates of psychiatric comorbidity [11]. These disorders are challenging for caregivers within the family system, as well as for health care professionals [12,13]. Moreover, EDs are egosyntonic—the person sees the disorder as part of themselves, might attribute positive valuations to ED consequences, and may perceive ED not only as an illness but also as a meaningful behavior [14]. Recovery from an ED is a long and potentially life-threatening process [15]. Despite the severity of the illness, patients with ED symptoms do not necessarily seek treatment [16].

At different levels of illness, patients have different motivations and goals for changing their ED-connected behavior. The stages of change model [17,18] proposes 6 stages of motivation for recovery. This model is particularly relevant to understanding the phases of EDs as relapses are regarded as integral parts of the change cycle, and an application to the ED context has been examined [18]. The first 2 stages are described as precontemplation (ie, nonexisting or limited intention to change behavior) and contemplation (ie, willingness to think about the change but not commit to it). These 2 stages resemble the earliest phases of ED as bounded by limited or lacking motivation to change ED-connected behavior. Following are the stages of preparation (ie, intention to change the behavior), action (ie, actively modifying the behavior), maintenance (ie,

work on the prevention of relapse), and termination (ie, zero temptation to relapse). This study presumes that people with EDs often switch between stages of illness, although the distinction between these stages might be blurred or even overlapping. Moreover, this study acknowledges that motivation plays an important role in the behavior of people with ED experience, including web-based praxes, and that motivations differ with respect to the stage of illness.

Prior research has explored the role of the internet and communication technologies in the lives of people with ED experience from various perspectives. On a general level, media are studied as one of the sociocultural risk factors that contribute to the etiology of EDs through the cultural ideals of appearance and weight [19]. Other lines of research have focused on exposure to ED-promoting websites. Such exposure is associated with a reduction in the number of calories, restriction in food consumption, greater body dissatisfaction, and a greater drive for leanness and musculature, especially among vulnerable individuals [20]. In contrast, the treatment of EDs has also moved to the web-based environment, as the anonymous internet provides a safe environment for people with EDs who may experience high levels of secrecy, shame, and stigma [21]. Internet-based treatments may offer self-assessment, self-monitoring, information about EDs, and advice for systematic treatment plans. People with ED experience also use the internet and visit web-based groups to seek information, get support, and share experiences [22].

Information Evaluation in the Web-Based Environment

Although tools for the assessment of web-based health information have been developed (eg, a study by Beaunoyer et al [23]), to the best of our knowledge, little attention has been paid to how web-based information is evaluated by users with health-related problems. This study aims to fill this gap by examining the information quality assessment by people with ED experience within their specific ED-relevant web-based activities. The construct of information quality is broad and lacks conceptual clarity. For instance, Tao et al [24] distinguished 5 dimensions of information quality on health websites: completeness, understandability, relevance, depth, and accuracy. However, other indicators of the quality of web-based information have also been identified, such as perceived aesthetics, credibility, reliability, security, consistency, usefulness, and worth [25-27]. This study defines information quality assessment as an evaluation of web-based information and materials, with a focus on both web content (including web-based information and features) and its design (the representation of the content for users) [28]. Accuracy, currency, and credibility may be among the content quality criteria, and aesthetics, cultural sensitivity, or accessibility be among the design-related criteria [29]. However, the definition of *quality* can also differ interindividually depending on the specific context and purpose of the information.

Another presumption of this study is that the web-based environment is highly diverse and can be assessed differently in terms of quality. For example, an environment (presumably) controlled by expert editors, such as websites, can be perceived

as more credible than a personal blog [30]. User characteristics, such as personal traits, abilities, previous knowledge, and topic familiarity, are also considered. The assessment process also depends on interindividual differences between users and the perceived type and context of information [31-34]. For example, the dual processing model of credibility focuses on individual factors and depicts how factors independent of message quality can affect our evaluation [35,36]. According to this model, when the user has the motivation and ability (eg, the level of literacy skills) to evaluate the quality of information, they are likely to use an analytical strategy to assess credibility systematically and rigorously. The lack of motivation inhibits users from putting effort into credibility evaluations. However, if they lack the ability and yet have motivation, they will rely more on peripheral cues (eg, the appearance of the site) and heuristics to form a judgment. Such heuristics can take the form of relying on the reputation of the source or endorsement from others [37].

User Characteristics Connected to EDs

Although research evidence about the specifics of web-based information evaluations made by people with ED experience is limited, their assessment may, in some moments, contrast that of people without ED experience. For example, their motivation for evaluation may differ according to the current state of their illness (ie, affected by the egosyntonic feature of the illness, where people with ED experience might see information congruent with their pro-ED values). Thus, to understand how people with EDs process and evaluate information, it is crucial to consider their psychological traits and cognitive characteristics. For example, a review by Cassin and von Ranson [38] revealed that both AN and BN are characterized by perfectionism, obsessive compulsiveness, neuroticism, negative emotionality, harm avoidance, low self-directedness, low cooperativeness, and traits associated with avoidant psychiatric disorders.

Moreover, some of the cognitive challenges may affect how people with the experience of EDs process web-based and offline information. Current research has demonstrated an attentional bias for disorder-salient stimuli (related to food and the body), which indicates that people with ED experience have a potential overall deficit in processing conflicting information [39]. Another cognitive deficit is a weak central coherence when attention is focused on detail, resulting in global understanding deficits [40]. Furthermore, poor set shifting (ie, a lower ability to move back and forth between tasks) results in cognitive inflexibility [41,42]. This inflexibility may manifest in rigid and concrete problem solving, reliance on strict habits and rules, and difficulties with multitasking [43].

Aims of the Study

To the best of our knowledge, no study has focused on the evaluation of web-based health information in the context of ED topics among people with an ED experience. This qualitative study intends to enhance knowledge about this topic and help us understand how women with ED experience evaluate web-based information. This study uses previous knowledge related to information evaluation. Specifically, it presumes that individual characteristics, including personality, abilities, and motivation, shape the formation of judgments about information.

Moreover, information assessment can be more or less thorough. Finally, this study considers that the web-based environment provides different cues that may be used in the process. On the basis of this knowledge, the following research questions were formulated: what ED-related characteristics of young women with the experience of EDs are involved in their evaluation of web-based information, and what criteria do young women with the experience of EDs use in evaluating web-based information?

Methods

Recruitment

The data were obtained from a research project that examined the role of new technologies among young people with ED experience in the Czech Republic. Respondents were recruited via leaflets handed out at universities in large Czech cities and in the waiting rooms of health care professionals working with people with EDs (mainly in hospitals and ED-focused nonprofit organizations in the Czech Republic). Owing to the seriousness of the illness, some respondents were available only via web-based means. Thus, the outreach was gradually expanded during the sampling process, with invitations to participate on websites relevant to EDs (both supporting ED behavior and the treatment of EDs). From the previous quantitative part of the research [44], we had a list of 307 Czech websites (including blogs and Facebook groups) that focused on healthy lifestyles (including fitness and nutrition) and professional help for EDs and promoted ED behavior (mainly pro-ED blogs and groups). These websites were found via search engines using keywords connected to a healthy lifestyle (ie, exercise, diet, and healthy eating), ED problematics (ie, professional help and informational websites), and ED promotion (ie, keywords identified in previous pro-ED research, such as *pro-ana*, *thinspo*, and *bonespo*). Finally, we posted an invitation on 15 websites (including Facebook groups) that focused mainly on ED information, recovery, and ED promotion and that ranked highest in website traffic. Research has shown that EDs are most prevalent among women [9], and risk factors are present in early adolescence, although anorexia and bulimia tend to emerge in late adolescence and early adulthood. However, the onset of EDs is individual [45]. Thus, the criterion for respondents was to be aged between 13 and 28 years and have experienced (now or in the past) a form of an ED.

Sample

The final sample comprised 30 Czech women aged 16 to 28 (mean 22.4, SD 3.9) years. Although EDs are increasing among men [46], and we actively recruited respondents of all genders, the recruitment of men was not successful. All participants claimed to experience or have experienced various EDs (AN 13/30, 43%; BN, 3/30, 10%; binge ED 1/30, 3%; or multiple ED diagnoses 13/30, 43%). Some of them had reached out for the help of health professionals and had an official diagnosis (27/30, 90%), whereas others did not (3/30, 10%). Respondents reported the presence of an ED in their current life (22/30, 73%) or that they were in full recovery (8/30, 27%). Experiences with the illness varied from 1 to 16 (mean 6.3, SD 4.5) years.

Procedure

A total of 30 semistructured interviews were conducted face to face (23/30, 77%) or via Skype web-based sessions (7/30, 23%), which lasted 41 to 118 (mean 61.0, SD 21.1) minutes. The interviews focused on the use of new technologies, including questions about the role of the internet in respondents' lives; for example, "What helps you orient in health-related online information?"; "From where do you retrieve the online information?"; and "What are the most common online activities?" The interviewers also asked about the criterion for information relevance and quality and on what cues respondents adopted the information and acted on it.

All participants were informed of the ethical aspects and purpose of the research, and they provided written informed consent. In the case of respondents aged <18 years, parents provided written consent. The interviewers had psychotherapy training and at least 2 years of psychotherapy practice as a condition for reducing potential stress among respondents.

Ethics Approval

This study was approved by the Ethical Committee of the Faculty of Social Studies of the Masaryk University, Brno, Czech Republic.

Data Analyses

The thematic analysis developed by Braun and Clarke [47] was used as the analytic method and was conducted by the first author of the study (HD). The inductive approach of analysis was used when the themes were content (data) driven and emerged from the interaction between the researcher and respondents, regardless of the specific questions. Therefore, researchers could capture the complexities of meaning within a text and understand the more tacit content.

During the analysis, we were inspired by the steps in the guidelines presented by Braun and Clarke [47] and Guest et al [48]. First, researchers became familiar with the data and text segmentation by rereading the transcripts and noting their initial ideas in a logbook. Passages related to information evaluation were segmented. In creating the initial codes, the authors entered the transcripts into the qualitative analytic software NVivo (version 10) and started to generate a codebook for codes and their labels. The code represents the specific, interesting, and essential elements of the text, and it has a greater level of abstraction than the themes [47,48]. The labels of these codes were mainly in the form of in vivo phrases used by respondents and a simplified description of the code content. These codes were discussed during research team meetings to create common categories for an initial category structure. Subsequently, continuing the analysis of the second half of the transcripts, the authors merged the codes into subthemes based on their similarities. Next, revisions were made to avoid the overlapping of the meanings. According to Chang et al [49], researchers included only subthemes with ≥3 respondents to prevent fragmentation. During the phase of defining and naming the themes, the authors reread the existing codes to better understand

their meaning and created corresponding labels and descriptions. Saturation was reached, with the occurrence of redundancy, after 30 interviews. The last 4 interviews confirmed saturation of the themes as they did not create new categories. Subsequently, the researchers checked whether the themes and subthemes were internally coherent and consistent. As some of the subthemes still overlapped, they merged some such subthemes. In addition, the themes were renamed to better correspond to their meanings. The final list of themes and subthemes is presented in the *Results* section.

The following steps were applied to ensure the validity of the results. First, all researchers followed an interview guide to standardize the data collection. Furthermore, remarks regarding the content of the interviews were discussed. Second, the first (HD) and third author (MS) consulted on emerging themes during the entire analytical process. Third, the study's third author (MS) conducted an audit comprising reading the related parts of the data and validating the final analyses. Fourth, the second author (HM) advised on the final presentation of the results to clarify the meaning of the themes. Examples supported the transparency of the interview results. Finally, the researchers applied the verbal labels of frequencies in the *Results* section to state how many respondents mentioned a particular subtheme. Instead of using the number of respondents, a verbal label (a pronoun connoting an indeterminate quantity) was attached to enhance the qualitative methodology. Inspired by the verbal counting of Sandelowski [50], the researchers operationally define, for example, *few* as something occurring among 3 to 8 respondents (see the *Results* section).

Results

Overview

A total of 4 themes and 10 subthemes were identified, as summarized in Table 1. The theme *respondent characteristics* represents respondents' characteristics that influenced their information evaluation. Themes of *content cues*, *characteristics of other users*, and *website cues* present the respondents' cues mentioned in their evaluation. The results cover the findings relevant to the entire information assessment process. It encompasses the initial phases of information seeking, including factors that affect preferences for diverse sources and further evaluation of the found information.

The specifics of the EDs phases emerged as important aspects of respondents' characteristics. Some respondents spontaneously categorized themselves as being in the *disorder stage* or in the *treatment process* during the interviews. Respondents who described themselves as being in the *disorder stage* had statements and descriptions that fit in the precontemplation and contemplation stages of the stages of change model [17], whereas those seeing themselves as being in the *treatment process* described more processes connected to the preparation, action, maintenance, and termination stages of the model. These dimensions were emphasized for each theme and subtheme, as shown in Table 1.

Table 1. The final list of themes and subthemes and their occurrence in ED[a] phase[b].

Theme and subtheme	ED phase (disorder stage and treatment process)	Frequency label
Respondent characteristics		
Motivation	Both	Most
Abilities and resources	Both	Few
Congruence between personal experience and information	Both	Some
Content cues		
Verification	Both	Few
Novelty	Only in the disorder stage	Few
Social information pooling	More in the disorder stage	Some
Characteristics of other users		
Source expertise	Both	Most
Similarity to respondent	Both	Most
Website cues		
Reputable look	Unclear	Few
Photographs of people relevant to ED	More in the disorder stage	Some

[a]ED: eating disorder.

[b]Frequency labels in Table 1 and further in the text describe how many respondents mentioned each subtheme. *Few* indicates 3 to 8 respondents, *some* indicates 9 to 17 respondents, *most* indicates 18 to 29 respondents, and *all* indicates 30 respondents.

Respondent Characteristics

Overview

This theme captures how the characteristics of respondents, including their web-based behavior patterns, are involved in information evaluation. Specifically, it addresses how respondents arrived at certain web-based information and how they evaluated it. The following subthemes emerged: *motivation, abilities and resources,* and *congruence between personal experience and information.*

Motivation

Most respondents described 2 main motivations relevant to their illness: maintaining their disorder in the disorder stage or getting cured during the treatment process. The particular type of motivation affected what information the respondent chose, as one respondent revealed the following:

> If someone, I think, has the motivation, that she can cure herself, then she is able to filter on the internet to what she wants to read, what she doesn't want to read, and what she lets influence her and what not. [R13]

For respondents in the disorder stage, the vision of a skinny body was so strong that they looked up and accumulated as much information as possible to maintain their goals. However, they were not concerned with its evaluation. They read and eventually "tried everything" (R2). One of the respondents piled up information instead of analyzing it, and although the information was labeled as nonsense, they read and used it anyway. They also talked about their passive role in choosing information when their disorder decided what was needed. A respondent experienced the following:

> I know what is right, what is really right, I know it. But of course, anorexia chooses what she likes, not how it's supposed to be like. So I believe more or less in almost everything. [R29]

In contrast, respondents in the treatment process did not read or seek proana and promia information as they were afraid of being pulled back into the disorder stage. However, they described their persistent sensitivity to disordered relevant information, such as diet commercials. Nevertheless, they used different information checks, as a respondent pointed out the following:

> When the person is in an acute stage, then she blindly follows what she wants to gather. Now I follow what I want to gather, that I want to be healthy. But in that acute phase, I don't think about if it will hurt me. And now I want to verify all information with someone responsible, with a professional who tells me "it is appropriate for you, it is not appropriate for you." [R6]

Abilities and Resources

For a few respondents, their current situation in terms of abilities and resources affected their information selection. For instance, respondents at the disorder stage chose exercise and weight reduction tips based on their physical state. Financials played a role as well, leading to choosing diet tips and menus suited to their monetary situation. One of the respondents in the disorder stage also mentioned that the criterion for information selection was the time spent applying particular advice to her life.

Information choice was also influenced by the information-seeking strategies and skills of the respondents.

Some actively used search engines and keywords, such as *anorexia*, *bulimia*, and *eating disorders* or questions such as "How to do away with bulimia" (R2) or "How to throw up" (R4). They then mainly clicked on the first link. Active searches also included clicking on links on blogs and forums that led to similar websites.

For others, disorder-related information had appeared unwelcomely when information was "jumping out" (R10) at them against their will. Respondents discussed how the information was "attacking them" (R27) and how it was almost impossible to avoid. For respondents in the treatment process, it was proana or promia blogs or information about eating. For respondents in the disorder stage, it was information about treatment or dieting.

Congruence Between Personal Experience and Information

The knowledge and experience gathered through their disorder helped some to assess the relevance of information about EDs. The information that corresponded to personal experiences was credible and relevant. For example, respondents in the disorder stage said that they did not trust diets as they already knew what weight loss strategies were good for them:

> *I was always using my experience. So, for example, when I knew that by that [caloric] intake I had lost weight, or by eating that food I had lost weight, then I simply ate it, because I had tested it. And I did not trust anything else.* [R8]

Respondents in the treatment process viewed healthy lifestyle information as appropriate because of their experience with professionals and with new nutrition information.

Content Cues

Overview

According to the respondents' evaluation of information, content cues described the qualities (in the sense of features) of the information found on websites. The following subthemes emerged within this theme: *verification*, *novelty*, and *social information pooling*.

Verification

Information was approved by a few respondents when it was consistent across websites.

However, some respondents needed to verify the web-based information in the offline environment by comparing the information with books or, as mentioned by respondents in the treatment process, via consultation with professionals.

Novelty

Another cue for some respondents was whether the information was new. This subtheme was mentioned more by respondents in the disorder stage:

> *How should I behave towards the food, how to hide the food, what should I avoid, how to deal with various situations, what to do and what not to do. Any new information was good information for me.* [R27]

Consequently, other users' long-term sharing of new posts was considered beneficial rather than the sharing of a few posts once in a while: information posted "every day or every other day was more credible" (R25).

Social Information Pooling

The subtheme most cited by some respondents was how others' experiences helped them judge information. This subtheme was mentioned more by respondents in the disorder stage and represented sharing opinions, recommendations, and feedback on desirable topics between respondents and other users (eg, comments below articles, reactions on forums, and liking some posts via social network sites). Respondents saw others' recommendations as helpful and worthy, although they had never met them on the web or offline:

> *It is weird that I took their advice a lot. One doesn't know the other person at all, but still follows what is written there.* [R7]

However, other respondents saw recommendations as valuable if they knew the users from the web-based environment (eg, following the blog of a friend).

In this process, the number of reactions was also significant. Testimonials, positive responses, and the number of *thumbs-ups* increased the chance of seeing information as trustful, whereas their absence had the opposite effect:

> *I read the comments if [pro-ana advice] works or not. But if there was no comment, then I did not trust it. Or I wouldn't try it unless there was something written there, some opinion.* [R7]

For respondents in the disorder stage, if some information worked for others (ie, was used and acted upon), it was good:

> *But when she reads "try this and that,"' then she says to herself "when others do it, it must be really cool."* [R4]

For respondents in the treatment process, social pooling was essential for the assessment of the treatment procedures and practice of helping professionals.

Characteristics of Other Web-Based Users

Overview

This theme captures how other users' attitudes, experiences, activities, and personal features are involved in the information evaluation of these users and the information presented by them. Other users were mostly seen as post contributors or members of web-based communities. Their characteristics were expressed in the subthemes of *source expertise* and *similarity to the respondent*.

Source Expertise

The perception of who is an expert in an ED field served as a hint for information assessment. Respondents distinguished whether the information was provided by users who are currently experiencing or had experienced EDs or by ED specialists. The specialists were mentioned without further explanation or specified as professionals, such as psychologists, physicians,

and nutrition specialists, and were connected to institutions, such as universities, hospitals, and ED treatment centers.

The most mentioned aspect for most respondents was whether users had experience with EDs. Those who did were seen as experts in ED problems and credible as they better understood the respondents' issues and feelings. Respondents who preferred this expertise sometimes set experience with EDs as a cornerstone:

> Certainly, there were some important basic factors. That person had to have an eating disorder, or at least not be OK with food, such as people with obesity. [R24]

Specifically, one of the respondents in the disorder stage said that she ignored notes by professionals on a self-help website and read only the text passages written by people with EDs.

In both stages, the mentoring and labeling of respondents by professionals discouraged respondents from looking up and reading the information presented by these professionals. Professionals could not be trustworthy for a respondent in the disorder stage as they wanted her to gain weight and, therefore, did not provide complete information. In contrast, the entire community of users with ED experience was considered trustworthy. The goodwill of this community was described as being open, welcoming, and accepting, creating the feeling of an alliance.

The valuable characteristics of ED specialists on the internet were described simply as being professional, providing verified information to respondents, understanding EDs, having healthy opinions about EDs, and being selflessly helping respondents. Those who saw professionals as credible (although some questioned their expertise) perceived the ED information written by experts and treatment groups supervised by them as trustworthy.

For some respondents, the view of the expertise cue had changed over time. Those who started treatment looked more or only for information written and shared by professionals. Respondents uncertain about their willingness to be cured could have ambivalent feelings, as one respondent pointed out the following:

> If I listened to the doctor, I would eat well. I know she gives me the right advice, that she is a professional. However, unfortunately, I adopt what suits me more. Rather from the Internet. [R29]

Similarity to Respondent

Most respondents wanted to know many details about other users to better evaluate whether they were similar and, consequently, whether they were trustworthy, which in some respondents led to trying to find as many details as possible.

Respondents specifically mentioned similarity cues connected to their illness, including the same type of ED, a similar stage of the disorder, similar problems, and similar attitudes toward food. The cues not strictly linked to ED were described as similarities in current mood, humor, age, and writing style.

Website Cues

Overview

This theme reveals a representation of the visual look and graphic components of the website. Two subthemes were identified: *reputable look* and *photographs of people relevant to ED*.

Reputable Look

Respondents named a reputable look as a good sign for further selection. This look was not defined by the specific features of a website but rather by broad general characteristics, such as clarity, lack of bias, and good organization.

Photographs of People Relevant to ED

Pictures, videos, advertisement photos of diet products, and especially photographs of people with experience of EDs were important for respondents, mostly in the disorder stage. Before and after images helped respondents assess the actual effect of others' aims and provided tips to lose or gain weight. Moreover, the photographs acted as proof that the people in them were real and not lying. Specifically, respondents in the disorder stage believed in what was presented in the photographs:

> I saw a picture of a woman with a gorgeous figure, so I believed that she is on that diet [from a commercial]. [R23]

A few respondents mentioned that they no longer took advertisement information for granted when they shifted to the treatment stage of ED.

Discussion

Principal Findings

The purpose of this study was to explore how young women with ED evaluate web-based information and how the specifics of their illness contribute to this evaluative process. Respondents mentioned several cues for information assessment within the websites' content, the characteristics of other users, and the website characteristics while stressing the influence of the current phase of their illness.

First, respondents' characteristics played an integral role in the evaluation, intervening in the entire process from exposure to information to final judgment. As Hargittai et al [32] suggested, obtaining information and its evaluation are more often handled as 2 separate research interests. However, our respondents holistically depicted these 2 steps. They explained how their *abilities* and *resources* affected this process and, importantly, what role *motivation* played in grounding their illness. *Motivation* turned out to be a vital part of the assessment, confirming the justified emphasis within the dual processing model of credibility by Metzger [36] and the stage of change model [17]. For respondents in the disorder stage, the motivation to lose weight was their biggest goal. They were not concerned about information evaluation and saw every piece of information as automatically good. This strategy could be a consequence of their *congruence between personal experience and information* (ie, EDs): the more experienced the respondents were, the more certain they were about the accuracy of their information

selection. Another explanation for the automatic assessment of information may be cognitive rigidity, which is more pronounced in people with ED experience (eg, see the study by Tchanturia and Hambrook [43]). As a result, when a particular type of information is already assessed as quality, additional information from a similar topic can also be considered appropriate without further evaluation. Although such quick evaluation strategies are convenient, they may generate bias. A study by Guardiola-Wanden-Berghe et al [51] suggested that the information quality of websites on dieting and EDs was poor. Viewing them as automatically believable may lead to the risk of adopting and behaving on harmful information.

The social element attached to quality cues was also prevalent throughout most themes. The *characteristics of others* was the most frequently mentioned theme, although web-based sources and other users are often masked or missing for evaluation [35]. However, reliance on social features is part of the evaluative process, as shown in other studies [37]. Similarly, *social information pooling* was a social-connected cue whereby other users' opinions and feedback were hints for information selection and evaluation. In particular, for respondents in the disorder stage, the information confirmed and approved by other users was seen as trustworthy. Weight reduction was the primary goal for respondents in the ED stage; therefore, they looked for fast and effective weight reduction information. However, this search could be limited in 2 aspects: as the subtheme *abilities and resources* revealed, they were tired and probably less inclined to put much energy into it; in addition, some of them did not perceive health professionals as credible. Thus, the opinions and reviews of others served as quick and less effortful tools for information evaluation. The more those testimonials were presented, the more chance the information was viewed as valid. These assumptions are similar to the bandwagon heuristic of Sundar [52], whereby people suppose that if many others think something is correct, then it must be. However, heuristics may lead to problems with crowd behavior, especially for young people with EDs, who are at greater risk of peer pressure [53]. Moreover, quality may be falsely connected to popularity when unpopular topics and information are discarded [2]. In addition, popular users may be perceived as falsely trustworthy. Thus, when and from whom people with EDs seek support and advice might be crucial for understanding their information evaluation and selection.

User ED expertise was another socially relevant cue in which information from others who had experience with EDs was perceived as appropriate. Boero and Pascoe [54] suggested that other users in pro-ED communities use their experiences to demonstrate their authenticity and offer advice about, for example, how to get through recovery programs without actual recovery. They also answered the questions of new members using their nutritional and medical knowledge. The respondents' positive perception of the information presented by these users may be partly understood by the Situated Identity Enactment Model of Cruwys et al [55], highlighting the role of socially bonded norms, identity, and context. Being a member of a social group may moderate conformity with the group's social norms. For example, these norms may be to follow a lifestyle presented by the group or to trust more experienced members. Being a member of an ED-experienced group is also vital in the treatment stage because of the transformation of illness identity to treatment identity [56]. If a member fails to follow the norms, they may be rejected [57]. Thus, one of the possible group norms may be to see ED-experienced group members as reliable, and respondents comforted by this norm seek to avoid group rejection. Studies suggest that communities of ED-experienced users represent a safe place for sharing experiences and a means of support and understanding [58]. The respondents in this study indeed saw their ED-related web-based communities as full of goodwill and, thus, were trustworthy.

The similar stage of an ED or the same ED type was another marker for user quality, reflecting studies showing that people seek and see information that matches what they already know [59,60]. Accordingly, respondents stated that users who were similar to respondents and supported respondents' opinions might have similar opinions and are plausible. Metzger and Flanagin [2] described a self-confirmation heuristic whereby credible information confirms pre-existing beliefs. In the case of an ED, this heuristic may enhance the egosyntonic feature of the illness, whereby respondents value their ED and see the information as good if it supports its maintenance. This confirmation of established opinions may place a burden on treatment, especially at the disorder stage. In the disorder stage, respondents ignored health professional advice and followed prodisorder tips, which enabled them to pursue an ED lifestyle. In contrast, similar experiences of recovery and treatment strategies may be supportive for those who decide to treat themselves. Consequently, health professionals should be aware of the identification needs of their clients and how their selectiveness may influence what is considered credible.

Finally, respondents also mentioned cues related to websites (specifically the reputable look and photographs associated with EDs); however, mentions of such elements were relatively scarce. This could suggest that, within the self-report context of the testimonies, respondents consciously reflected that website cues were not as significant as to be named and instead focused on social aspects in the overall evaluation.

Limitations of the Study

This study had several limitations. First, the sample was relatively homogenous as it comprised only women, although they differed in age, region, and illness experience.

Thus, the testimonies might not represent other demographics, such as men or different ethnicities. Next, the respondents reported the condition of the ED diagnosis, and not all of them had experience with professional health care. It is possible that although respondents claimed to have an ED themselves, they would not meet the diagnostic criteria of psychological diagnostic manuals, such as the Diagnostic and Statistical Manual of Mental Disorders. Similarly, experience with technology was not measured and might have varied for different illnesses and ages. In addition, the 2 stages of EDs (disorder vs treatment process) derived from the respondents were often blurred, unclear, sometimes overlapping, or even missing. Although some processes of these stages might resemble the stage of change model [17], respondents were not further asked for their motivation to change. Finally, respondents' browsing

history was self-reported, and researchers did not retrieve the browsing history data. Thus, respondents' internet use and information were unique to each and based and customized on their previous web-based activities, which can vary substantially among individuals. For future studies, we suggest using more accurate measures to use the level of change and consider different contexts and settings for respondents' information evaluation.

Conclusions and Implications

This study showed that the web-based information evaluation reported by women with the experience of EDs is a complex process. The assessment is influenced by current personal characteristics related to the illness (mainly the motivation to maintain or cure the ED) using cues associated with information content, other users, and website look. The study findings have important implications for health professionals, who should ask their clients questions about web-based communities and their needs to understand what information and sources they choose.

They should support clients by consulting their judgments and uncertainties about information evaluation. Further investigation into the role of consulting web-based information with professionals in information assessment by users might benefit future therapeutic practices. Who and what is positive regarding quality and how it changes within the ED phases may help to understand the illness. Moreover, quality cues may serve as merits for designing an ideal website for health providers who use the internet for prevention and intervention. Future research might benefit from the experimental design of these websites and their evaluation by people with ED experience. For example, the personal stories of people experiencing EDs, their tips about treatment, and web-based peer groups on websites may increase the perceived quality of information and providers. However, the perceived quality of information might not be associated with the direct application and use of this information in an offline environment. Thus, the dynamism of the transfer between perceptions of information quality and acting on this information might be another research interest for the future.

Conflicts of Interest

None declared.

References

1. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Annual Rev Inf Sci Technol 2008 Oct 24;41(1):307-364. [doi: 10.1002/aris.2007.1440410114]

2. Metzger MJ, Flanagin AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmatics 2013 Dec;59:210-220. [doi: 10.1016/j.pragma.2013.07.012]

3. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):e1635-e1643. [doi: 10.1542/peds.2006-1133] [Medline: 17142493]

4. Schlegl S, Bürger C, Schmidt L, Herbst N, Voderholzer U. The potential of technology-based psychological interventions for anorexia and bulimia nervosa: a systematic review and recommendations for future research. J Med Internet Res 2015 Mar 31;17(3):e85 [FREE Full text] [doi: 10.2196/jmir.3554] [Medline: 25840591]

5. ter Huurne ED, de Haan HA, Postel MG, van der Palen J, VanDerNagel JE, DeJong CA. Web-based cognitive behavioral therapy for female patients with eating disorders: randomized controlled trial. J Med Internet Res 2015 Jun 18;17(6):e152 [FREE Full text] [doi: 10.2196/jmir.3946] [Medline: 26088580]

6. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]

7. Maine M, McGilley BH, Bunnell DW. Treatment of Eating Disorders Bridging the Research - Practice Gap Book, 1st edition. Amsterdam, The Netherlands: Elsevier; 2010.

8. Ágh T, Kovács G, Supina D, Pawaskar M, Herman BK, Vokó Z, et al. A systematic review of the health-related quality of life and economic burdens of anorexia nervosa, bulimia nervosa, and binge eating disorder. Eat Weight Disord 2016 Sep;21(3):353-364 [FREE Full text] [doi: 10.1007/s40519-016-0264-x] [Medline: 26942768]

9. Nagl M, Jacobi C, Paul M, Beesdo-Baum K, Höfler M, Lieb R, et al. Prevalence, incidence, and natural course of anorexia and bulimia nervosa among adolescents and young adults. Eur Child Adolesc Psychiatry 2016 Aug;25(8):903-918. [doi: 10.1007/s00787-015-0808-z] [Medline: 26754944]

10. The number of people with eating disorders is increasing in the Czech Republic. NewsBeezer. URL: https://czechpoints.com/eating-disorders-are-on-the-rise-in-the-czech-republic/ [accessed 2022-04-20]

11. Ulfvebrand S, Birgegård A, Norring C, Högdahl L, von Hausswolff-Juhlin Y. Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. Psychiatry Res 2015 Dec 15;230(2):294-299. [doi: 10.1016/j.psychres.2015.09.008] [Medline: 26416590]

12. Seah XY, Tham XC, Kamaruzaman NR, Yobas PK. Knowledge, attitudes and challenges of healthcare professionals managing people with eating disorders: a literature review. Arch Psychiatr Nurs 2017 Feb;31(1):125-136. [doi: 10.1016/j.apnu.2016.09.002] [Medline: 28104050]

13. Stefanini MC, Troiani MR, Caselli M, Dirindelli P, Lucarelli S, Caini S, et al. Living with someone with an eating disorder: factors affecting the caregivers' burden. Eat Weight Disord 2019 Dec;24(6):1209-1214. [doi: 10.1007/s40519-018-0480-7] [Medline: 29368292]

14. Clinical presentation of anorexia nervosabulimia nervosa. In: Eating Disorders and Obesity, Second Edition A Comprehensive Handbook. New York, USA: The Guilford Press; 2002.

15. Rome E, Ammerman S, Rosen D, Keller R, Lock J, Mammel K, et al. Children and adolescents with eating disorders: the state of the art. Pediatrics 2003 Jan;111(1):e98-108. [doi: 10.1542/peds.111.1.e98] [Medline: 12509603]

16. Hart LM, Granillo MT, Jorm AF, Paxton SJ. Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. Clin Psychol Rev 2011 Jul;31(5):727-735. [doi: 10.1016/j.cpr.2011.03.004] [Medline: 21501580]

17. DiClemente C, Prochaska J. Toward a comprehensive model of change. In: Treating Addictive Behaviors. Boston, MA, USA: Springer; 1986.

18. Hasler G, Delsignore A, Milos G, Buddeberg C, Schnyder U. Application of Prochaska's transtheoretical model of change to patients with eating disorders. J Psychosom Res 2004 Jul;57(1):67-72. [doi: 10.1016/S0022-3999(03)00562-2] [Medline: 15256297]

19. Keel PK, Forney KJ. Psychosocial risk factors for eating disorders. Int J Eat Disord 2013 Jul;46(5):433-439. [doi: 10.1002/eat.22094] [Medline: 23658086]

20. Rodgers RF, Lowy AS, Halperin DM, Franko DL. A meta-analysis examining the influence of pro-eating disorder websites on body image and eating pathology. Eur Eat Disord Rev 2016 Jan;24(1):3-8. [doi: 10.1002/erv.2390] [Medline: 26230192]

21. Murphy R, Frost S, Webster P, Schmidt U. An evaluation of web-based information. Int J Eat Disord 2004 Mar;35(2):145-154. [doi: 10.1002/eat.10245] [Medline: 14994351]

22. McCormack A. Individuals with eating disorders and the use of online support groups as a form of social support. Comput Inform Nurs 2010;28(1):12-19. [doi: 10.1097/NCN.0b013e3181c04b06] [Medline: 19940616]

23. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: evaluation, tools, and strategies. Patient Educ Couns 2017 Feb;100(2):183-189. [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]

24. Tao D, LeRouge C, Smith KJ, De Leo G. Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. JMIR Hum Factors 2017 Oct 06;4(4):e25 [FREE Full text] [doi: 10.2196/humanfactors.6455] [Medline: 28986336]

25. Kim H, Park SY, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011 Sep;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]

26. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]

27. Knight S, Burn J. Developing a framework for assessing information quality on the world wide web. Informing Sci J 2005;8:159-172. [doi: 10.28945/493]

28. Huizingh E. The content and design of web sites: an empirical study. Inf Manag 2000 Apr;37(3):123-134 [FREE Full text] [doi: 10.1016/s0378-7206(99)00044-0]

29. 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]

30. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Commun Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]

31. Fogg BJ. Prominence-interpretation theory: explaining how people assess credibility online. In: Proceedings of the Extended Abstracts on Human Factors in Computing Systems. 2003 Presented at: CHI03: Human Factors in Computing Systems; Apr 5 - 10, 2003; Ft. Lauderdale Florida USA. [doi: 10.1145/765891.765951]

32. Hargittai E, Fullerton L, Menchen-Trevino E, Thomas KY. Trust online: young adults' evaluation of web content. Int J Commun 2010;4(1):468-494 [FREE Full text]

33. Hilligoss B, Rieh SY. Developing a unifying framework of credibility assessment: construct, heuristics, and interaction in context. Inf Process Manag 2008 Jul;44(4):1467-1484. [doi: 10.1016/j.ipm.2007.10.001]

34. Rafalak M, Abramczuk K, Wierzbicki A. Incredible: is (almost) all web content trustworthy? analysis of psychological factors related to website credibility evaluation. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14: 23rd International World Wide Web Conference; Apr 7-11, 2014; Seoul, Korea. [doi: 10.1145/2567948.2578997]

35. Digital Media, Youth, and Credibility. Cambridge, MA, USA: MIT Press; 2008.

36. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Nov;58(13):2078-2091. [doi: 10.1002/asi.20672]

37. Metzger MJ, Flanagin AJ, Medders RB. Social and heuristic approaches to credibility evaluation online. J Commun 2010 Aug 19;60(3):413-439. [doi: 10.1111/j.1460-2466.2010.01488.x]

38. Cassin SE, von Ranson KM. Personality and eating disorders: a decade in review. Clin Psychol Rev 2005 Nov;25(7):895-916. [doi: 10.1016/j.cpr.2005.04.012] [Medline: 16099563]

39. Aspen V, Darcy AM, Lock J. A review of attention biases in women with eating disorders. Cogn Emot 2013;27(5):820-838 [FREE Full text] [doi: 10.1080/02699931.2012.749777] [Medline: 23228135]

40. Lang K, Roberts M, Harrison A, Lopez C, Goddard E, Khondoker M, et al. Central coherence in eating disorders: a synthesis of studies using the Rey Osterrieth complex figure test. PLoS One 2016;11(11):e0165467 [FREE Full text] [doi: 10.1371/journal.pone.0165467] [Medline: 27806073]

41. Wu M, Brockmeyer T, Hartmann M, Skunde M, Herzog W, Friederich H. Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: a systematic review and meta-analysis. Psychol Med 2014 Dec;44(16):3365-3385. [doi: 10.1017/S0033291714000294] [Medline: 25066267]

42. Kanakam N, Raoult C, Collier D, Treasure J. Set shifting and central coherence as neurocognitive endophenotypes in eating disorders: a preliminary investigation in twins. World J Biol Psychiatry 2013 Aug;14(6):464-475. [doi: 10.3109/15622975.2012.665478] [Medline: 22630167]

43. Tchanturia K, Hambrook D. Cognitive remediation therapy for anorexia nervosa. In: The Treatment of Eating Disorders: A Clinical Handbook. New York, USA: The Guilford Press; 2011.

44. Almenara CA, Machackova H, Smahel D. Sociodemographic, attitudinal, and behavioral correlates of using nutrition, weight loss, and fitness websites: an online survey. J Med Internet Res 2019 Apr 04;21(4):e10189 [FREE Full text] [doi: 10.2196/10189] [Medline: 30946018]

45. Rohde P, Stice E, Marti CN. Development and predictive effects of eating disorder risk factors during adolescence: implications for prevention efforts. Int J Eat Disord 2015 Mar 06;48(2):187-198 [FREE Full text] [doi: 10.1002/eat.22270] [Medline: 24599841]

46. Mitchison D, Hay P, Slewa-Younan S, Mond J. The changing demographic profile of eating disorder behaviors in the community. BMC Public Health 2014 Sep 11;14:943 [FREE Full text] [doi: 10.1186/1471-2458-14-943] [Medline: 25213544]

47. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

48. Guest G, MacQueen KM, Namey E. Applied Thematic Analysis. Los Angeles, CA, USA: SAGE Publications, Inc; 2014.

49. Chang Y, Voils CI, Sandelowski M, Hasselblad V, Crandell JL. Transforming verbal counts in reports of qualitative descriptive studies into numbers. West J Nurs Res 2009 Nov;31(7):837-852 [FREE Full text] [doi: 10.1177/0193945909334434] [Medline: 19448052]

50. Sandelowski M. Real qualitative researchers do not count: the use of numbers in qualitative research. Res Nurs Health 2001 Jun;24(3):230-240. [doi: 10.1002/nur.1025] [Medline: 11526621]

51. Guardiola-Wanden-Berghe R, Gil-Pérez JD, Sanz-Valero J, Wanden-Berghe C. Evaluating the quality of websites relating to diet and eating disorders. Health Info Libr J 2011 Dec;28(4):294-301 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00961.x] [Medline: 22051128]

52. Sundar SS. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Digital Media, Youth, and Credibility. Cambridge, MA, USA: The MIT Press; 2008.

53. Ferguson CJ, Muñoz ME, Garza A, Galindo M. Concurrent and prospective analyses of peer, television and social media influences on body dissatisfaction, eating disorder symptoms and life satisfaction in adolescent girls. J Youth Adolesc 2014 Jan;43(1):1-14. [doi: 10.1007/s10964-012-9898-9] [Medline: 23344652]

54. Boero N, Pascoe C. Pro-anorexia communities and online interaction: bringing the pro-ANA body online. Body Soc 2012 May 24;18(2):27-57. [doi: 10.1177/1357034X12440827]

55. Cruwys T, Platow M, Rieger E, Byrne DG, Haslam SA. The social psychology of disordered eating: the situated identity enactment model. Eur Rev Soc Psychol 2016 Oct 24;27(1):160-195. [doi: 10.1080/10463283.2016.1229891]

56. McNamara N, Parsons H. 'Everyone here wants everyone else to get better': the role of social identity in eating disorder recovery. Br J Soc Psychol 2016 Dec;55(4):662-680. [doi: 10.1111/bjso.12161] [Medline: 27667140]

57. Giles D. Constructing identities in cyberspace: the case of eating disorders. Br J Soc Psychol 2006 Sep;45(Pt 3):463-477. [doi: 10.1348/014466605X53596] [Medline: 16984715]

58. Yeshua-Katz D, Martins N. Communicating stigma: the pro-ana paradox. Health Commun 2013;28(5):499-508. [doi: 10.1080/10410236.2012.699889] [Medline: 22873763]

59. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]

60. Senkowski V, Branscum P. How college students search the internet for weight control and weight management information: an observational study. Am J Health Educ 2015 Jul 06;46(4):231-240. [doi: 10.1080/19325037.2015.1044139]

Abbreviations

AN: anorexia nervosa
BN: bulimia nervosa
ED: eating disorder

XSL•FO
RenderX

Original Paper

Machine Learning–Based Text Analysis to Predict Severely Injured Patients in Emergency Medical Dispatch: Model Development and Validation

Kuan-Chen Chin[1*], MD; Yu-Chia Cheng[2*], BS; Jen-Tang Sun[3], MD, MS; Chih-Yen Ou[2], BS; Chun-Hua Hu[4], EMTP; Ming-Chi Tsai[4], EMTP; Matthew Huei-Ming Ma[5,6], MD, PhD; Wen-Chu Chiang[5,6*], MD, PhD; Albert Y Chen[2*], PhD

[1]Department of Emergency Medicine, Taipei Hospital, Ministry of Health and Welfare, New Taipei City, Taiwan

[2]Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan

[3]Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan

[4]Emergency Medical Service Division, Taipei City Fire Department, Taipei City, Taiwan

[5]Department of Emergency Medicine, National Taiwan University Hospital, Taipei City, Taiwan

[6]Department of Emergency Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yunlin County, Taiwan

[*]these authors contributed equally

Corresponding Author:
Albert Y Chen, PhD
Department of Civil Engineering
National Taiwan University
No 1, Section 4, Roosevelt Rd
Taipei City, 106
Taiwan
Phone: 886 2 3366 4255
Email: AlbertChen@ntu.edu.tw

Abstract

Background: Early recognition of severely injured patients in prehospital settings is of paramount importance for timely treatment and transportation of patients to further treatment facilities. The dispatching accuracy has seldom been addressed in previous studies.

Objective: In this study, we aimed to build a machine learning–based model through text mining of emergency calls for the automated identification of severely injured patients after a road accident.

Methods: Audio recordings of road accidents in Taipei City, Taiwan, in 2018 were obtained and randomly sampled. Data on call transfers or non-Mandarin speeches were excluded. To predict cases of severe trauma identified on-site by emergency medical technicians, all included cases were evaluated by both humans (6 dispatchers) and a machine learning model, that is, a prehospital-activated major trauma (PAMT) model. The PAMT model was developed using term frequency–inverse document frequency, rule-based classification, and a Bernoulli naïve Bayes classifier. Repeated random subsampling cross-validation was applied to evaluate the robustness of the model. The prediction performance of dispatchers and the PAMT model, in severe cases, was compared. Performance was indicated by sensitivity, specificity, positive predictive value, negative predictive value, and accuracy.

Results: Although the mean sensitivity and negative predictive value obtained by the PAMT model were higher than those of dispatchers, they obtained higher mean specificity, positive predictive value, and accuracy. The mean accuracy of the PAMT model, from certainty level 0 (lowest certainty) to level 6 (highest certainty), was higher except for levels 5 and 6. The overall performances of the dispatchers and the PAMT model were similar; however, the PAMT model had higher accuracy in cases where the dispatchers were less certain of their judgments.

Conclusions: A machine learning–based model, called the PAMT model, was developed to predict severe road accident trauma. The results of our study suggest that the accuracy of the PAMT model is not superior to that of the participating dispatchers; however, it may assist dispatchers when they lack confidence while making a judgment.

XSL•FO
RenderX

Introduction

Background

Trauma is a leading cause of accidental death globally. According to the World Health Organization, injuries contribute to >5 million deaths each year. Road traffic accidents accounted for most injuries and were the ninth leading cause of death in 2012 [1]. Severe trauma is a time-sensitive emergency condition. Prompt transport is beneficial for patients with neurotrauma and penetrating injuries with unstable hemodynamic features [2]. Delays in transportation are associated with poor functional outcome [3].

Prehospital triage allows severely ill patients to receive appropriate time-sensitive management. For cardiac arrest and stroke victims, dispatchers can obtain critical information on the phone, such as the patient's level of consciousness, breath patterns, or prehospital stroke scales [4,5]. However, no standardized questions have been designed for dispatchers when they encounter severe trauma. Only a few studies on helicopter emergency medical services have addressed the accuracy of dispatch for trauma victims [6]. Current trauma scales for predicting severity require either physiological or anatomical assessments [7]. Therefore, a victim's condition cannot be identified or evaluated until the first batch of emergency medical technicians (EMTs) arrives at the scene.

Motivation

Content analysis has been conducted on emergency calls to discover the factors that affect dispatch and have the potential to assist prehospital triage [8,9]. Specifically, text classification has demonstrated the effectiveness of classifying events recorded during phone calls [10]. In addition, natural language processing has been used in emergency medicine. Text mining techniques have been used to predict the triage level, length of stay, disposition, and mortality in emergency department patients [11-16]. A textual analysis–based machine learning framework was developed to assist dispatchers during the prehospital phase in out-of-hospital cardiac arrest (OHCA) recognition; this framework has been commercialized [17-20]. These techniques make it possible to stratify the risk to patients when structured questions are unavailable, similar to the assessment of trauma patients over the phone.

The classic process of text classification includes text preprocessing, feature extraction, and classifier construction. Text preprocessing aims to remove noise and effectively retrieve information through text cleaning and organization [21]. Common feature extraction approaches can be loosely divided into two domains: word frequency and semantics [22,23]. Machine and deep learning models, such as k-nearest neighbors, decision trees, support vector machines, multilayer perceptron classifiers, and naïve Bayes, are widely used as classifiers [24-28].

Aim

We hypothesized that severe trauma cases could be recognized based on the content of communication between callers and call takers during emergency calls. The main research question and objective of this study was to develop a machine learning–based model through text mining of emergency calls to automatically identify severely injured patients in road accidents. We focused on road accidents instead of all trauma cases because they are the major cause of trauma, and compared with other types of injuries, the content of emergency calls for road accidents is homogeneous. As there are no suitable previous studies for comparison, our second objective was to compare the results of the model with 6 participating dispatchers' judgment.

Methods

Study Design and Setting

This paper describes a cross-sectional study on identifying severely injured patients in road accidents by analyzing Mandarin text of emergency calls using machine learning. The results were compared with those of human judgment. We defined severely injured patients as those who fit the major trauma criteria of the EMT trauma triage protocol, that is, prehospital-activated major trauma (PAMT).

Data Acquisition

Data were obtained from the Taipei Trauma Registry, which is a database of trauma accident information from 8 out of 18 hospitals with first aid capabilities. A random sample of one-fourth of the total cases considered as PAMT in 2018 was retrieved. After excluding cases without complete information, 92 PAMT patients (92 of 377 registered cases) were enrolled. As control cases, 3 consecutive non-PAMT road accident calls were matched with each PAMT on the same day from the dispatch system. If the number of non-PAMT cases to be matched on a given day was insufficient, only 1 or 2 calls were included. A total of 92 PAMT calls and 255 non-PAMT calls were considered in this study. The exclusion criteria were as follows: the caller was not by the side of the victim, the caller did not speak Mandarin, the accident was not vehicle-related, and the calls did not provide sufficient information. The final data for analysis included 114 cases in total, which comprised 42 PAMT and 72 non-PAMT cases (Figure 1).

**Figure 1.** Data acquisition and study design. PAMT: prehospital-activated major trauma.
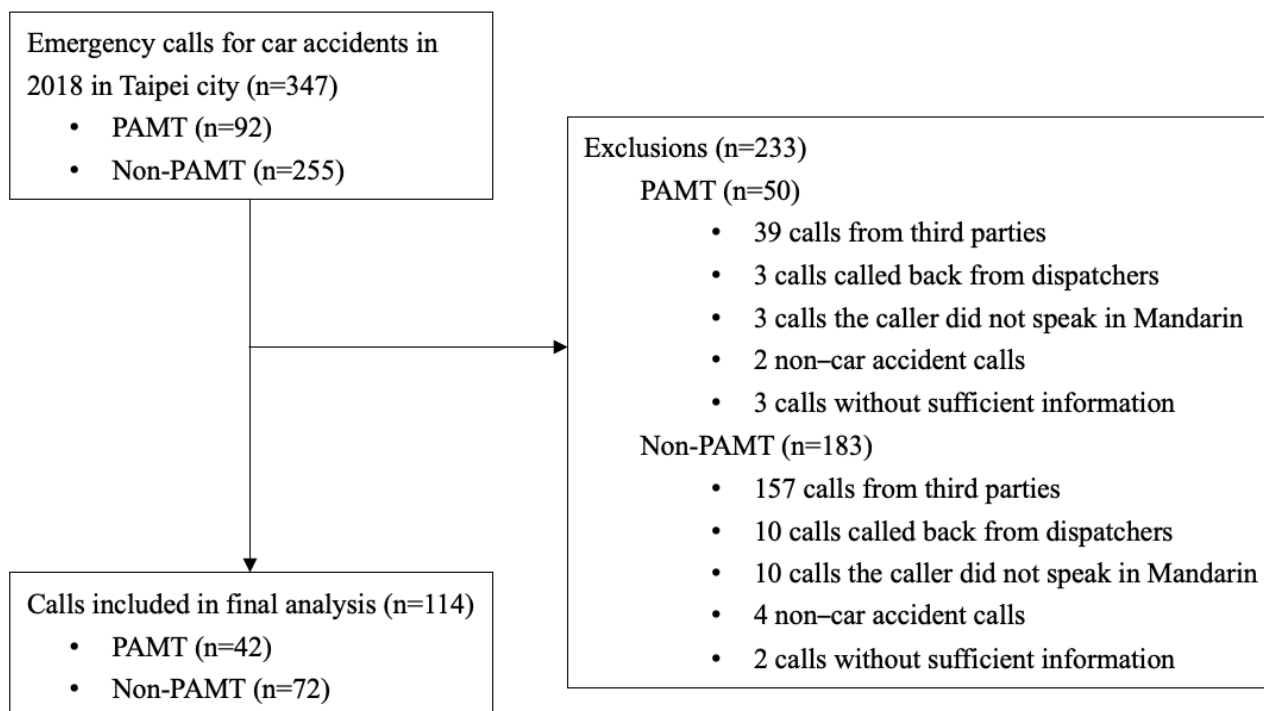
Ethics Approval

This study was approved by the institutional review board of the National Taiwan University Hospital (case number 201902043RINB).

Model Development

As shown in Figure 2, formal model development comprises four steps: (1) text preprocessing, (2) feature engineering, (3) model classification, and (4) model enhancement, which was conducted to improve model performance.

Figure 2. Model development. PAMT: prehospital activated major trauma; TF-IDF: term frequency–inverse document frequency; KNN: k-nearest neighbors; SVM: support vector machine; MNB: multinomial naïve Bayes; BNB: Bernoulli naïve Bayes; MLP: multilayer perceptron; SENS: sensitivity; SPEC: specificity; PPV: positive predictive value; NPV: negative predictive value; ACC: accuracy. Repeated random subsampling-cross validation (RRS-CV) for 100 times were performed in the step of model enhancement. All training data include 39 PAMT and 65 non-PAMT cases; testing data included 3 PAMT and 7 non-PAMT cases.
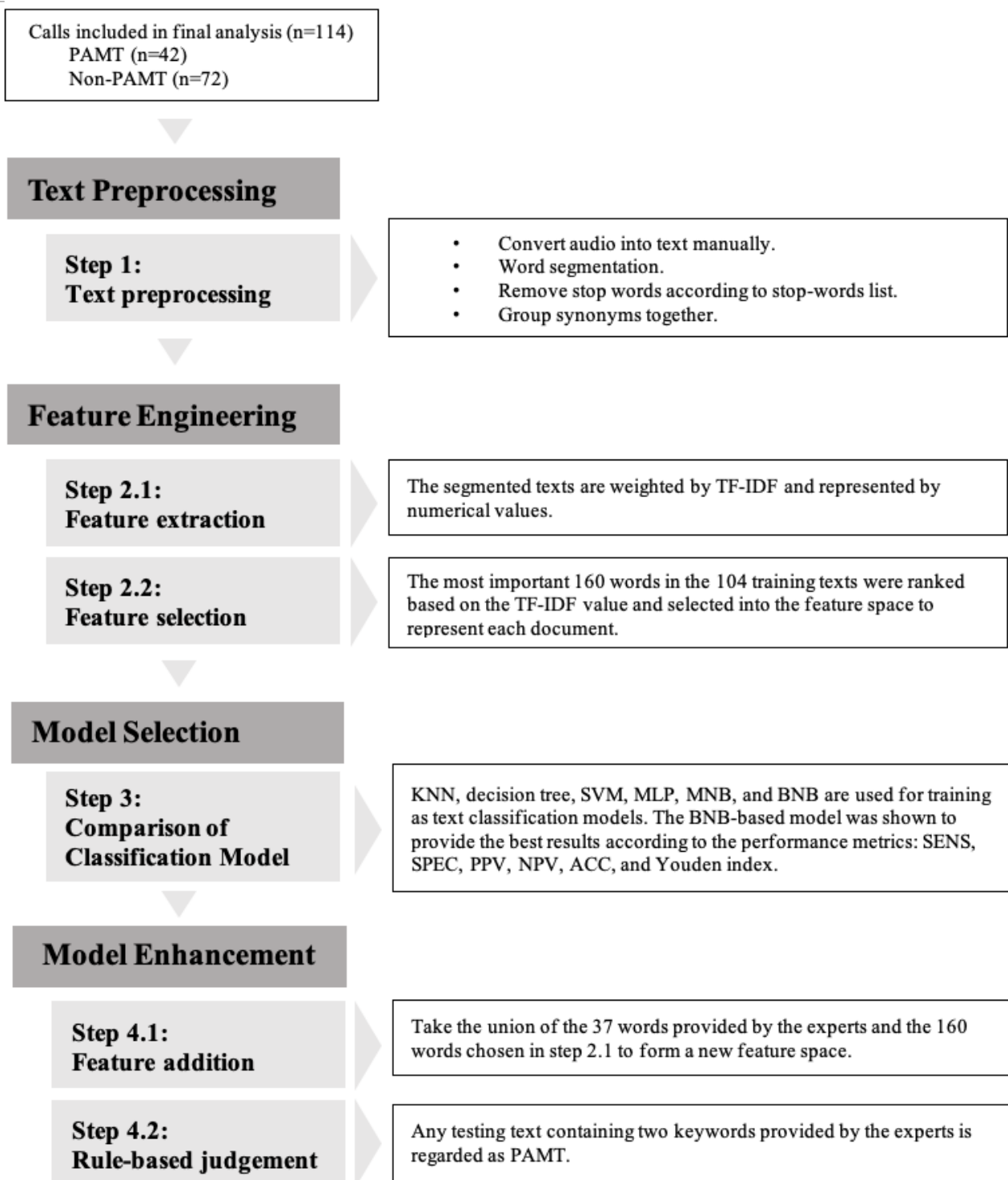
Text Preprocessing (Step 1)

The purpose of text preprocessing is to organize the data such that useful information can be retrieved. This process includes word segmentation, stop word removal, and synonym grouping (Figure 2). First, each emergency call was manually converted into a text form. The continuous text string was then segmented into words, which were the shortest units of meaning, consisting of at least one character. Segmentation was performed using the Chinese word segmentation system developed by the Institute of Information Science and the Institute of Linguistics of Academia Sinica [29,30]. To eliminate segmentation errors caused by ambiguous Chinese compound words, a dictionary of special terms with specific weights was manually constructed based on experience and trial and error. The segmentation

system refers to the weight required to force certain words to merge or separate. Subsequently, stop words were removed to remove insignificant words, such as conjunctions, pronouns, and articles. Then, synonyms were grouped and regarded as the same word, potentially reducing the model overfitting to specific words, thus providing a means for bias-variance control. From >27,000 characters in the original 114 texts, approximately 7000 different word meanings were identified.

Feature Engineering (Step 2)

In feature engineering, the segmented words were transformed into a machine-readable format by feature extraction (step 2.1, Figure 2). As emergency calls are often short, and conversations are urgent, important words are frequently mentioned (Multimedia Appendix 1). Thus, we used term frequency–inverse document frequency (TF-IDF) to weigh each word. The TF-IDF calculation consists of two sections: TF and IDF (Multimedia Appendix 2). TF illustrates the word frequency, whereas IDF explains the rarity of words appearing in the entire document. A higher frequency of occurrence of a word in one specific text indicates its importance. In contrast, a higher frequency of occurrence of the word in the entire body of texts lowers its importance. By considering these 2 frequencies simultaneously, we ranked all words by importance to conduct feature selection (step 2.2, Figure 2). The most important 160 out of 7000 words were chosen based on the experiments. The selected features were placed in a feature space to reduce the number of dimensions and to make the results more explanatory. The feature space included the selected features used to develop the model.

Model Selection (Step 3)

For model selection, we evaluated several commonly used machine learning models for text classification, including k-nearest neighbors, decision tree, support vector machine, multilayer perceptron, multinomial naïve Bayes, and Bernoulli naïve Bayes (BNB). Repeated random subsampling cross-validation (RRS-CV) was conducted 100 times to avoid overfitting and to obtain more stable and reliable classification results. RRS-CV splits samples in a randomized and repeated manner without replacement. The performance of the different models used for comparison was the average of 100 RRS-CV scores. According to Table 1, among these, the BNB-based model achieved the best results. The BNB classifier, which is a supervised learning model, is based on Bayes' theorem. It assumes that each input variable is independent of the other variables. According to the BNB equation in Multimedia Appendix 2, the calculation concentrates on binary information of whether the word appears in a document. The Boolean expression of the selected features forms the feature vector for each document. The category estimation of a document depends on the maximum a posteriori of each class k, which consists of the likelihood of the document being given by class k and its prior probability. The category with the highest maximum a posteriori labeled the classified documents. To avoid a zero-probability situation, Laplace smoothing was used to set the additive smoothing parameter to one. Consequently, no hyperparameter tuning was required for BNB. Compared with other text classification models, the BNB model has the advantages of simplicity, efficient computational speed, and ability to achieve a high level of accuracy without hyperparameter tuning. Furthermore, this model is suitable for processing small-scale data and short texts [31,32]. The results and hyperparameter tuning of other models are presented in Multimedia Appendix 3.

Table 1. Comparison of machine learning models.

Model	SENS[a] (%)	SPEC[b] (%)	PPV[c] (%)	NPV[d] (%)	ACC[e] (%)	Youden index
KNN[f]	18.7	89.0	32.6	72.1	67.9	0.077
Decision tree	32.7	76.0	35.9	72.9	63.0	0.087
SVM[g]	55.7	74.0	49.3	80.3	68.5	0.297
MNB[h]	19.0	96.1	42.2	73.8	73.0	0.151
BNB[i]	*53.0* [j]	*86.7* [j]	*67.0* [j]	*81.6* [j]	*76.6* [j]	*0.397* [j]
MLP[k]	53.7	79.0	55.6	80.6	71.4	0.327

[a]SENS: sensitivity.

[b]SPEC: specificity.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]ACC: accuracy.

[f]KNN: k-nearest neighbors.

[g]SVM: support vector machine.

[h]MNB: multinomial naïve Bayes.

[i]BNB: Bernoulli naïve Bayes.

[j]BNB-based model achieved the best ACC and Youden index.

[k]MLP: multilayer perceptron.

For the split of training and validation data, we set a fixed ratio of PAMT to non-PAMT cases in the validation data. As shown in Figure 3, when the amount of training data becomes larger than that of the validation data, the training score gradually decreases and the validation score increases. The 2 lines were closest when the training and validation data sizes were 104 and 10, respectively. The convergence illustrates that, at this number of training samples, adding more training data does not significantly improve the classification performance. Therefore, for all text classification models, 104 texts were randomly selected as training data and the remaining 10 texts were used

as validation data (Figure 2). The training data included 39 PAMT and 65 non-PAMT cases, and the validation data included 3 PAMT and 7 non-PAMT cases. The ground truth of model classification is the on-scene judgment of the EMT, which is presented in the form of binary labels. Figure 4 shows the scalability of the BNB-based model. As the training data increased, the model-fitting time fluctuated moderately around 0.002 seconds but significantly increased when the training data size was >104. In addition, 104 training data points with 10 validation data points had the highest validation score and the third shortest model-fitting time (Figure 5).
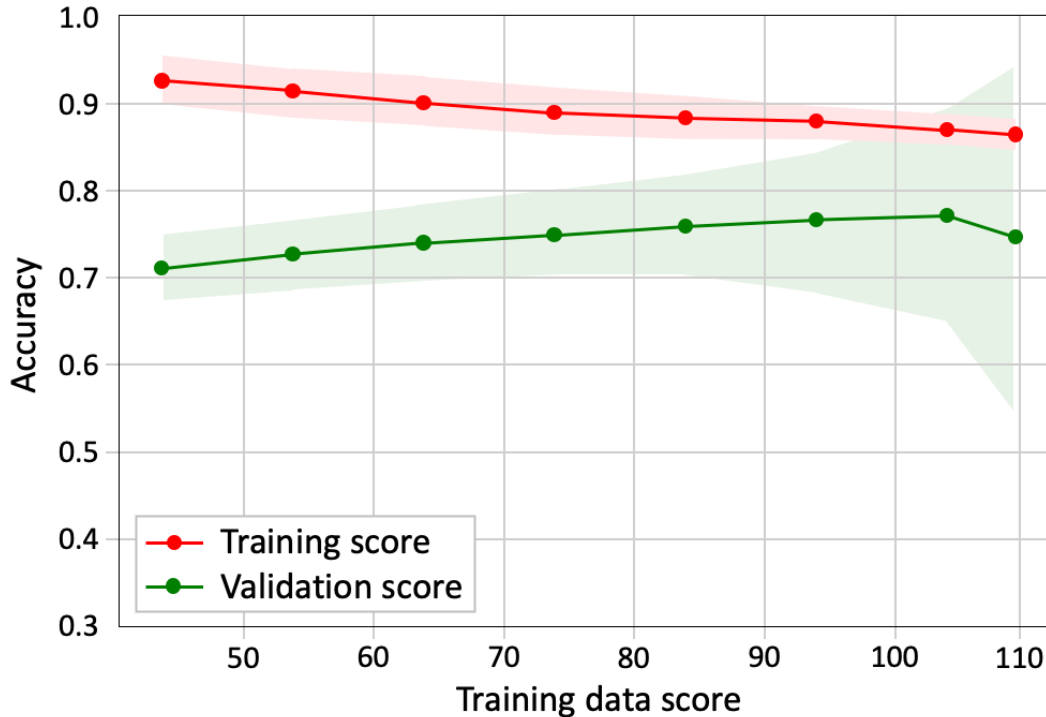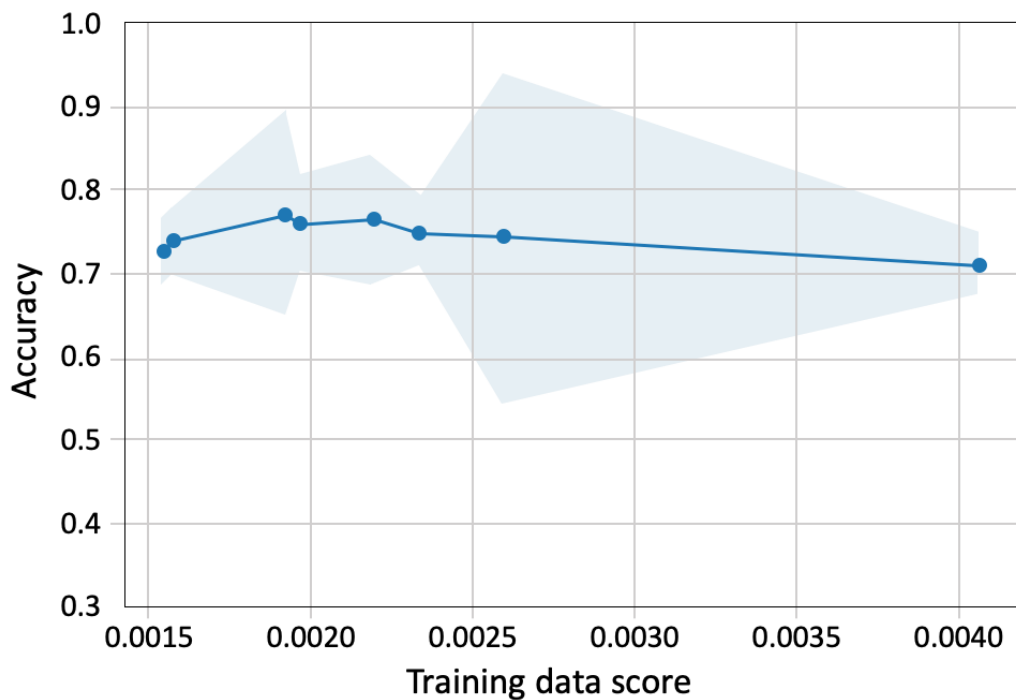
Figure 3. Learning curve of the Bernoulli naïve Bayes.

Figure 4. Scalability of the Bernoulli naïve Bayes.

Figure 5. Performance of the Bernoulli naïve Bayes.

Model Enhancement (Step 4)

To optimize the final performance of our model, we enhanced the BNB-based models using feature addition (step 4.1) and rule-based judgment (step 4.2). In feature addition, we gathered 37 keywords provided by the experts and combined them with the 160 words chosen in step 2.2 to form a new feature space (Figure 2). The experts included 6 participating dispatchers and 2 emergency physicians. After they had listened to the 114 audio recordings, they were asked, "Which keyword in an emergency

call indicates whether a patient is a PAMT or non-PAMT patient?" They then provided keywords based on their personal experience. The 37 keywords were expected to expand the important feature set, which may be limited by the small amount of data. The feature space created by the union of 160 and 37 words was used to develop enhanced models. Although important features must be included, their contribution to the classification may be small if their frequencies are not significant. Therefore, a rule-based judgment (step 4.2) was designed to highlight the importance of the 37 suggested

keywords. Specifically, any text used in the validation that contained at least 2 of the 37 words provided by the experts was classified as PAMT. Texts that did not fit this rule were further examined by a BNB classifier (Figure 2).

The enhanced BNB-based model was compared with various derivative models based on combinations of different steps. The 4 derivative versions of the BNB-based model are presented in Table 2. Model A comprised manually selected features and rule-based judgment. Model B was a classical text classification model that included TF-IDF feature extraction and selection with BNB classification. Model C comprised feature engineering steps and manual feature addition with BNB classification. Finally, we named the best version as the PAMT model. It comprises steps 1 to 4.2, including text preprocessing, feature engineering, model classification, and both model enhancement approaches.

Table 2. BNB-based models of different combinations of steps.

Model	Performance						Steps included[a]				BNB[b] classification
	SENS[c] (%)	SPEC[d] (%)	PPV[e] (%)	NPV[f] (%)	ACC[g] (%)	Youden index	1	2	4.1	4.2	
Model A	54.7	82.1	56.8	80.9	73.9	0.368	✓		✓	✓	
Model B	53.0	86.7	67.0	81.6	76.6	0.397	✓	✓			✓
Model C	54.0	87.3	67.8	82.1	77.3	0.413	✓	✓	✓		✓
PAMT[h] model	68.0	78.0	60.6	85.8	75.0	0.460	✓	✓	✓	✓	✓

[a]Step 1, text preprocessing; step 2, term frequency–inverse document frequency feature extraction and selection; step 4.1, manual feature addition; step 4.2, rule-based judgment.

[b]BNB: Bernoulli naïve Bayes.

[c]SENS: sensitivity.

[d]SPEC: specificity.

[e]PPV: positive predictive value.

[f]NPV: negative predictive value.

[g]ACC: accuracy.

[h]PAMT: prehospital-activated major trauma.

Human Participants

For a reference comparison with the PAMT model, we conducted a survey to collect severe trauma judgments from 6 volunteer dispatchers. They were from the fire departments of Taipei City and New Taipei City (Table 3). The participants were asked to listen to 114 road accident audio clips. As we focused on text analysis, the participants were not allowed to receive any information other than the text. Therefore, the audio clips were transcribed into a computer-synthesized voice using a text-to-speech tool. The audio clips were played randomly in both female and male voices. In this way, the tone, speed, and emotions of the speech were neutralized. While listening to the clips, each participant classified the cases as PAMT or non-PAMT depending on their personal experience and intuition. They also shared information regarding their certainty (certain or uncertain) in each case.

Table 3. Profiles of the participating dispatchers.

Participant	Sex	Age (years), range	Service city	EMT[a] experience (year)	Dispatch experience (year)
A	Male	30-39	New Taipei City	13	6
B	Female	40-49	New Taipei City	10	2
C	Male	30-39	New Taipei City	14	1
D	Male	30-39	New Taipei City	10	1
E	Male	30-39	Taipei City	10	4
F	Male	30-39	Taipei City	9	4

[a]EMT: emergency medicine technician.

Data Analysis

The analysis determined the accuracy, positive predictive value, negative predictive value, sensitivity, and specificity of the PAMT model prediction and average judgments of the participants [33,34].

Accuracy refers to the proportion of correctly predicted PAMT and non-PAMT cases. The proportion of cases with true-predicted PAMT and non-PAMT results can be presented as positive predictive value and negative predictive value, respectively. sensitivity and specificity represent the ability of a classification system to correctly identify PAMT and

non-PAMT cases, respectively. The Youden index was calculated using different models and can be expressed as the sum of sensitivity and specificity minus 1.

All 114 cases were categorized into certainty levels from 0 to 6, depending on how many participants regarded a case as *certain*. For example, a case with certainty level 4 indicated that 4 participants were certain of their judgment, whereas the other two were not. The accuracy was also calculated for different certainty levels.

Data management and statistical analyses were performed using Python (Python Software Foundation) and Excel (Microsoft Corporation).

Results

Sample

In total, 114 patients were included in the final analysis. The transcribed texts ranged from 84 to 652 characters, with a mean of 241.4 (SD 106.7) characters; the mean character count of PAMT cases was greater than that of non-PAMT cases (266, SD 102 vs 227, SD 107). The transcribed computer-synthesized audio ranged from 24 to 145 seconds in length, with a mean of 58.9 (SD 24.5) seconds, and the mean call length of PAMT cases was longer than that of non-PAMT (64, SD 24 vs 54, SD 24 seconds) cases (Multimedia Appendix 4).

Outcome Data

In this study, the machine learning model was trained on a random sample of 104 cases and validated on the remaining 10 cases. RRS-CV was conducted 100 times to obtain greater unbiased validation results; moreover, no external data were

used to test the performance of the trained models. According to Table 1, BNB outperformed the other models because it had the highest overall metrics: accuracy (76.6%) and Youden index (0.397). The mean sensitivity, specificity, positive predictive value, and negative predictive value for BNB were 53.0%, 86.7%, 67.0%, and 81.6%, respectively. As there was still room for improvement, model enhancement was performed based on BNB to increase the performance. The enhanced BNB-based model, known as the PAMT model, exhibited the best performance. Its Youden index was 0.460, and it achieved a mean sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of 68.0%, 78.0%, 60.6%, 85.8%, and 75.0%, respectively (Table 2). The performance of model C, which was only enhanced by adding the features provided by the 6 volunteer dispatchers, was ranked after the PAMT model. The mean sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and Youden index of model C were 54.0%, 87.3%, 67.8%, 82.1%, 77.3%, and 0.413, respectively. Model A contained only the features provided by the experts and was classified based on rule-based judgment. It achieved the worst results (sensitivity 54.7%; specificity 82.1%; positive predictive value 56.8%; negative predictive value 80.9%; accuracy 73.9%; Youden index 0.368).

In contrast, the mean sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of the 6 participants were 63.1%, 85.0%, 71.7%, 80.3%, and 76.8%, respectively (Multimedia Appendix 5). The PAMT model with the best performance had a higher sensitivity and negative predictive value but a lower specificity, positive predictive value, and accuracy than the participants. Overall, the PAMT model did not surpass the performance of the participating dispatchers (Figure 6).
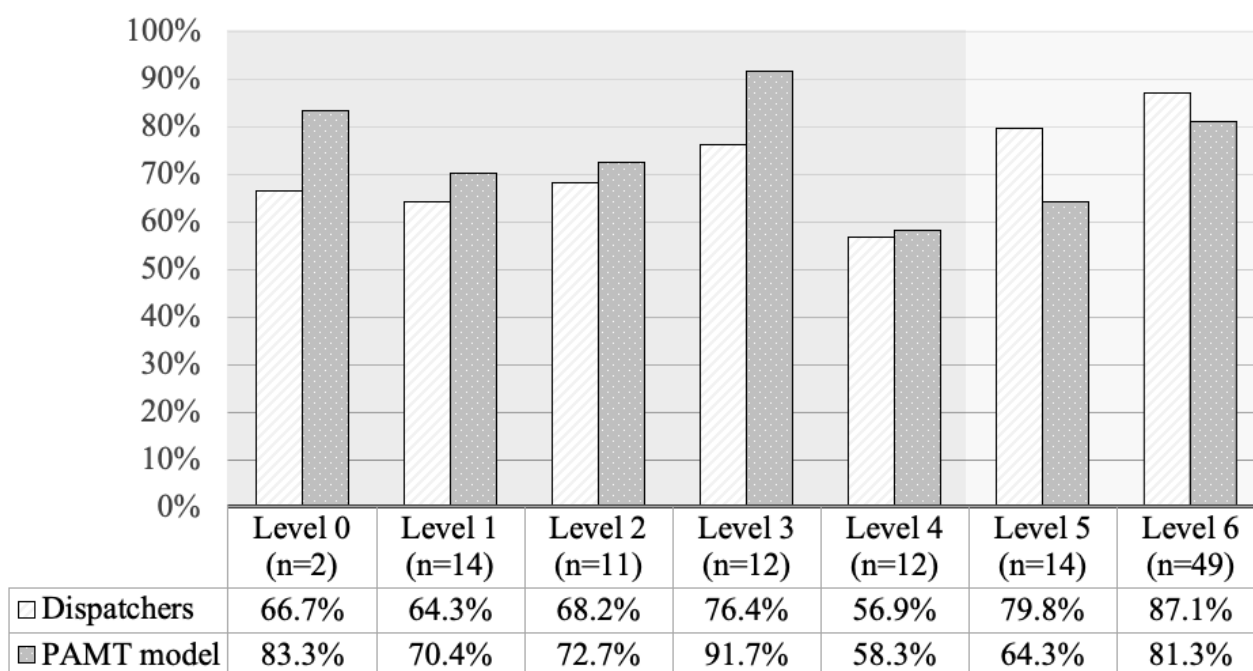
Figure 6. Overall performance of participating dispatchers versus prehospital-activated major trauma (PAMT) model. ACC: accuracy; NPV: negative predictive value; PAMT: prehospital activated major trauma; PPV: positive predictive value; SENS: sensitivity; SPEC: specificity.

	SENS	SPEC	PPV	NPV	ACC
□ Dispatchers (n=114)	63.1%	85.0%	71.7%	80.3%	76.8%
▨ PAMT model (n=114)	68.0%	78.0%	60.6%	85.8%	75.0%
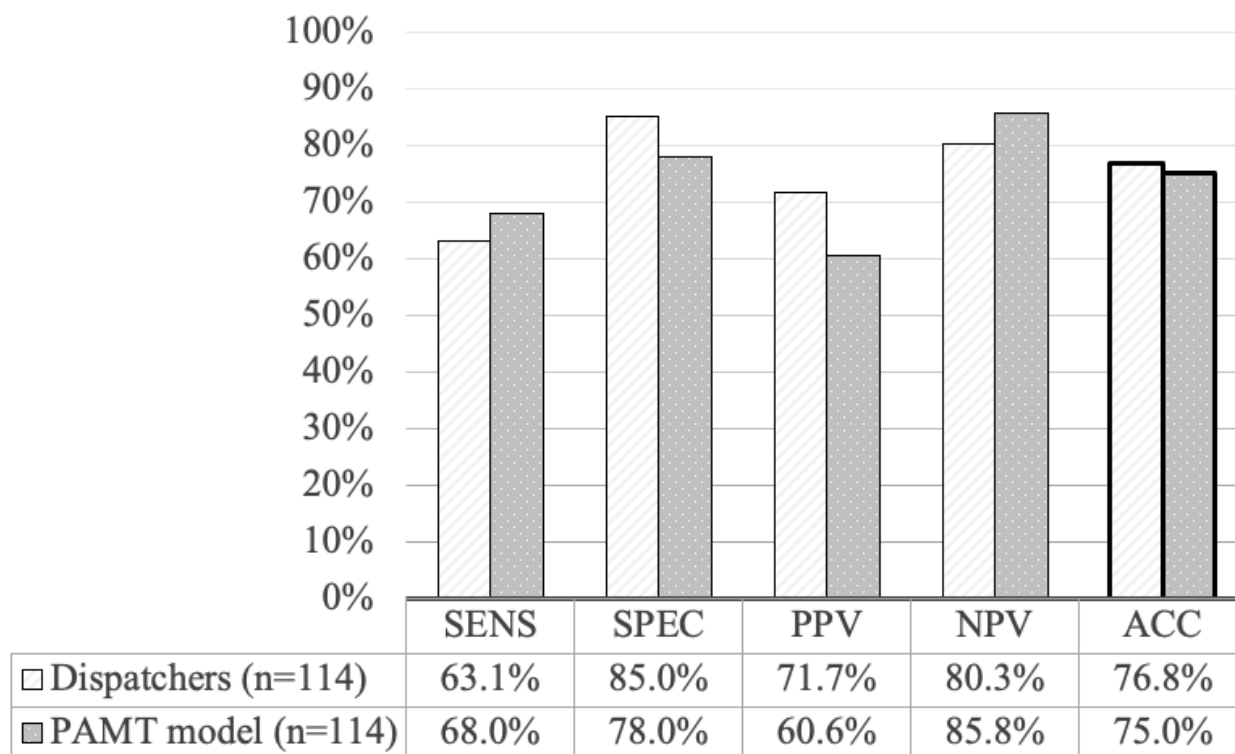
In the subgroup analysis, as shown in Figure 7, the mean accuracy of the participants at certainty levels from 0 to 6 was 66.7%, 64.3%, 68.2%, 76.4%, 56.9%, 79.8%, and 87.1%. The mean accuracy of the PAMT model at certainty levels from 0 to 6 was 83.3%, 70.4%, 72.7%, 91.7%, 58.3%, 64.3%, and 81.3%. After all cases were categorized based on different certainty levels, the accuracy of the participants for levels 0 to 6 generally increased, except for level 4, whereas the accuracy of the PAMT model did not show such a linear pattern. The results of the PAMT model did not display a clear trend; that is, they were affected by the certainty level because the BNB model classified cases according to the feature distribution. If we define levels 5 and 6 as *certain cases* and levels 0 to 4 as *uncertain cases*, we can observe that, although the accuracy of the PAMT model was lower than that of the participants in *certain* cases (77.52% vs 85.48%), it was greater than the accuracy of the participants in *uncertain* cases (73.57% vs 66.34%; Figure 7).

Figure 7. Accuracy of predicting prehospital-activated major trauma (PAMT) by participating dispatchers and PAMT model over different certainty levels.

	Level 0 (n=2)	Level 1 (n=14)	Level 2 (n=11)	Level 3 (n=12)	Level 4 (n=12)	Level 5 (n=14)	Level 6 (n=49)
☐ Dispatchers	66.7%	64.3%	68.2%	76.4%	56.9%	79.8%	87.1%
▦ PAMT model	83.3%	70.4%	72.7%	91.7%	58.3%	64.3%	81.3%

Discussion

Principal Findings

Our study makes 3 major contributions to the field. First, this is the primary study to use a machine learning–based model to identify severely injured patients during the dispatch phase. Second, the overall performance of the model was similar to that of human dispatchers (Figure 6). Third, the model produced favorable results for cases in which dispatchers were uncertain (Figure 7).

With no suitable previous studies as a reference, we enrolled 6 volunteer dispatchers in our study. Their judgment was regarded as a reference for comparison with the models. Although such a small sample size cannot represent all dispatchers, we were still able to observe heterogeneity in human performance. As shown in Multimedia Appendix 5, three participants (A, B, and E) had a high specificity and low sensitivity, whereas the other three (C, D, and F) had more balanced figures between specificity and sensitivity. We can speculate that different experiences may affect judgment, and that the policy each participant chose, either aggressive or conservative, also made a difference. With the assistance of the proposed model, which is more stable and adjustable, it is possible to narrow the range of human discrepancies and decrease the uncertainty.

The proposed machine learning models are text classification models. As important words were repeatedly mentioned in often short and intermittent emergency calls (Multimedia Appendix 1), the frequency-based feature extraction method, TF-IDF, demonstrated the ability to select representative words in severe trauma calls. In addition, feature correlation analysis was performed for these words (Multimedia Appendix 6). Features with high correlation coefficients were words that frequently appeared together in Mandarin, or in the question required to

be asked during a call. In contrast, low correlation words indicated that they appeared independently. Despite varying degrees of correlation, all the selected features are meaningful and have the potential to be keywords for judging PAMT. Therefore, the occurrence of these words was the main input for the machine learning models. Furthermore, we analyzed the length of the texts and the accuracy between the PAMT and participants. As each text was represented by a feature vector formed by word occurrences, the original length may be one of the factors affecting accuracy. Multimedia Appendix 7 presents further results.

To explore why machine learning performed better in classifying uncertain cases, we compared the words suggested by experts and the words selected by the model. Of the 37 words provided by the experts, 23 were regarded as keywords specifically for PAMT. In Multimedia Appendix 8, we compare these 23 words and the top 23 decisive words selected by the model that were most likely to occur in the PAMT texts. In the left column, most words are aggregated in the "Patient status" and "Patient basic information" categories; few are in the "Geographic information" and "Auxiliary words and other information" categories. In contrast, the words in the right column are grouped not only in the "Patient status" category but also in "Geographic information" and "Auxiliary words and other information." This phenomenon shows that the participants focused more on the situations and injury mechanisms of the patients, whereas the proposed model was able to capture other information such as the location of an accident or wording in a conversation. In uncertain cases, there may be fewer obvious keywords for PAMT, which is possibly why the proposed model is more helpful.

In addition to the PAMT model, we tried different feature combinations and classification approaches to develop three other models. Models A, B, and C refer to manual feature

addition with rule-based judgment, TF-IDF feature engineering with BNB classification, and TF-IDF feature engineering plus manual feature addition with BNB classification, respectively (Table 2). The PAMT model consisted of steps 2.1 to 4.2. It is important to consider sensitivity and specificity while developing a triage tool; therefore, we chose the model with the highest Youden index as our final model, which was the PAMT model. The sensitivity of the PAMT model was also the highest, making it suitable for use as a triage tool.

Our results demonstrate that it is necessary to combine machine learning (steps 2.1 and 2.2) and human experience (steps 4.1 and 4.2) to develop a prehospital dispatching triage tool (Table 2). A purely manual model using the features provided by experts with rule-based judgment, such as model A, or a classical machine learning–based text classification model, such as model B, did not perform sufficiently well. Although the features of model C are composed of the TF-IDF selection and are provided by experts, without rule-based judgment to increase the importance of these keywords, it failed to outperform the PAMT model. Rule-based judgment makes the added feature of experts suggesting words more significant in classification, which is a complement of limited data. Although the best classification performance of the BNB model indicates that the occurrence of words in a call is key to identifying PAMT cases, there is currently no machine learning model that can completely replace human dispatchers.

Comparison With Prior Work

Abundant research has been conducted on field triage and prognosis prediction using prehospital data for the early recognition of severely injured trauma patients. However, the dispatching accuracy has seldom been addressed in previous studies [35,36]. The predictors used in these studies, either physiological data or injury mechanisms, were difficult to acquire through telephone calls. In the few studies regarding the accuracy of dispatching, most dealt with helicopter emergency medical services dispatching [6,37-40]. In another study, all trauma emergency calls were included and compared between clinicians and nonclinicians in a prehospital critical care team in Scotland [41]. The sensitivity of the two groups, in the study, for identifying major trauma (injury severity score>15) were 0.112 and 0.259 and the specificity was 0.998 and 0.995. Our model had a significantly higher sensitivity and lower specificity. However, the results varied as the gold standards differed. We chose the judgment of the on-scene EMT as the gold standard as it represents comprehensive prehospital information, whereas injury severity score is prognostic data that can only be obtained in the hospital. Moreover, a higher sensitivity, which avoids undertriage, allows us to apply the model as an early triage tool that can determine the priority of dispatching based on patient severity.

For a machine learning dispatch support system, a commercialized model for the recognition of OHCA through dispatching was proposed [17]. The model consists of an automatic speech recognition and textual analysis. In 2 retrospective studies conducted in Denmark and Sweden, positive results were reported in terms of both accuracy and time [18,19]. In a randomized controlled trial in Denmark, the performance of this model surpassed human recognition; however, no significant improvements were found in dispatchers' ability to recognize OHCA with model assistance [20]. Although there are numerous differences in recognizing OHCA and severe trauma, this model also uses machine learning–based text analysis. Another machine learning–based voice analysis model was proposed to recognize the emotional state of OHCA callers [42]. Although the goal differed from our approach, the study also had a small sample size, and the data source was the audio of emergency calls. It is reasonable to expect that the performance of future models may improve with a combination of semantic and emotional analyses.

Limitations

Our study had several limitations. First, human intervention is required for text preprocessing. Conversations in the recordings were often in fragments without a complete grammatical structure and contained specific terms. A customized dictionary for word segmentation, stop word removal, and synonym grouping must be constructed according to the specific medical domain, regional features, and culture. Although there are references, applying these procedures requires researchers to fully comprehend phone conversations [29]. We assumed that everyone who understood the conversation would process the audio and text materials in the same way; otherwise, the features we used in the later steps would be different. This limitation can be overcome by replacing this step with an automatic program [10]. Second, the dispatchers listened to 114 audio clips before providing the keywords. Although they did not know the answers and were asked to provide their opinions based on their experience, they might have chosen words from the audio that they had just heard. Third, owing to strict protocols and the administration's concern for this novel study concept, the recordings of the emergency calls were not allowed to be copied, and only a limited number of crews were permitted to access them over a short period. Given that the preprocessing steps were labor intensive, we could not enroll a large sample size. To compensate for this shortage, we randomly sampled the PAMTs during the entire year of 2018 in the Taipei Trauma Registry. For text classification, the main factors that affect the results may not be determined by data size [43]. Another innovative study with a small amount of data contributed to specific fields [42]. This study serves as a proof of concept and aims to reveal the potential of this methodology for target applications. Nevertheless, some advanced text classification models, such as deep learning models with semantic feature extraction methods [44], may be limited by the size and characteristics of the data. Therefore, research should be conducted on a larger scale with more participants and integrated data to develop a more mature model for actual deployment.

Conclusions

The results of our study suggest that the applied machine learning model is not superior to dispatchers in identifying road accident calls in severe trauma cases; however, the model can assist dispatchers when they lack confidence in the judgment of the calls. A study conducted on a larger scale is required for further model development and validation.

Acknowledgments

Authors' Contributions

KCC and YCC contributed equally as first authors, and WCC and AYC contributed equally as corresponding authors. KCC contributed to the formal analysis, visualization, and original draft. YCC contributed to the data curation, formal analysis, methodology, and original draft. JTS contributed to the conceptualization, data curation, methodology, resources, visualization, and review and editing of the manuscript. CYO contributed to data curation. CHH contributed to data curation and resources. MCT contributed to data curation and resources. MHMM contributed to funding acquisition, project administration, resources, and supervision. WCC contributed to conceptualization, data curation, methodology, visualization review, editing, and supervision. AYC contributed to conceptualization, formal analysis, funding acquisition, project administration, editing, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example texts.
[DOCX File , 18 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Equations and Python script used in the model.
[DOCX File , 29 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Comparison of machine learning models.
[DOCX File , 52 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Descriptive statistics for audio and text files.
[DOCX File , 18 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Profiles and predictive performances of the participating dispatchers.
[DOCX File , 15 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Feature correlation analysis.
[DOCX File , 55 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Relation of the length of text and accuracy.
[DOCX File , 19 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Representative keywords for prehospital-activated major trauma (PAMT) chosen by experts and the PAMT model.

[DOCX File , 23 KB-Multimedia Appendix 8]

References

1. Injuries and violence: the facts 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/149798/9789241508018_eng.pdf [accessed 2022-05-19]

2. Harmsen AM, Giannakopoulos GF, Moerbeek PR, Jansma EP, Bonjer HJ, Bloemers FW. The influence of prehospital time on trauma patients outcome: a systematic review. Injury 2015 Apr;46(4):602-609. [doi: 10.1016/j.injury.2015.01.008] [Medline: 25627482]

3. Chen CH, Shin SD, Sun JT, Jamaluddin SF, Tanaka H, Song KJ, et al. Association between prehospital time and outcome of trauma patients in 4 Asian countries: a cross-national, multicenter cohort study. PLoS Med 2020 Oct 6;17(10):e1003360 [FREE Full text] [doi: 10.1371/journal.pmed.1003360] [Medline: 33022018]

4. Drennan IR, Geri G, Brooks S, Couper K, Hatanaka T, Kudenchuk P, Basic Life Support (BLS), Pediatric Life Support (PLS) and Education, Implementation and Teams (EIT) Taskforces of the International Liaison Committee on Resuscitation (ILCOR), BLS Task Force, Pediatric Task Force, EIT Task Force. Diagnosis of out-of-hospital cardiac arrest by emergency medical dispatch: a diagnostic systematic review. Resuscitation 2021 Feb;159:85-96. [doi: 10.1016/j.resuscitation.2020.11.025] [Medline: 33253767]

5. Zhelev Z, Walker G, Henschke N, Fridhandler J, Yip S. Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack. Cochrane Database Syst Rev 2019 Apr 09;4(4):CD011427 [FREE Full text] [doi: 10.1002/14651858.CD011427.pub2] [Medline: 30964558]

6. Bohm K, Kurland L. The accuracy of medical dispatch - a systematic review. Scand J Trauma Resusc Emerg Med 2018 Nov 09;26(1):94 [FREE Full text] [doi: 10.1186/s13049-018-0528-8] [Medline: 30413213]

7. Gianola S, Castellini G, Biffi A, Porcu G, Fabbri A, Ruggieri MP, Italian National Institute of Health guideline working group. Accuracy of pre-hospital triage tools for major trauma: a systematic review with meta-analysis and net clinical benefit. World J Emerg Surg 2021 Jun 10;16(1):31 [FREE Full text] [doi: 10.1186/s13017-021-00372-1] [Medline: 34112209]

8. Richards CT, Wang B, Markul E, Albarran F, Rottman D, Aggarwal NT, et al. Identifying key words in 9-1-1 calls for stroke: a mixed methods approach. Prehosp Emerg Care 2017;21(6):761-766 [FREE Full text] [doi: 10.1080/10903127.2017.1332124] [Medline: 28661784]

9. Riou M, Ball S, Williams TA, Whiteside A, O'Halloran KL, Bray J, et al. The linguistic and interactional factors impacting recognition and dispatch in emergency calls for out-of-hospital cardiac arrest: a mixed-method linguistic analysis study protocol. BMJ Open 2017 Jul 09;7(7):e016510 [FREE Full text] [doi: 10.1136/bmjopen-2017-016510] [Medline: 28694349]

10. Trujillo A, Orellana M, Acosta MI. Design of emergency call record support system applying natural language processing techniques. In: Proceedings of the 6th Conference on Information and Communication Technologies of Ecuador. 2019 Presented at: TIC.EC '19; November 27-29, 2019; Cuenca City, Ecuador p. 53-65 URL: https://doi.org/10.1007/978-3-030-35740-5_4 [doi: 10.1007/978-3-030-35740-5_4]

11. Choi SW, Ko T, Hong KJ, Kim KH. Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients. Healthc Inform Res 2019 Oct;25(4):305-312 [FREE Full text] [doi: 10.4258/hir.2019.25.4.305] [Medline: 31777674]

12. Chen CH, Hsieh JG, Cheng SL, Lin YL, Lin PH, Jeng JH. Early short-term prediction of emergency department length of stay using natural language processing for low-acuity outpatients. Am J Emerg Med 2020 Nov;38(11):2368-2373. [doi: 10.1016/j.ajem.2020.03.019] [Medline: 32216994]

13. Sterling NW, Patzer RE, Di M, Schrager JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. Int J Med Inform 2019 Sep;129:184-188. [doi: 10.1016/j.ijmedinf.2019.06.008] [Medline: 31445253]

14. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med 2017 Oct 26;56(5):377-389. [doi: 10.3414/ME17-01-0024] [Medline: 28816338]

15. Lucini FR, Fogliatto FS, da Silveira GJ, Neyeloff JL, Anzanello MJ, Kuchenbecker RS, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. Int J Med Inform 2017 Apr;100:1-8. [doi: 10.1016/j.ijmedinf.2017.01.001] [Medline: 28241931]

16. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. PLoS One 2020 Apr 2;15(4):e0230876 [FREE Full text] [doi: 10.1371/journal.pone.0230876] [Medline: 32240233]

17. AI for Patient Consultations. Corti. URL: https://www.corti.ai/ [accessed 2021-09-12]

18. Blomberg SN, Folke F, Ersbøll AK, Christensen HC, Torp-Pedersen C, Sayre MR, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019 May;138:322-329 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.01.015] [Medline: 30664917]

19. Byrsell F, Claesson A, Ringh M, Svensson L, Jonsson M, Nordberg P, et al. Machine learning can support dispatchers to better and faster recognize out-of-hospital cardiac arrest during emergency calls: a retrospective study. Resuscitation 2021 May;162:218-226 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.02.041] [Medline: 33689794]

20. Blomberg SN, Christensen HC, Lippert F, Ersbøll AK, Torp-Petersen C, Sayre MR, et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial. JAMA Netw Open 2021 Jan 04;4(1):e2032320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32320] [Medline: 33404620]

21. Uysal AK, Gunal S. The impact of preprocessing on text classification. Inf Process Manag 2014 Jan;50(1):104-112 [FREE Full text] [doi: 10.1016/j.ipm.2013.08.006]

22. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Doc 1972 Jan;28(1):11-21 [FREE Full text] [doi: 10.1108/eb026526]

23. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the 2013 International Conference on Learning Representations. 2013 Presented at: ICLR '13; May 2-4, 2013; Scottsdale, AZ, USA.

24. Kamath CN, Bukhari SS, Dengel A. Comparative study between traditional machine learning and deep learning approaches for text classification. In: Proceedings of the 2018 ACM Symposium on Document Engineering. 2018 Presented at: DocEng '18; August 28-31, 2018; Halifax, Canada p. 1-11 URL: https://doi.org/10.1145/3209280.3209526 [doi: 10.1145/3209280.3209526]

25. Mccallum A, Nigam K. A Comparison of Event Models for Naive Bayes Text Classification. Association for the Advancement of Artificial Intelligence. 1998. URL: https://www.cs.cmu.edu/~knigam/papers/multinomial-aaaiws98.pdf [accessed 2022-05-19]

26. Trstenjak B, Mikac S, Donko D. KNN with TF-IDF based framework for text categorization. Procedia Eng 2014;69:1356-1364 [FREE Full text] [doi: 10.1016/j.proeng.2014.03.129]

27. Lodhi H, Saunders C, Shawe-Taylor J, Cristianini N, Watkins C. Text classification using string kernels. J Mach Learn Res 2002 Feb;2:419-444.

28. Li Y, Wang X, Xu P. Chinese text classification model based on deep learning. Future Internet 2018 Nov 20;10(11):113 [FREE Full text] [doi: 10.3390/fi10110113]

29. Ma WY, Chen KJ. Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. In: Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing. 2003 Presented at: SIGHAN '03; July 11-12, 2003; Sapporo, Japan p. 168-171 URL: https://doi.org/10.3115/1119250.1119276 [doi: 10.3115/1119250.1119276]

30. Li PH, Fu TJ, Ma WY. Why Attention? Analyze BiLSTM deficiency and its remedies in the case of NER. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020 Presented at: AAAI '20; February 7-12, 2020; New York, NY, USA p. 8236-8244. [doi: 10.1609/aaai.v34i05.6338]

31. Venkatesh, Ranjitha KV. Classification and optimization scheme for text data using machine learning naïve Bayes classifier. In: Proceedings of the 2018 IEEE World Symposium on Communication Engineering. 2018 Presented at: WSCE '18; December 28-30, 2018; Singapore, Singapore p. 33-36 URL: https://doi.org/10.1109/WSCE.2018.8690536 [doi: 10.1109/wsce.2018.8690536]

32. Sheshasaayee A, Thailambal G. Comparison of classification algorithms in text mining. Int J Pure Appl Math 2017;116(22):425-433.

33. Chatterjee A, Gerdes MW, Prinz A, Martinez S. A comparative study to analyze the performance of advanced pattern recognition algorithms for multi-class classification. In: Proceedings of the 2020 Conference on Emerging Technologies in Data Mining and Information Security. 2020 Presented at: IEMIS '20; July 2-4, 2020; Kolkata, India p. 111-124 URL: https://doi.org/10.1007/978-981-15-9774-9_11 [doi: 10.1007/978-981-15-9774-9_11]

34. Chatterjee A, Gerdes MW, Martinez SG. Identification of risk factors associated with obesity and overweight-a machine learning overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]

35. Sewalt CA, Venema E, Wiegers EJ, Lecky FE, Schuit SC, den Hartog D, et al. Trauma models to identify major trauma and mortality in the prehospital setting. Br J Surg 2020 Mar;107(4):373-380 [FREE Full text] [doi: 10.1002/bjs.11304] [Medline: 31503341]

36. Atiksawedparit P, Rattanasiri S, Sittichanbuncha Y, McEvoy M, Suriyawongpaisal P, Attia J, et al. Prehospital prediction of severe injury in road traffic injuries: a multicenter cross-sectional study. Injury 2019 Sep;50(9):1499-1506 [FREE Full text] [doi: 10.1016/j.injury.2019.05.028] [Medline: 31174870]

37. Giannakopoulos GF, Bloemers FW, Lubbers WD, Christiaans HM, van Exter P, de Lange-de Klerk ES, et al. Criteria for cancelling helicopter emergency medical services (HEMS) dispatches. Emerg Med J 2012 Jul;29(7):582-586. [doi: 10.1136/emj.2011.112896] [Medline: 21785150]

38. Wilmer I, Chalk G, Davies GE, Weaver AE, Lockey DJ. Air ambulance tasking: mechanism of injury, telephone interrogation or ambulance crew assessment? Emerg Med J 2015 Oct;32(10):813-816. [doi: 10.1136/emermed-2013-203204] [Medline: 25527473]

39. Coats TJ, Newton A. Call selection for the Helicopter Emergency Medical Service: implications for ambulance control. J R Soc Med 1994 Apr;87(4):208-210 [FREE Full text] [Medline: 8182675]

40. Cameron S, Pereira P, Mulcahy R, Seymour J. Helicopter primary retrieval: tasking who should do it? Emerg Med Australas 2005 Aug;17(4):387-391. [doi: 10.1111/j.1742-6723.2005.00762.x] [Medline: 16091103]

XSL•FO

RenderX

41.   Sinclair N, Swinton PA, Donald M, Curatolo L, Lindle P, Jones S, et al. Clinician tasking in ambulance control improves the identification of major trauma patients and pre-hospital critical care team tasking. Injury 2018 May;49(5):897-902. [doi: 10.1016/j.injury.2018.03.034] [Medline: 29622470]

42.   Chin KC, Hsieh TC, Chiang WC, Chien YC, Sun JT, Lin HY, et al. Early recognition of a caller's emotion in out-of-hospital cardiac arrest dispatching: an artificial intelligence approach. Resuscitation 2021 Oct;167:144-150. [doi: 10.1016/j.resuscitation.2021.08.032] [Medline: 34461203]

43.   Ong MS, Magrabi F, Coiera E. Automated categorisation of clinical incident reports using statistical text classification. Qual Saf Health Care 2010 Dec;19(6):e55. [doi: 10.1136/qshc.2009.036657] [Medline: 20724392]

44.   Lee JY, Dernoncourt F. Sequential short-text classification with recurrent and convolutional neural networks. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016 Presented at: NAACL HLT '16; June 12-17, 2016; San Diego, CA, USA p. 515-520. [doi: 10.18653/v1/n16-1062]

Abbreviations

BNB: Bernoulli naïve Bayes
EMT: emergency medical technician
OHCA: out-of-hospital cardiac arrest
PAMT: prehospital-activated major trauma
RRS-CV: repeated random subsampling cross-validation
TF-IDF: term frequency–inverse document frequency

Possible Impact of a 12-Month Web- and Smartphone-Based Program to Improve Long-term Physical Activity in Patients Attending Spa Therapy: Randomized Controlled Trial

Florie Fillol[1], PhD; Ludivine Paris[2], PhD; Sébastien Pascal[1], MSc; Aurélien Mulliez[3], MSc; Christian-François Roques[4], MD; Sylvie Rousset[5], PhD; Martine Duclos[5,6,7], MD, PhD

[1]Biomouv SAS Inc, Paris, France

[2]University of Auvergne, Clermont-Ferrand, France

[3]Biostatistics Unit (Clinical Research and Innovation Direction), University-Hospital Clermont-Ferrand, Clermont-Ferrand, France

[4]Physical and Rehabilitation Medicine, Paul Sabatier University, Toulouse University, Toulouse, France

[5]Human Nutrition Unity, Centre de Recherche en Nutrition Humaine Auvergne, French National Institute for Agriculture, Food and Environment (INRAE), Clermont-Ferrand, France

[6]Department of Sport Medicine and Functional Explorations, University-Hospital Clermont-Ferrand, G. Montpied Hospital, Clermont-Ferrand, France

[7]Unité fonctionnelle de Recherche Médecine, Clermont University, University of Auvergne, Clermont-Ferrand, France

Corresponding Author:
Florie Fillol, PhD
Biomouv SAS Inc
259 Rue Saint Honoré
Paris, 75001
France
Phone: 33 0664421389
Email: florie.fillol@gmail.com

Abstract

Background: Lack of physical activity (PA) and sedentary behaviors are leading risk factors for noncommunicable diseases (NCDs). Web- and smartphone-based interventions are effective in increasing PA in older adults and in patients with NCD. In many countries, spa therapy, commonly prescribed to patients with NCD, represents an ideal context to initiating lifestyle changes.

Objective: This study aimed to evaluate, in patients attending spa therapy, the effectiveness of an intervention combining a face-to-face coaching and, when returning home, a web- and smartphone-based PA program on the achievement of PA guidelines (PAG) 12 months after the end of spa therapy.

Methods: This was a 12-month, prospective, parallel-group randomized controlled trial. Patients were enrolled during spa therapy and randomized 1:1 to intervention or control group who received PA usual advice. From the end of spa therapy, PA, weight, waist circumference, and quality of life of the participants were assessed by phone every 2 months. Primary outcome was meeting PAG (PA ≥600 metabolic equivalent of task) at 12 months. Secondary outcomes were meeting current PAG at 6 months; sedentary time, weight, waist circumference, PA, and quality of life at 6 and 12 months. Objective use data of the web- and smartphone-based PA program were collected. Analytic methods included intention to treat and constrained longitudinal data analyses.

Results: The study sample included 228 participants (n=176, 77.2% females) with a mean age of 62.4 (SD 6.7) years and a mean BMI of 28.2 (SD 4.2) kg/m². Approximately 53.9% (123/228) of the participants were retired. No group differences were found for any baseline variable. At 12 months, the proportion of patients achieving PAG was significantly higher in intervention group than in the control group (81% vs 67% respectively, odds ratio 2.34, 95% CI 1.02-5.38; *P*=.045). No difference between intervention and control group was found neither in achieving PAG at 6 months nor for sedentary time, weight, and waist circumference at 6 and 12 months. Regarding quality of life, the physical component subscale score was significantly higher at 12 months in the intervention group than in the control group (mean difference: 4.1, 95% CI 1.9-6.3; *P*<.001). The mean duration use of the program was 7.1 (SD 4.5) months. Attrition rate during the first 2 months was 20.4% (23/113) whereas 39.8% (45/113) of the participants used the program for at least 10 months.

Conclusions: PA increased in both the intervention group and the control group. However, at 12 months, more participants met PAG in the intervention group compared with the controls. This indicates that the web- and smartphone-based program could have maintained PA in the intervention group. In addition, a spa therapy seems to be an ideal time and framework to implement PA education.

Trial Registration: ClinicalTrials.gov NCT02694796; https://clinicaltrials.gov/ct2/show/NCT02694796

Introduction

Background

Lack of physical activity (PA) and excess sedentary behaviors are now recognized as leading risk factors for noncommunicable diseases (NCDs), such as cardiovascular diseases, chronic obstructive pulmonary disease, cancers, and type 2 diabetes, which, taken together, are the primary causes of death worldwide [1]. Insufficient PA, or physical inactivity, is defined as a level of PA below the recommended 150 minutes of moderate PA per week, and sedentary behaviors are defined as "any waking behaviors characterized by an energy expenditure <1.5 metabolic equivalents of task (METs), while in a sitting, reclining or lying posture" [2]. In 2017, of the 41 million NCD-related deaths, 1.26 million were because of insufficient PA [3]. In 2016, >1 in 4 (27.5%) adults worldwide were physically inactive [4]. In 2015, a national survey in France showed that among adults aged 55 to 74 years, 42.2% of women and 28.4% of men did not achieve the recommendation of 150 minutes of moderate PA per week [5]. Engaging people in healthy behaviors such as stopping tobacco use, reducing alcohol consumption, adopting healthy diets, increasing PA, and limiting sedentary time is crucial to tackling the rise of NCDs [1]. Although the health benefits of PA are widely recognized [6], engaging older adults and those with NCDs in long-term lifestyle modifications is very challenging. Although many studies have shown the benefits of PA interventions on the health of patients with NCDs [7], a decrease in PA adherence is frequently observed in the long term, leading to a loss of the acquired health benefits [8].

To maintain adherence to PA, information and communication technologies appear to be promising tools that provide personalized follow-up, real-time feedback, and recommendations. Recent reviews and meta-analyses have found that web- and smartphone-based interventions are effective in increasing PA in the general population [9,10], in older adults [11,12], and in patients with an NCD [13,14]. However, another systematic review [15] suggested that multicomponent interventions, where the use of an app was one of several intervention components such as physical education, provision of PA equipment, parental education, face-to-face counselling, might be more effective than stand-alone app interventions.

In many countries (continental Europe, Japan, China, South America, and North Africa), a course of spa therapy is accepted as treatment by the health insurance system and is commonly prescribed to patients with chronic diseases such as rheumatic conditions, respiratory diseases, and skin diseases and patients convalescing from cancer, as well as to those who are overweight or obese. In France, the 3-week courses of therapy delivered in spa centers are reimbursed by the national social security. The context and environment of a stay in a spa therapy center have been shown to be conducive to educating patients about their disease and initiating lifestyle changes, including increasing PA, through patient therapeutic education in PA programs [16-20].

Objectives

We hypothesized that an intervention combining individual face-to-face coaching during spa therapy with a subsequent 12-month web- and smartphone-based PA program would improve PA in patients undergoing spa therapy. The main objective of this study was to evaluate the effectiveness of the intervention compared with the usual advice (ie, standard advice on PA provided during spa therapy) on the achievement of PA recommendations 12 months after the end of spa therapy. Secondary outcomes were to evaluate, throughout the 12-month follow-up, at 6 and 12 months, the effectiveness of the intervention on PA, sedentary time, weight, waist circumference, quality of life of the patients, and engagement with the program (the number of performed PA sessions and frequency of use of the program).

Methods

Study Design

This was a 12-month, prospective, parallel-group, open, multicenter, single-blinded randomized controlled trial (RCT) that enrolled patients attending a 3-week spa therapy treatment. It evaluated the effectiveness of individual face-to-face PA coaching during the stay at the spa therapy facility followed by a 12-month web- and smartphone-based PA program, including a connected wrist pedometer and a connected weighing scale. Participants were randomized 1:1 to either the intervention group or the control group. The participants were enrolled in 1 of 8 French spa therapy facilities: Amélie-les-Bains, Bourbon-Lancy, Brides-les-Bains, Le-Boulou, Chaudes-Aigues, Eugénie-les-Bains, Vals-les-Bains, and Vichy.

Participants and Recruitment

Enrollment and follow-ups were conducted between September 2015 and December 2017. Patients were recruited through posters and flyers displayed in spa therapy facilities and spa physicians' surgeries. A PA instructor was allocated to each spa center to prescreen all potential patients and evaluate their eligibility. Spa physicians participating in the study could also

refer their patients to the PA instructor for prescreening. The inclusion criteria were as follows: age of 50 to 79 years, diagnosis of a stabilized chronic disease (cardiovascular disease, obesity, type 2 diabetes, chronic obstructive pulmonary disease, rheumatic conditions, and breast cancer), BMI between >19 kg/m^2 and <35 kg/m^2, undertaking PA for <150 minutes per week, and having smartphone access to the internet. Exclusion criteria included having a cardiac pacemaker, nonstabilized chronic disease, locomotor disability, evolving metastatic cancer, or a contraindication to PA. Eligible participants underwent a medical examination with the spa physician, who after checking that they could safely follow the study protocol, included them in the trial after the participants provided informed consent. Randomization of the participants to the intervention or control group was stratified by gender and center (thermal spa resort) and performed by the spa PA instructor using a centralized secured management system, REDCap (Research Electronic Data Capture; Vanderbilt University).

Intervention and Control

The intervention comprised a 1-hour individual coaching session with a PA instructor during the 3-week spa therapy stay in 1 of the 8 spa care facilities and then access to the web- and mobile-based PA program and associated connected devices for the 12 months following the end of the spa therapy. All PA instructors received the same training and used the same material. The first part of the consultation aimed to introduce or remind the participants of the benefits of PA for health and disease management. The PA instructor provided advice on how to reach the recommended level of PA and examples of PA adapted to the patient's particular condition. Subsequently, the PA instructor presented the automated web and mobile-based PA program (Thermactive, BIOMOUV SAS Inc) together with the use of connected devices (weighing scales and wrist pedometer; Multimedia Appendices 1 and 2). The PA instructor downloaded the mobile app onto the patient's smartphone and showed him or her how to log into the mobile app and connect and use the weighing scale. The PA instructor also explained access to the website and showed participants the main functionalities of the program. The patients were registered in the program by the PA instructor who completed a web-based questionnaire to determine the patient's PA profile: age, weight, height, physical fitness (endurance, strength, flexibility, and balance measured by the PA instructor), PA, joint disabilities, and pathology. The patient also declared his availability for PA sessions, his PA preferences, and his sports material (such as dumbbells, yoga mats, bands, and wrist weights). Participants in the intervention group followed the web- and mobile-based PA program for 12 months from the end of their 3-week stay in the spa therapy center.

The automated program aimed to help patients achieve recommended levels of PA in 2 ways: by proposing personalized and structured PA sessions and by increasing daily PA (number of steps). The PA sessions were automatically generated based on the patient's profile. To generate personalized PA sessions, an algorithm was developed to select and associate exercises from a database of >1500 different exercises. Each exercise was classified according to its nature (aerobic, strengthening, and balance), part of the body concerned (leg, arm, and trunk), exercise intensity, and duration. The algorithm selected exercises appropriate to a patient's physical capacity and availability and constructed a PA session adapted to the patient. Each PA session comprised 3 phases: a 5-minute warm-up period; either 10 to 35 minutes of exercise to develop muscle strength and flexibility or 10 to 50 minutes of endurance during walking or cycling (mixing continuous and intermittent effort); and finally, a 5-minute recovery phase comprising stretching and relaxation or a return to calm after walking sessions. The PA sessions were either automatically compiled videos or PDF files. The program of PA sessions followed international guidelines regarding the number of sessions per week, resting periods, type of exercise (resistance and endurance), duration, and intensity of each exercise [21]. For each participant, their PA sessions evolved during the course of the intervention taking into account the number of PA sessions completed (recorded by the patient) and any difficulty perceived at the end of the PA sessions (collected using a Borg scale [22]). To increase daily PA, the program generated a daily goal of the number of steps to be achieved based on data from the pedometer over 7 consecutive days. The achievement of these goals determined the subsequent goals, and every day, participants received a notification on their mobile app about the achievement of their personal goals. They also received emails about new PA sessions available on the website and emails reminding them whether a PA session had not been performed and inviting them to do it when possible. Participants had the possibility to record or add activities on the mobile app, which were not planned in the program, such as walking, cycling, swimming, or fitness sessions. The website and the mobile app also allowed participants to record their daily PA and amount of sedentary time to visualize their evolution over time.

Patients allocated to the control group received the usual advice on PA and a booklet providing advice and examples of PA suited to their pathology. At the end of the 12-month follow-up period, the patients included in the control group received free connected devices and access to the Thermactive program for 12 months.

Measurements and Follow-up

Data collected during the study and follow-up were recorded using an electronic case report form in a centralized secured management system, REDCap.

Demographic variables of the participants, including sex, age, weight, waist circumference, highest level of formal education (high school or less and higher education), occupation (nonworking [retired or unemployed], manager [artisan or intellectual profession], and employee [employee, intermediate occupation, and worker]), condition treated by spa therapy, medical-surgical and family history, medical treatments, physical fitness, PA, and quality of life were collected at baseline (month 0 [M0]) by the PA instructor. PA was assessed using the validated International Physical Activity Questionnaire (IPAQ)-short version [23]. The IPAQ measures the frequency (days per week) and duration (minutes) of PA during the past 7 days in the following domains: work, transportation, work at home, and leisure activities [23]. Different levels of PA

(walking, moderate, vigorous, and total) were calculated and expressed in METs minutes per week (a product of PA intensity and duration). PA was classified as low (<600 MET minutes per week), moderate (600-3000 METs), or high (>3000 METs) [23]. Meeting current PA guidelines (PAG) was defined as a total PA of ≥600 METs [23].

At inclusion (M0), physical fitness was evaluated in both groups using validated physical fitness field tests from *Eurofit for Adults* [24] and the *Senior Fitness Tests* [25], with the 6-minute walk test to assess cardiorespiratory fitness (endurance), the arm curl test, the 30-second chair stand test for muscle strength, the lateral side–bending test for flexibility and patients' balance by the one-leg standing test, and the timed up and go test for balance. Quality of life was assessed using the Short Form Health Survey-12 (SF-12; version 2) [26]. The SF-12 assesses limitations in role functioning with 12 items. It consists of 2 subscales measuring physical health (physical component subscale [PCS]) and mental health (mental component subscale). The presence and severity of different impairments over the past 4 weeks are rated. Subscale scores can vary between 0 and 100, with higher scores indicating less impairment or greater health well-being.

From the end of the 3-week spa therapy, PA, body weight, waist circumference, and quality of life of the participants in both groups were assessed at month 2 (M2), month 4, month 6 (M6), month 8, month 10, and month 12 (M12) by interviewers (masked to the participant's randomization group) by phone. Data were collected every 2 months to avoid a loss to follow-up and to allow more precise measurement of change in outcome over time.

To limit missing data, participants were contacted 3 times for each follow-up phone interview. First, the participants were contacted by email to plan the phone interview; in case of no answer, an SMS text message was sent to his or her cell phone within 7 days, and after failing to respond within 3 days of the SMS text message, he or she was contacted directly by phone. The interviewer tried to contact nonresponders for 1 month after the theoretical follow-up date.

Outcomes

The primary outcome was meeting the current PAG at 12 months after the end of spa therapy, defined as reporting total PA ≥600 METs [23] measured by the IPAQ short form.

Secondary outcomes were meeting the current PAG at 6 months after the end of spa therapy; sedentary time, weight, waist circumference, PA, and quality of life at 6 and 12 months; and changes in these indicators evaluated every 2 months during the 12-month follow-up.

The use of the program was evaluated by the number of connections to the Thermactive website, number of PA sessions conducted (structured PA sessions+recorded PA sessions), and number of months for which use of the program was maintained.

Sample Size

With a risk of 0.05, a power (1-b) of 0.90, and assuming a detectable difference in patients meeting the PAG of 15%

between the 2 groups [27], the sample size required was 462, with 231 participants in each study arm.

Statistics

Continuous variables were described as mean (SD or 95% CI) or median (IQR). The normal distribution of continuous variables was checked using the Shapiro-Wilk test. To compare between-group differences, a Student t test (2-tailed) was used for variables with normal distribution; otherwise, the Mann-Whitney test was used. Categorical variables are presented as frequencies and percentages and were compared between groups using a chi-square test. To test effectiveness, the data were analyzed using intention-to-treat principles [28]. As all randomized patients were included in the analyses and considering that assessment every 2 months should limit the loss to follow-up, attrition was not considered to increase the sample size [29]. To compare between-group differences from baseline for repeated outcomes, a constrained longitudinal data analysis (CLDA) was used. This mixed model is a constrained full-likelihood approach, whereby both the baseline and postbaseline values are modeled as dependent variables (the constrained model assumes that both the baseline and postbaseline measurements are jointly multivariate and normally distributed as the baseline value is treated as part of the response vector), and the true baseline values are constrained to be the same for the 2 treatment groups. Such methods based on maximum likelihood are consistent under the *missing at random* assumption. This model allows the inclusion of patients for whom either the baseline or postbaseline measurements are missing, thereby increasing efficiency [30]. Hence, this analysis provides an adjustment for the observed baseline difference in estimating the intervention effects. Time was treated as a categorical variable so that no restriction was imposed on the trajectory of the means over time. In addition to adjusting for baseline covariates, the analysis model was also adjusted for the intervention, time, sex, and interaction of time and intervention. Random effects at the patient and center levels were also included. The results are expressed as odds ratios (ORs) with 95% CI and P values for categorical variables and as differences in mean change from baseline to 1 year with 95% CI for continuous variables. All statistical tests were 2-sided, and $P<.05$ was considered statistically significant. Data were analyzed using Stata 12.

Safety

All serious adverse events (AEs) were recorded and notified to the French clinical trials pharmacovigilance system.

Ethics Approval

The trial, funded by Association Française pour la Recherche Thermale (grant number 2015-02), a nonprofit independent organization, was approved by the National Agency for the Safety of Medicine and Health Products and the regional ethics committee (Comité de Protection des Personnes Sud-Est N 6; registration number: CPP AU1196; registration number IDRCB:2015-A00855-44) and registered at ClinicalTrials.gov (NCT02694796) before enrollment of the participants began.

Results

Patients

Recruitment was conducted from September 2015 to December 2016. Of the 304 patients screened, 230 (75.6%) were enrolled and randomly assigned to either the control group (n=114, 49.6%) or intervention group (n=116, 50.4%; Figure 1). After randomization, 0.9% (2/230) of patients (1 in each group) withdrew their participation; thus, a total of 228 patients were included in the analyses. Patient characteristics are presented in Table 1. More than 3 participants out of 4 were women (176/228, 77.2%). The mean age of the sample was 62.4 (SD 6.7) years, and 53.9% (123/228) of the participants were retired. The 2 groups did not differ in any variable recorded at baseline.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. IPAQ: International Physical Activity Questionnaire.
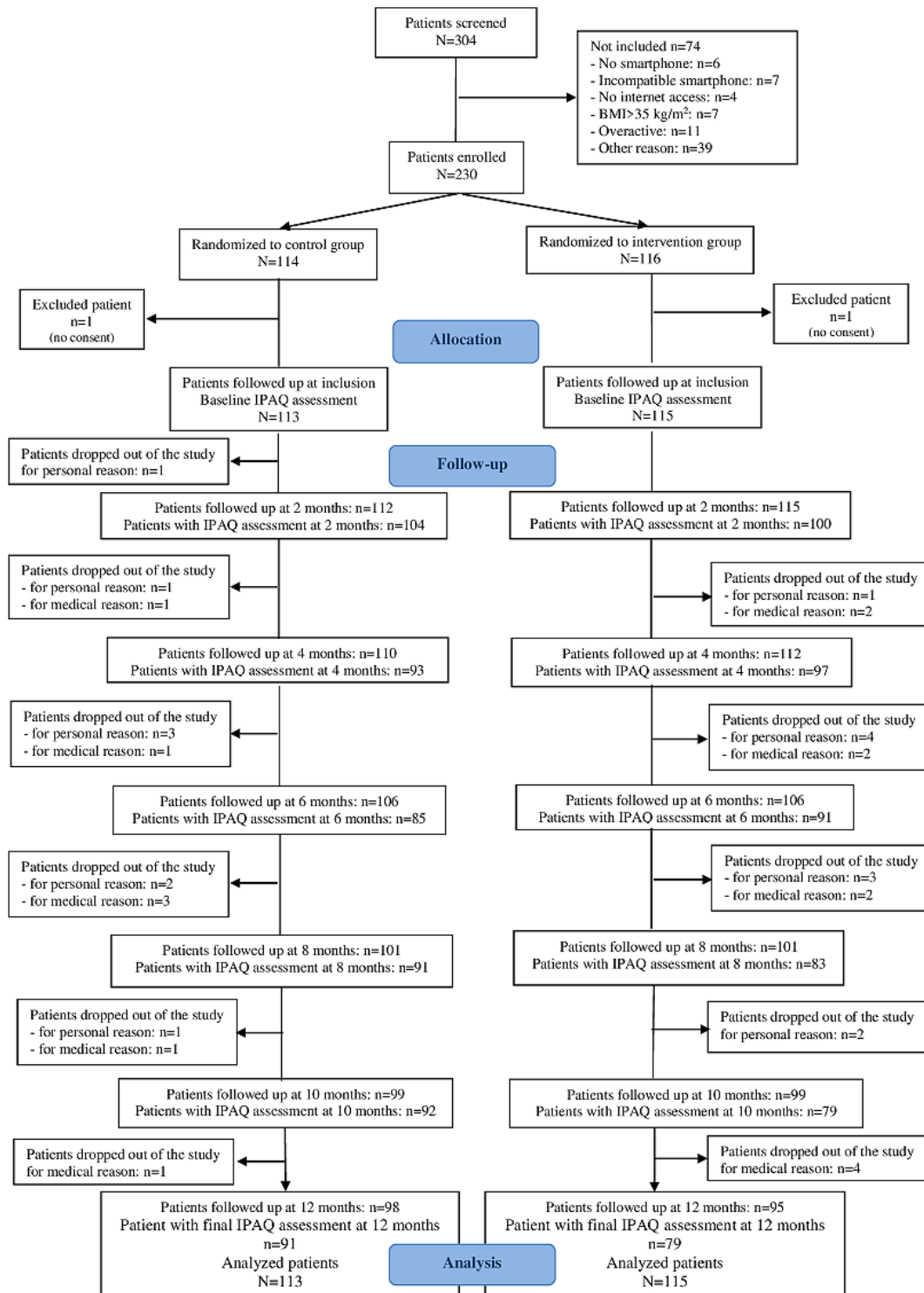
Table 1. Baseline characteristics of the participants in the control and intervention group (N=228).

Characteristics	Control group (n=113)	Intervention group (n=115)	Total
Female, n (%)	86 (76.1)	90 (78.3)	176 (77.2)
Age (years), mean (SD)	62.3 (6.9)	62.6 (6.6)	62.5 (6.7)
Weight (kg), mean (SD)	76.3 (15.1)	77.0 (14.3)	76.7 (14.6)
BMI, mean (SD)	28.3 (4.4)	28.2 (4.0)	28.3 (4.2)
Waist circumference (cm), mean (SD)	96.7 (13.6)	95.0 (12.4)	95.8 (13.0)
Educational level, n (%)			
High school or less	62 (54.9)	55 (47.8)	117 (51.3)
Higher education	51 (45.1)	60 (52.2)	111 (48.7)
Occupation, n (%)			
Nonworking (retired, unemployed, housewife or househusband, disability, or long-term leave)	77 (68.1)	74 (64.3)	151 (66.2)
Manager (artisan, trader, senior executive, or intellectual profession)	18 (15.9)	16 (13.9)	34 (14.9)
Employee (intermediate occupation or worker)	17 (15)	24 (20.9)	41 (18)
Indication for spa treatment, n (%)			
Arthrosis	82 (72.6)	83 (72.2)	165 (72.4)
Cardiovascular diseases	14 (12.4)	11 (9.6)	25 (11)
Obesity	17 (15)	19 (16.5)	36 (15.8)
Diabetes (type 1 and type 2)	8 (7.1)	12 (10.4)	20 (8.8)
COPD[a]	2 (1.8)	1 (0.9)	3 (1.3)
Cancer	3 (2.7)	4 (3.5)	7 (3.1)
Other	26 (23)	22 (19.1)	48 (21.1)
Physical fitness, mean (SD)			
Resting heart rate (beats per minute)	70.1 (9.8)	71.1 (10.7)	70.6 (10.3)
6-minute walk test (minutes)	463 (97.6)	464.6 (94.6)	463.8 (95.9)
Arm curl test (number of flexions)	22 (6.9)	22.1 (7.2)	22.0 (7.0)
30-second chair stand test (number of up-and-down)	14.3 (4.1)	13.8 (4.4)	14.0 (4.2)
Lateral side–bending test (right side; cm)	15.9 (4.2)	15.2 (3.5)	15.6 (3.9)
Lateral side–bending test (left side; cm)	16 (4.3)	15.4 (3.7)	15.7 (4.0)
One-leg standing test (seconds)	6.2 (9.0)	6.0 (6.3)	6.1 (7.8)
Timed up and go test (seconds)	6.2 (1.5)	6.2 (1.7)	6.2 (1.6)
PA[b] (IPAQ[c]; MET[d] minutes per week), median (IQR)			
Continuous score for vigorous intensity	0 (0-960)	0 (0-480)	0 (0-960)
Continuous score for moderate intensity	120 (0-240)	240 (0-360)	130 (0-360)
Continuous score for walking	198 (66-396)	198 (66-346.5)	198 (66-396)
Continuous score for overall activity	396 (198-664)	419 (238-720)	396 (198-686)
Sedentary time (IPAQ; minutes), median (IQR)			
Time spent sitting on a week day	300 (240-420)	360 (270-480)	360 (240-480)
Time spent sitting on a weekend day	300 (240-360)	300 (240-360)	300 (240-360)
Time spent watching television on a week day	120 (120-180)	120 (120-180)	120 (120-180)
Time spent watching television on a weekend day	120 (120-240)	120 (120-180)	120 (120-180)
Time spent in front of computer or tablet on a week day	120 (60-180)	120 (60-240)	120 (60-210)
Time spent in front of computer or tablet on a weekend day	60 (30-120)	60 (45-150)	60 (30-120)

Characteristics	Control group (n=113)	Intervention group (n=115)	Total
Quality of life (SF-12[e]; 0-100), mean (SD)			
Physical health (PCS[f])	43.2 (8.5)	43.3 (8.5)	43.2 (8.5)
Mental health (MCS[g])	47.0 (9.5)	48.1 (8.9)	47.6 (9.2)

[a]COPD: chronic obstructive pulmonary disease.

[b]PA: physical activity.

[c]IPAQ: International Physical Activity Questionnaire.

[d]MET: metabolic equivalent of task.

[e]SF-12: Short Form Health Survey-12.

[f]PCS: physical component subscale.

[g]MCS: mental component subscale.

Primary Outcome

The change in the percentage achieving PAG marginal values according to CLDA modeling for each group is presented in Figure 2, and the statistical comparison between the groups for PAG achievement is shown in Table 2. The achievement of PAG significantly increased in both groups from M0 to M12 (Table 2), with the greatest increase occurring between M0 and M2 (Figure 2). At 12 months, the proportion of patients achieving PAG was significantly higher in the intervention group than in the control group (64/79, 81% vs 61/91, 67%, respectively; Figure 2; OR 2.34, 95% CI 1.02-5.38; P=.045; Table 2). The CLDA analysis also showed that significantly fewer women achieved PAG than men (P=.005; Table 2).

Figure 2. Change in the percentage of PA guidelines achievement (total PA MET≥600) marginal values according to constrained longitudinal data analysis model for each group over time. MET: metabolic equivalent of task; PA: physical activity.
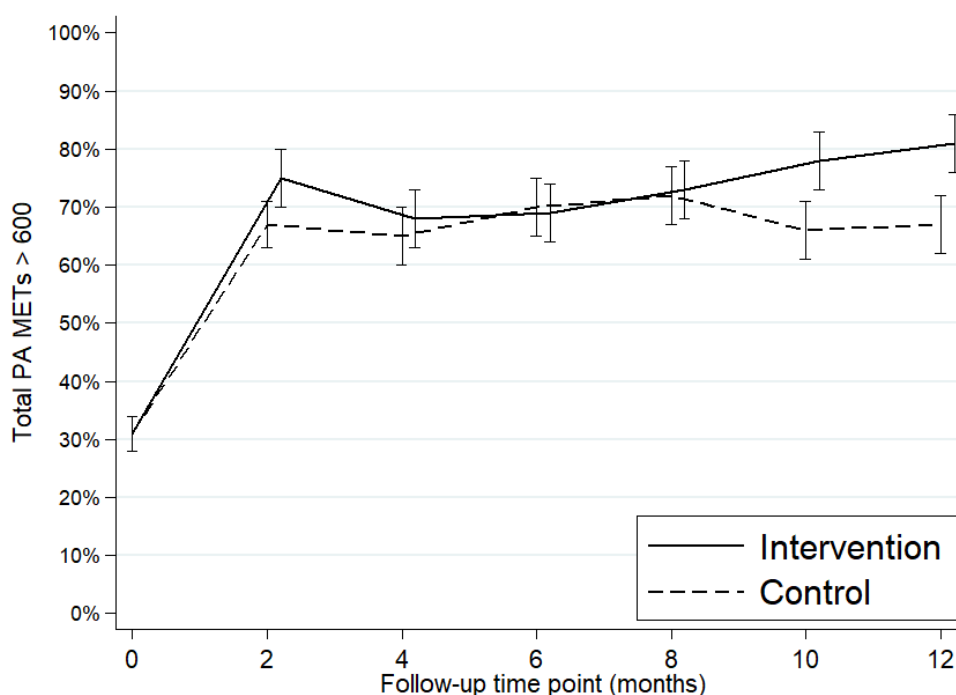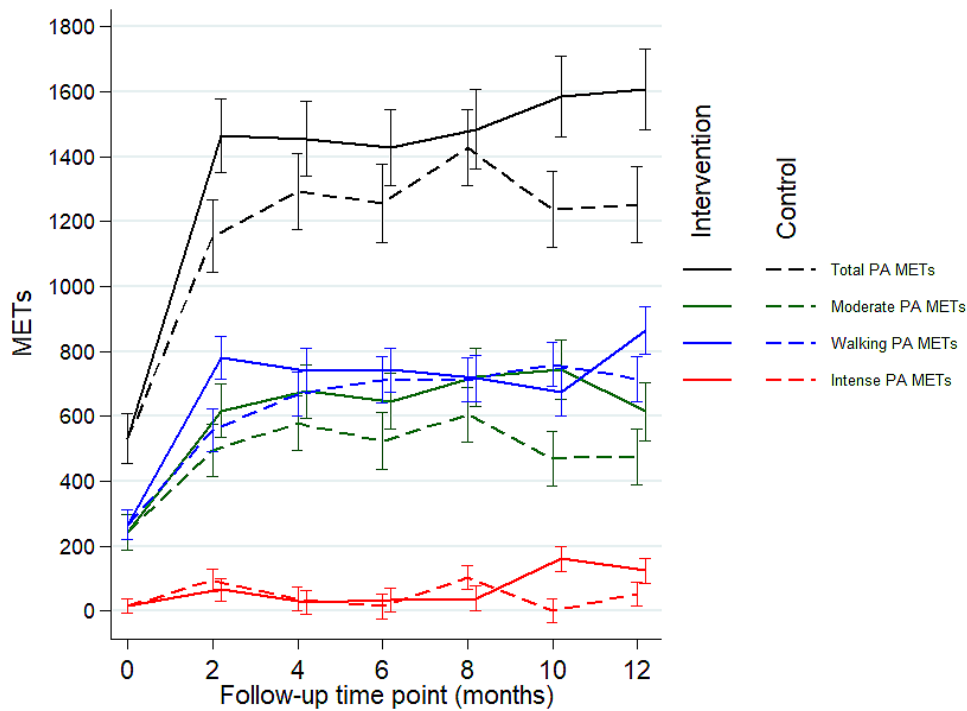
Table 2. Constrained longitudinal data analysis model of the achievement of physical activity guidelines (total physical activity metabolic equivalents of task ≥600) over time.

Characteristics	Odds ratio (95% CI)	P value
Female	0.52 (0.33-0.83)	.005
Inclusion visit (month 0)	N/A[a]	N/A
2-month visit (month 2)	6.3 (3.55-11.17)	<.001
4-month visit (month 4)	5.49 (3.05-9.87)	<.001
6-month visit (month 6)	7.41 (3.97-13.85)	<.001
8-month visit (month 8)	8.37 (4.5-15.55)	<.001
10-month visit (month 10)	5.79 (3.2-10.48)	<.001
12-month visit (month 12)	6.29 (3.45-11.46)	<.001
Intervention group×month 2	1.58 (0.77-3.23)	.21
Intervention group×month 4	1.18 (0.58-2.41)	.65
Intervention group×month 6	0.95 (0.45-2.01)	.89
Intervention group×month 8	1.04 (0.47-2.26)	.93
Intervention group×month 10	2.12 (0.95-4.74)	.07
Intervention group×month 12	2.34 (1.02-5.38)	.045

[a]N/A: not applicable.

Secondary Outcomes

PA and Sedentary Times

At 6 months follow-up, the achievement of PAG did not differ between the intervention and control groups (63/91, 69.2% and 59/84, 70.2% of patients reached the PAG, respectively; Figure 2; OR 0.95, 95% CI 0.45-2.01; P=.89; Table 2). Regarding the PA level (Figure 3), the IPAQ score of total PA at M12 was significantly higher in the intervention group than in the control group (intervention group total PA 1618 METs, 95% CI 1491-1744 METs vs control group total PA 1275 METs, 95% CI 1140-1385 METs; P=.04), whereas no significant difference was observed at M6 (intervention group total PA 1427 METs, 95% CI 1303-1564 METs vs control group total PA 1274 METs, 95% CI 1146-1392 METs; P=.30).

There were no statistically significant differences between the 2 groups at M6 or M12 regarding the IPAQ scores for walking, moderate, and vigorous PA (Figure 3) or for sitting time or time spent in front of a screen (television or computer) during weekdays or weekends (Figure 4).

Nevertheless, the time spent in front of a screen (computer or television) decreased significantly over the follow-up in both the groups during both weekdays and weekends (Table 3).

Figure 3. International Physical Activity Questionnaire scores for total, moderate, intense, and walking physical activity margin values according to constrained longitudinal data analysis model for each group over time. MET: metabolic equivalent of task; PA: physical activity.

Figure 4. Sedentary times marginal values according to constrained longitudinal data analysis model for each group over time.
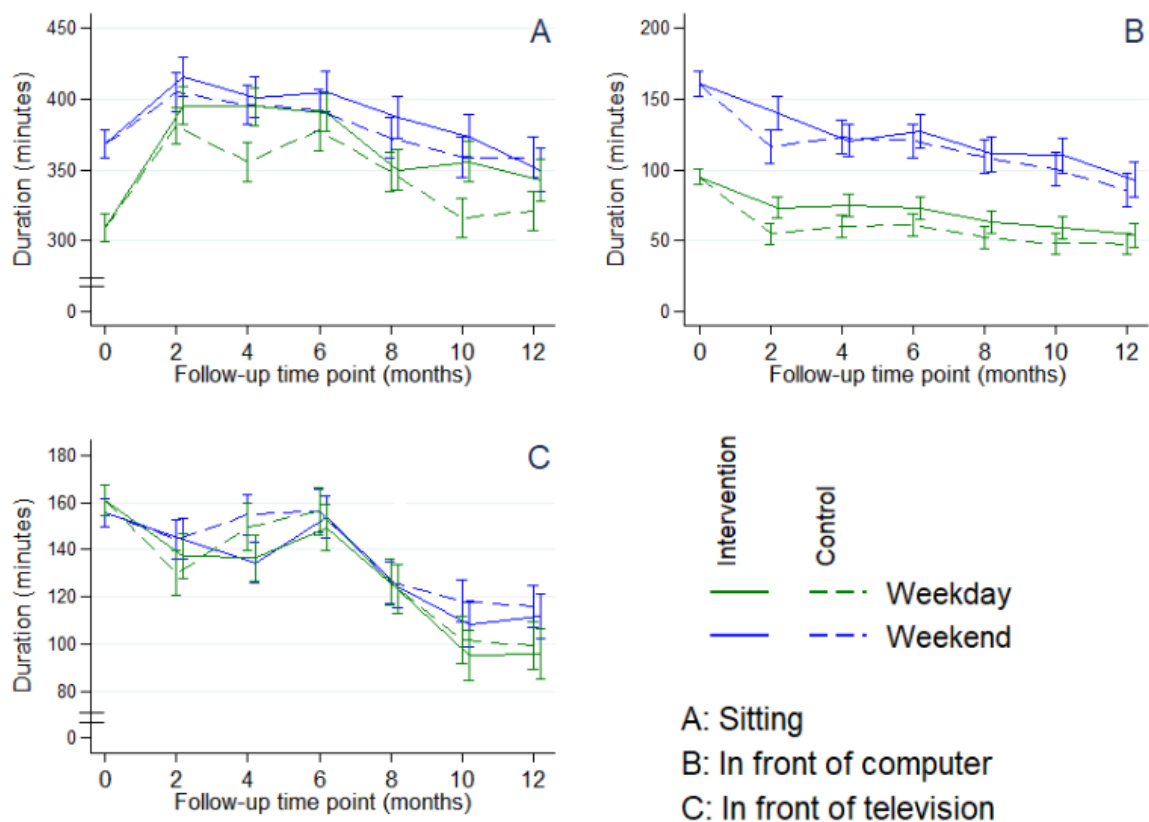
Table 3. Change in time spent in front of screens (computer or television) for control and intervention groups pooled.

Duration of time	Change between month 12 and month 0, mean (SE; 95% CI)	P value
Time spent in front of a computer during the week (minutes)	−75.2 (10.2; −95.3 to −55.2)	<.001
Time spent in front of a computer during the weekend (minutes)	−47.2 (7.8; −62.6 to −31.8)	<.001
Time spent in front of a television during the week (minutes)	−39.7 (9.3; −58.0 to −21.4)	<.001
Time spent in front of a television during the weekend (minutes)	−61.9 (10.8; −83.0 to −40.7)	<.001

Body Weight and Waist Circumference

There was no statistically significant difference between the 2 groups for body weight or waist circumference at M6 and M12 (Figure 5). However, the mean waist circumference for the 2 groups had significantly decreased at 6 months by 1.9 cm (95% CI −3.0 to −0.8 cm; P=.001) and at 12 months by 2.4 cm (95% CI −3.5 to −1.3 cm; P<.001).

Figure 5. Weight and waist circumference marginal values according to constrained longitudinal data analysis model for each group over time.
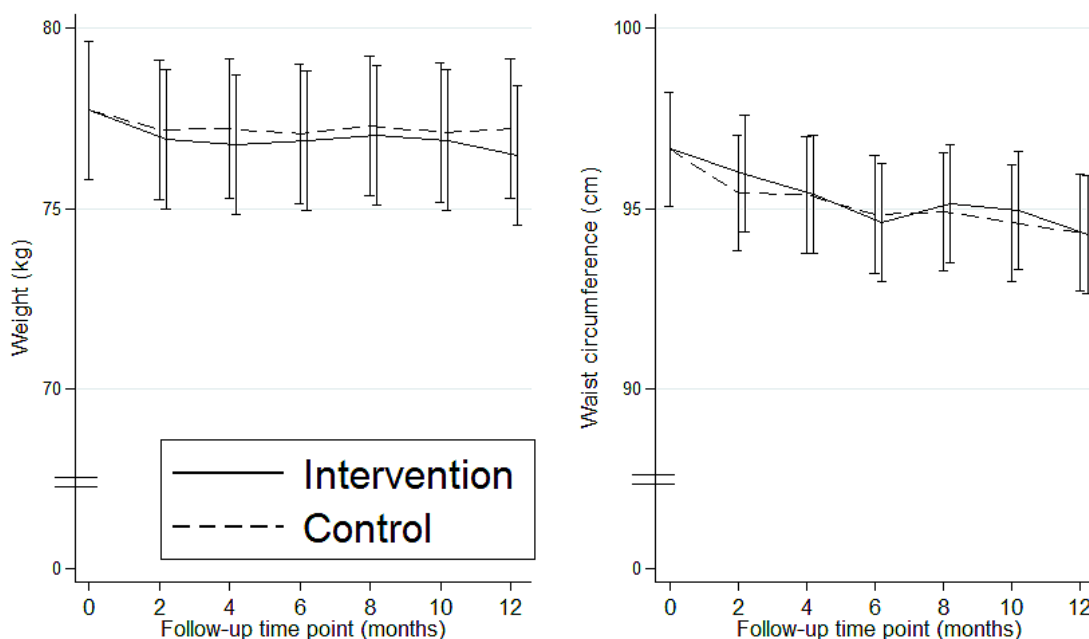
Quality of Life

The quality of life assessment showed that the PCS score was significantly higher at M12 in the intervention group than in the control group (Figure 6; mean difference at M12 4.1, 95% CI 1.9-6.3; P<.001). At M6, the PCS score tended to be higher in the intervention group than in the control group (PCS score 2.1, 95% CI 0.0-4.3; P=.06). There were no statistically significant differences between the 2 groups in the mental component subscale score at M6 or M12 (Figure 6).

Figure 6. SF-12 scores (PCS and MCS) marginal values to constrained longitudinal data analysis model for each group over time. MCS: mental component subscale; PCS: physical component subscale; SF-12: Short Form Health Survey-12.
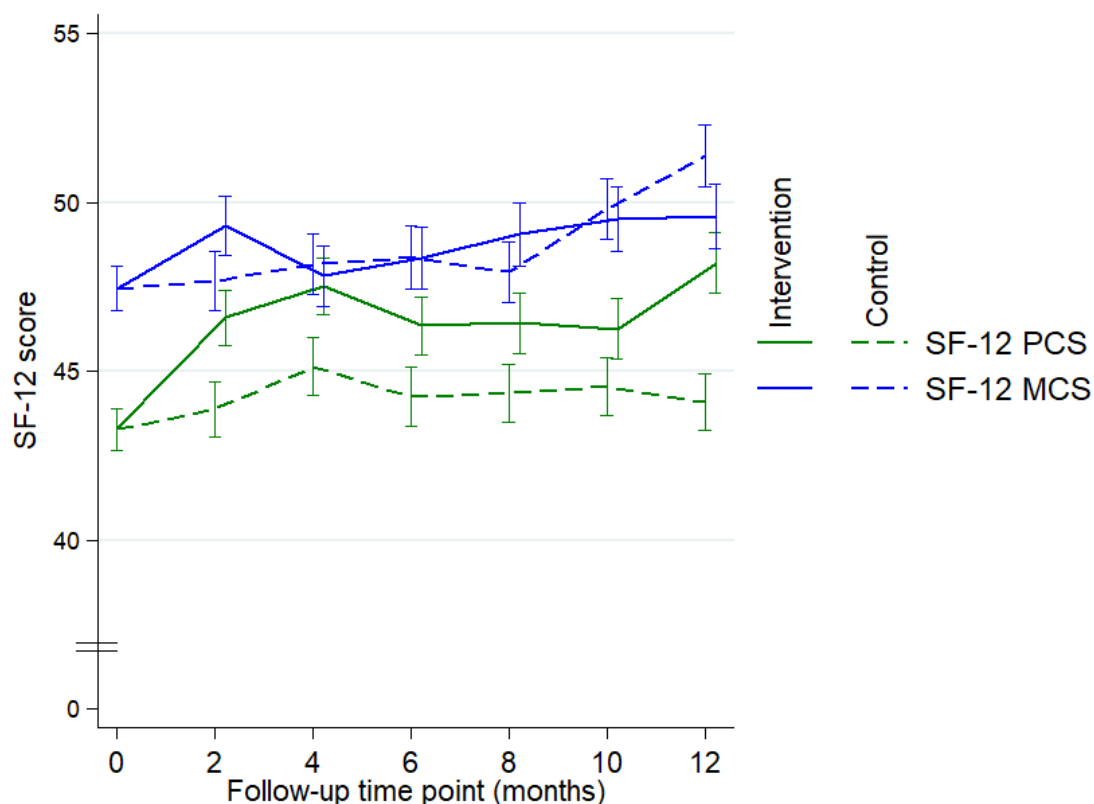
Use of the Program

Monitoring of the program use results is presented in Table 4. The patients used the program for an average of 7.1 (SD 4.5) months. Approximately 20.4% (23/113) dropped out of the program before 2 months of use; however, 39.8% (45/113) of the participants used the program for ≥10 months (Table 4). Among the participants, 62.8% (71/113) had at least one structured PA session.

Table 4. Use of the program (N=115).

Characteristics	Intervention group
Logging into the program	
Patients who logged into the program at least once, n (%)	113 (98.3)
Total number of log-ins into the program, N	16,325
Number of log-ins by patients, mean (SD)	143.2 (179.4)
Number of log-ins by patients, median (IQR)	76 (24.3-208.8)
Duration of program use	
Duration of use (months), mean (SD)	7.1 (4.5)
Patients who used the program for <2 months, n (%)	23 (20.4)
Patients who used the program for 2 to 4 months, n (%)	14 (12.4)
Patients who used the program between 4 and 6 months, n (%)	13 (11.5)
Patients who used the program between 6 and 8 months, n (%)	8 (7.1)
Patients who used the program between 8 and 10 months, n (%)	10 (8.8)
Patients who used the program for >10 months, n (%)	45 (39.8)
Total PA[a] sessions (recorded+structured)	
Patients who conducted at least one PA session, n (%)	81 (71.7)
Total number of PA sessions conducted, N	2588
Number of PA sessions conducted, median (IQR)	16 (3-47)
Structured PA sessions	
Patients who conducted at least one structured PA session, n (%)	71 (62.8)
Total number of structured PA sessions conducted, N	1836
Number of structured PA sessions conducted, median (IQR)	8 (2-34)
Patients who conducted <1 structured PA session by month of use, n (%)	25 (35.2)
Patients who conducted 1 to 4 structured PA sessions by month of use, n (%)	26 (36.6)
Patients who conducted 4 to 8 structured PA sessions by month of use, n (%)	16 (22.5)
Patients who conducted >8 structured PA sessions by month of use, n (%)	4 (5.6)

[a]PA: physical activity.

Safety

AEs recorded during the study are presented in Table 5. None of the severe AEs were attributed to the intervention. One patient reported an aggravation of lymphedema in the left arm because of wearing a wrist pedometer. This adverse effect was resolved by physiotherapy.

Table 5. Adverse events recorded during the follow-up (N=228).

Adverse events	Control group (n=113), n (%)	Intervention group (n=115), n (%)
Adverse events	102 (49.6)	70 (60.9)
Severe adverse events	13 (11.5)	11 (9.6)
Increased arthrosis	2 (1.8)	0 (0)
Hospitalizations or care for a disorder unrelated to the spa indication	11 (9.7)	11 (9.6)

Discussion

Principal Findings and Comparison With Prior Work

This RCT aimed to assess the effectiveness of an intervention, including an initial face-to-face coaching and a web- and mobile-based PA program, to meet PAG among patients attending a 3-week spa therapy treatment. The results showed that significantly more participants met the PAG at the 12-month follow-up in the intervention group than in the controls; however, no difference was observed between the 2 groups for reaching PAG at 6 months. The intervention significantly improved the physical component of the quality of life at 12 months. Sedentary times and waist circumference were

significantly reduced in both groups at 6 and 12 months of follow-up without significant differences between the groups.

The level of PA increased in both groups but was significantly higher at 12 months in the intervention group. The increase in PA in the control group might be explained by the usual advice on PA and lifestyle changes provided during the 3-week spa therapy by health care professionals. Indeed, a number of studies have shown that the context and environment of spa treatments represent an opportunity to educate patients on their chronic diseases and initiate behavioral changes [16-20], such as PA.

Our analyses showed that the effect of usual advice on PA in the control group was the highest during the first 2 months after the spa therapy; subsequently, this tended to stabilize and finally slightly decreased after 8 months. Although the PA in the intervention group followed the same dynamic for the first 8 months, it increased after 8 months and became significantly higher at 12 months.

The maintenance of the level of PA to reach the PAG at 12 months in the intervention group could be explained by the web- and mobile-based PA program. This result is in line with the results observed in other RCTs aimed at improving PA among older adults using web-based PA interventions [31,32]. A systematic review and meta-analysis evaluated the effects of eHealth interventions on promoting PA in older adults [12]. The results of this meta-analysis showed that the effects of the eHealth intervention (vs controls) on PA time measured by questionnaires and objective wearable devices on energy expenditure and step counts were all significant with minimal heterogeneity.

Our findings also highlight that the intervention significantly improved the physical component of quality of life at 12 months, which is consistent with the increase in physical abilities because of the improvement in PA level. Limited studies have reported on the effect of web- or mobile-based PA interventions on quality of life among older adults. A randomized control trial that included 235 participants indicated that after 3 months, an internet-based intervention aimed at increasing PA significantly improved the quality of life of inactive older adults [33]. Another study conducted by Irvine et al [34] also showed a significant improvement in the SF-12 PCS among sedentary older adults aged >55 years who engaged in a web-based PA program.

Our results indicate that waist circumference was significantly reduced in both groups at 6 and 12 months of follow-up without a significant difference between the groups.

A meta-analysis [35], including 31 RCTs, emphasized that internet-based interventions showed a significant reduction in waist circumference (mean change −2.99 cm, 95% CI −3.68 to −2.30 cm; I^2=93.3%) compared with minimal interventions such as information-only groups. Our findings indicate a similar mean change in waist circumference in the 2 groups (−2.4 cm; 95% CI −3.5 to −1.3 cm). Therefore, this reduction did not seem to be explained by the intervention. The inclusion in a research study and the focus on their medical conditions should motivate them to adopt better health behaviors. The time spent sitting was higher at M2, month 4, and M6 in both groups than that at baseline. This could be because of fatigue related to the increase in PA [36], which induced compensatory time spent being sedentary, probably at the expense of light PA (unassessed by the IPAQ questionnaire, but which can represent most PA in older adults). This hypothesis should be confirmed in future studies.

Our results also indicate that men were more likely to successfully reach the PAG than women. The present findings are consistent with those of previous studies. Blanchard et al [37] evaluated PA levels in patients with heart disease over 12 months (with or without cardiac rehabilitation) and showed a more pronounced decline in PA over time in women than in men. Jenkins and Gortner [38] specifically examined gender disparity in PA in people living with heart disease who did not receive cardiac rehabilitation. The results showed that men walked significantly more than women at 1, 2, 6, and 12 months after hospitalization. However, analyzing the determinants of parameters that establish which factors predict which participants are successful in reaching PAG was not a part of our research question. Such a determinant analysis will be performed in forthcoming studies and will address different research questions with the ultimate aim of better targeting different populations.

Limitations and Strengths

The effect of the intervention on maintaining long-term PA and reaching PAG needs to be viewed cautiously as, despite an extension of the enrollment period, the a priori sample size was not met. Two main reasons explain the difficulties in including participants in the trial. First, it appeared that many patients with a web connection and smartphone were already meeting the PAG. Second, we encountered difficulties in the recruitment of qualified PA instructors who played an essential role in the prescreening of participants and face-to-face coaching of the intervention group.

Another limitation of our trial is the self-reported assessment measures, making them potentially subject to social desirability bias [39]. Furthermore, the Hawthorne effect [40] (referring to a tendency in some individuals to alter their behavior in response to their awareness of being observed) along with contamination bias could also affect the magnitude of the differences observed in the results. However, the contamination bias cannot call into question our main result as it reduced the size of the difference between the 2 groups. Therefore, we can hypothesize that without contamination bias, the difference between the 2 groups would have been greater.

The Hawthorne effect and the repeated assessment of outcomes every 2 months could motivate participants to become more active, leading them to overestimate the report of PA and consequently bias our findings. Although this bias could have occurred in both the control and intervention groups and, therefore, would not bias the comparison between the 2 groups, the proportion of participants achieving PAG might be overestimated. Moreover, we cannot exclude that participants in the intervention group may be influenced by the expectation that they will perform better as they received the promising PA program, especially at the end of the program, resulting in an overestimation of their PA level.

A greater number of patients was assessed at M12 in the control group (91/113, 80.5%) than in the intervention group (79/115, 68.7%). One of the reasons for this higher compliance of the control group may be the promise to have free access to the program at the end of the follow-up.

The use of the program can be considered satisfactory as patients used the program for an average of 7.1 (SD 4.5) months; 78.3% (90/115) of the patients used the program for at least 2 months and 39.1% (45/115) for at least 10 months. Approximately 61.7% (71/115) of patients reported engaging in structured PA sessions (median 8 sessions), emphasizing the clear interest of participants in the value of the program, as well as its acceptability and usability. Indeed, the attrition rate for web and smartphone interventions in PA is often quite high [41] (ranging from 30% [42] to 80% [43]), and declining rates of engagement over time are often reported by researcher-led web-based health interventions.

In a secondary analysis of a randomized trial [42], attrition at 3 months of a 100-day PA intervention delivered via an app ranged from 32% to 39%. Another RCT found that 80% of participants ceased using a web-based PA intervention by week 80 (20 months), and the attrition rate was approximately 70% to 75% at 12 months [43].

The percentage of patients who stopped using the web application and mobile app before 4 months was 32.8% (37/113) in this study.

Thus, the attrition rate observed in our study was consistent with that reported in the literature.

A recent study [44] examined the effect of individualized follow-up with an app for 1 year on peak oxygen uptake in patients undergoing cardiac rehabilitation. The results of this study showed high levels of use of the app in the intervention group: 84% (46/55) of the patients used it to set and achieve personal goals and tasks. The intervention group improved in the peak oxygen uptake to a larger extent than the control group (without the app). Adherence to app use was more than twice the web and app adherence estimated in this study (45/113, 39.8%). This could be mainly explained by the fact that in the study of Lunde et al [44], monitoring and feedback were provided by a real person to the patients, whereas in our study, the PA program was fully automated. The authors explained that the high level of individualization (having a real person behind the app, as well as quite simple technology) may have been crucial to maintaining adherence to app use.

Therefore, adherence in the long term (>10 months) to the web- and mobile-based PA program studied here would be enhanced by introducing engagement with a real PA instructor in the follow-up of the patients.

In our analyses (not shown in the manuscript), we compared the *respondents* and those with *missing* data at 12 months by baseline characteristics.

Those with *missing* data differed from the *respondents* by the baseline declaration of *high PA* and *sitting time during the weekend*. The proportion of those with missing data who declared practicing high PA at baseline was higher than the proportion of the *respondents* (4/58, 6.9% vs 3/170, 1.8%, respectively; P=.05). For sitting time, those with *missing* data declared, at baseline, to spend less time sitting during the weekend than the *respondents* (280 vs 320 minutes, respectively; P=.046). Nevertheless, those with *missing* data were more frequent in the intervention group than in the control group (36/115, 31.3% vs 22/113, 19.5%, respectively). Therefore, if we hypothesized that those with *missing* data were more active than the *respondents*, the level of PA of the intervention group would have been higher if we had been able to collect data from those with *missing* data.

Finally, in the present analyses, we did not investigate the determinants of which participants were adherent to the program. Such analyses, along with the presentation of the results on the step counts, will be the topic of ongoing analyses.

This clinical trial provided results on the PA of participants attending spa treatment. The generalizability of the results to the general population of older adults with NCDs without spa treatment or rehabilitation programs might be limited. Thus, attending a spa treatment or rehabilitation program is proof of interest in one's health.

To our knowledge, this study is the first to combine education during a spa treatment and the use of a web- and mobile-based PA program over a 12-month follow-up. The 3-week stay at a spa resort favors the building of strong relations and exchanges with health care professionals and other patients and has an educational dimension [16,20]. This could help explain why various studies have shown that coaching and information on PA administered during spa therapy produces a lasting benefit on PA [17-20] in the intervention groups and also produces an improvement in the controls [20]. Thus, these findings could partly explain why no large differences in PA were observed among patients receiving information in different forms. Moreover, the periodic follow-up by the interviewers in the 2 groups could also be a potential reason for the increasing motivation to practice PA.

Conclusions

The limitations, especially the impossibility of reaching the required sample size, indicate that it is necessary to interpret the results with caution. Nonetheless, this study demonstrates the potential of a web- and mobile-based PA program associated with an initial face-to-face coaching during a spa treatment to maintain PA in older adults over a 12-month period to achieve PAG and improve quality of life. A spa treatment appears to offer the ideal time and setting to implement education in PA and initiate patients to the use of web- and mobile-based PA programs.

Increasing PA and reducing the excessive sedentariness of inactive patients reduce the risk of NCD aggravation and pain in some nonmalignant chronic conditions, favoring a lasting improvement in personal physical capacity and quality of life.

Acknowledgments

Authors' Contributions

FF designed the study, supervised the trial progress and implementation, wrote the manuscript (*Introduction*, *Methods*, *Results and Discussion* sections), and revised the manuscript. LP participated in the management of logistics and monitoring of data. SP conceived and designed the digital program and implemented the web platform. AM implemented the statistical analysis and revised the manuscript. SR revised the manuscript. CFR revised the manuscript. MD designed the study, supervised the progression of the trial, and revised the manuscript.

Conflicts of Interest

FF was an employee of Biomouv SAS Inc and SP was the chief executive officer of Biomouv SAS Inc, who provided the web and smartphone-based physical activity program. CFR was the president of the Association Française pour la Recherche Thermale Scientific Committee.

Multimedia Appendix 1

Use guide of the website.
[PPT File (Microsoft PowerPoint Presentation), 3915 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Use guide of the mobile app.
[PPT File (Microsoft PowerPoint Presentation), 2890 KB-Multimedia Appendix 2]

Multimedia Appendix 3

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1219 KB-Multimedia Appendix 3]

References

1. Global status report on noncommunicable diseases 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/handle/10665/148114 [accessed 2021-10-05]
2. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018 Oct 5;6:288 [FREE Full text] [doi: 10.3389/fpubh.2018.00288] [Medline: 30345266]
3. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
4. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
5. Étude de santé sur l'environnement, la biosurveillance, l'activité physique et la nutrition (Esteban), 2014-2016. Volet Nutrition. Chapitre Activité physique et sédentarité. 2e édition. Santé publique France. 2020 Feb 13. URL: https://www.santepubliquefrance.fr/import/etude-de-sante-sur-l-environnement-la-biosurveillance-l-activite-physique-et-la-nutrition-esteban-2014-2016.-volet-nutrition.-chapitre-activit [accessed 2021-10-05]
6. Leavitt MO. 2008 Physical Activity Guidelines for Americans. U.S. Department of Health and Human Services. 2008. URL: https://health.gov/sites/default/files/2019-09/paguide.pdf [accessed 2021-10-05]
7. Macera CA, Cavanaugh A, Bellettiere J. State of the art review: physical activity and older adults. Am J Lifestyle Med 2016 Jun 23;11(1):42-57 [FREE Full text] [doi: 10.1177/1559827615571897] [Medline: 30202313]

8. Dorn J, Naughton J, Imamura D, Trevisan M. Correlates of compliance in a randomized exercise trial in myocardial infarction patients. Med Sci Sports Exerc 2001 Jul;33(7):1081-1089. [doi: 10.1097/00005768-200107000-00003] [Medline: 11445753]

9. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]

10. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]

11. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]

12. Kwan RY, Salihu D, Lee PH, Tse M, Cheung DS, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020 Apr 21;17:7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]

13. Boekhout JM, Berendsen BA, Peels DA, Bolman CA, Lechner L. Evaluation of a computer-tailored healthy ageing intervention to promote physical activity among single older adults with a chronic disease. Int J Environ Res Public Health 2018 Feb 15;15(2):346 [FREE Full text] [doi: 10.3390/ijerph15020346] [Medline: 29462862]

14. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]

15. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

16. Gremeaux V, Benaïm C, Poiraudeau S, Hérisson C, Dupeyron A, Coudeyre E. Evaluation of the benefits of low back pain patients' education workshops during spa therapy. Joint Bone Spine 2013 Jan;80(1):82-87. [doi: 10.1016/j.jbspin.2011.12.014] [Medline: 22342470]

17. Kwiatkowski F, Mouret-Reynier MA, Duclos M, Leger-Enreille A, Bridon F, Hahn T, et al. Long term improved quality of life by a 2-week group physical and educational intervention shortly after breast cancer chemotherapy completion. Results of the 'Programme of Accompanying women after breast Cancer treatment completion in Thermal resorts' (PACThe) randomised clinical trial of 251 patients. Eur J Cancer 2013 May;49(7):1530-1538. [doi: 10.1016/j.ejca.2012.12.021] [Medline: 23352440]

18. Gin H, Demeaux JL, Grelaud A, Grolleau A, Droz-Perroteau C, Robinson P, et al. Observation of the long-term effects of lifestyle intervention during balneotherapy in metabolic syndrome. Therapie 2013;68(3):163-167. [doi: 10.2515/therapie/2013025] [Medline: 23886461]

19. Maitre J, Guinhouya B, Darrieutort N, Paillard T. Physical education in a thermal spa resort to maintain an active lifestyle at home: a one-year self-controlled follow-up pilot study. Evid Based Complement Alternat Med 2017;2017:1058419 [FREE Full text] [doi: 10.1155/2017/1058419] [Medline: 28539960]

20. Schnebelen-Berthier C, Negro N, Jaruga A, Roques CF, Lecerf JM. Long term effect of spa therapy combined with patient education program on subjects with overweight and obesity - a controlled study. Obes Res Clin Pract 2019;13(5):492-498. [doi: 10.1016/j.orcp.2019.06.005] [Medline: 31383564]

21. Jones CJ, Rose DJ. International guidelines for training physical activity instructors for older adults. J Aging Phys Act 2004 Jan;12(1):1-2. [doi: 10.1123/japa.12.1.1] [Medline: 15211016]

22. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. Scand J Work Environ Health 1990;16 Suppl 1:55-58 [FREE Full text] [doi: 10.5271/sjweh.1815] [Medline: 2345867]

23. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

24. Ojad P, Tuxworth B. Eurofit for Adults: Assessment of Health-related Fitness. New York, NY, USA: Council of Europe, Committee for the Development of Sport; 1995.

25. Fournier J, Vuillemin A, Le Cren F. Mesure de la condition physique chez les personnes âgées. Évaluation de la condition physique des seniors : adaptation française de la batterie américaine « Senior Fitness Test ». Sci Sports 2012 Sep;27(4):254-259. [doi: 10.1016/j.scispo.2012.07.005]

26. Ware JE, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, MA, USA: The Health Institute, New England Medical Center; 1995.

27. Kolt GS, Schofield GM, Kerse N, Garrett N, Ashton T, Patel A. Healthy Steps trial: pedometer-based advice and physical activity for low-active older adults. Ann Fam Med 2012;10(3):206-212 [FREE Full text] [doi: 10.1370/afm.1345] [Medline: 22585884]

28. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]

XSL•FO

RenderX

29. Wittes J. Sample size calculations for randomized controlled trials. Epidemiol Rev 2002;24(1):39-53. [doi: 10.1093/epirev/24.1.39] [Medline: 12119854]

30. Lu K. On efficiency of constrained longitudinal data analysis versus longitudinal analysis of covariance. Biometrics 2010 Sep;66(3):891-896. [doi: 10.1111/j.1541-0420.2009.01332.x] [Medline: 19764951]

31. Alley SJ, Kolt GS, Duncan MJ, Caperchione CM, Savage TN, Maeder AJ, et al. The effectiveness of a Web 2.0 physical activity intervention in older adults - a randomised controlled trial. Int J Behav Nutr Phys Act 2018 Jan 12;15(1):4 [FREE Full text] [doi: 10.1186/s12966-017-0641-5] [Medline: 29329587]

32. Rowley TW, Lenz EK, Swartz AM, Miller NE, Maeda H, Strath SJ. Efficacy of an individually tailored, Internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]

33. Broekhuizen K, de Gelder J, Wijsman CA, Wijsman LW, Westendorp RG, Verhagen E, et al. An Internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016 Apr 27;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]

34. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013 Feb 05;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]

35. Seo DC, Niu J. Evaluation of Internet-based interventions on waist circumference reduction: a meta-analysis. J Med Internet Res 2015 Jul 21;17(7):e181 [FREE Full text] [doi: 10.2196/jmir.3921] [Medline: 26199208]

36. Melanson EL, Keadle SK, Donnelly JE, Braun B, King NA. Resistance to exercise-induced weight loss: compensatory behavioral adaptations. Med Sci Sports Exerc 2013 Aug;45(8):1600-1609 [FREE Full text] [doi: 10.1249/MSS.0b013e31828ba942] [Medline: 23470300]

37. Blanchard CM, Reid RD, Morrin LI, Beaton LJ, Pipe A, Courneya KS, et al. Barrier self-efficacy and physical activity over a 12-month period in men and women who do and do not attend cardiac rehabilitation. Rehabil Psychol 2007;52(1):65-73 [FREE Full text] [doi: 10.1037/0090-5550.52.1.65]

38. Jenkins LS, Gortner SR. Correlates of self-efficacy expectation and prediction of walking behavior in cardiac surgery elders. Ann Behav Med 1998;20(2):99-103. [doi: 10.1007/BF02884455] [Medline: 9989315]

39. Fisher RJ. Social desirability bias and the validity of indirect questioning. J Consum Res 1993 Sep;20(2):303-315. [doi: 10.1086/209351]

40. Parsons HM. What happened at Hawthorne?: new evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science 1974 Mar 08;183(4128):922-932. [doi: 10.1126/science.183.4128.922] [Medline: 17756742]

41. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

42. 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]

43. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]

44. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]

Abbreviations

AE: adverse event
CLDA: constrained longitudinal data analysis
IPAQ: International Physical Activity Questionnaire
M0: month 0
M2: month 2
M6: month 6
M12: month 12
MET: metabolic equivalent of task
NCD: noncommunicable disease
OR: odds ratio
PA: physical activity
PAG: physical activity guidelines
PCS: physical component subscale
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

SF-12: Short Form Health Survey-12