JMIR Public Health and Surveillance

Impact Factor (2023): 3.5
Volume 7 (2021), Issue 1 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Original Paper

Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study

Stephanie Aboueid[1], MSc, RD; Samantha Meyer[1], PhD; James R Wallace[1], PhD; Shreya Mahajan[1], MSc; Ashok Chaurasia[1], PhD

School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Stephanie Aboueid, MSc, RD
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 966 0530468122
Email: seaboueid@uwaterloo.ca

Abstract

Background: Young adults often browse the internet for self-triage and diagnosis. More sophisticated digital platforms such as symptom checkers have recently become pervasive; however, little is known about their use.

Objective: The aim of this study was to understand young adults' (18-34 years old) perspectives on the use of the Google search engine versus a symptom checker, as well as to identify the barriers and enablers for using a symptom checker for self-triage and self-diagnosis.

Methods: A qualitative descriptive case study research design was used. Semistructured interviews were conducted with 24 young adults enrolled in a university in Ontario, Canada. All participants were given a clinical vignette and were asked to use a symptom checker (WebMD Symptom Checker or Babylon Health) while thinking out loud, and were asked questions regarding their experience. Interviews were audio-recorded, transcribed, and imported into the NVivo software program. Inductive thematic analysis was conducted independently by two researchers.

Results: Using the Google search engine was perceived to be faster and more customizable (ie, ability to enter symptoms freely in the search engine) than a symptom checker; however, a symptom checker was perceived to be useful for a more personalized assessment. After having used a symptom checker, most of the participants believed that the platform needed improvement in the areas of accuracy, security and privacy, and medical jargon used. Given these limitations, most participants believed that symptom checkers could be more useful for self-triage than for self-diagnosis. Interestingly, more than half of the participants were not aware of symptom checkers prior to this study and most believed that this lack of awareness about the existence of symptom checkers hindered their use.

Conclusions: Awareness related to the existence of symptom checkers and their integration into the health care system are required to maximize benefits related to these platforms. Addressing the barriers identified in this study is likely to increase the acceptance and use of symptom checkers by young adults.

(JMIR Public Health Surveill 2021;7(1):e22637) doi:10.2196/22637

KEYWORDS

self-assessment; symptom checkers; self-triage; self-diagnosis; young adults; digital platforms; internet; user experience; Google search

Introduction

Seeking online health information through search engines is common [1,2]; however, it can have negative effects on individuals due to the lack of reliable information and lack of health literacy or expertise of those seeking health information [3,4]. In addition to information overload, navigating the internet can be problematic due to the use of overly technical language as well as the high volume of irrelevant content returned from search engine results, the confusing layout of many web pages, and the lack of quality requirements for publishing online content [5,6]. These limitations coupled with a shortage in the health workforce globally [7] have led to the development of novel digital platforms that allow users to self-assess their symptoms. Using a question-and-answer chat format, symptom checkers prompt users to enter their symptoms and health information based on the level of care required, and a potential list of diagnoses is provided [8]. These functions allow users to self-triage (ie, assess whether they should seek medical services based on the severity of symptoms) or self-diagnose (ie, identify a health condition in oneself). Despite the mounting number of symptom checkers available and the adoption of this technology by various credible health institutions and entities such as the UK National Health Service (NHS) and the government of Australia [9,10], knowledge surrounding this technology is limited [11]. The scarce literature on symptom checker accuracy suggests that the quality of diagnostic and triage advice differs based on the digital platform used [12] with those enabled by artificial intelligence having a higher percentage of listing the correct diagnosis first [13].

In assessing symptom checker benefits, it is important to understand user perspectives on the platform after actual use rather than simply assessing their accuracy in fictitious situations [14,15]. A study that examined patients' experiences using the symptom checker "Isabel" found that the platform was most commonly used to better understand the causes of symptoms, followed by deciding whether or not to seek care [14]. Most of the patients in that study (274/304, 90.1%) reported receiving useful information for their health problems and reported that they would use the symptom checker again (278/304, 91.4%) [14]. These findings are in line with another study that examined perspectives on use of the "Ada" symptom checker, which showed that most of the participants (443/503, 88.1%) would recommend the platform to a relative [16]. An important factor that seems to influence symptom checker acceptance is age, with younger populations exhibiting higher acceptance [14,15]. A UK-based study that engaged 1071 patients found that more than 70% of individuals between the ages of 18 and 39 years would use a symptom checker as compared to only 51% aged between 55 and 69 years [15].

A common theme that emerged from most studies examining symptom checkers is the importance of gathering user perspectives on the use of the platform to enable high acceptance (and use) as well as to prevent lost investments [17]. Given that young adults (between 18 and 34 years of age) may be the user group most accepting of such technology—and thus the ideal target group—we sought to maximize acceptance and use in this population by understanding the factors that would enable or hinder its use for self-triage and self-diagnosis. Given the relatively new emergence of symptom checkers and the prevalent use of the Google search engine (Dr. Google) for assessing symptom severity, the aim of this study was to gather university students' perspectives after having used a symptom checker on (i) using the internet's search engine versus a symptom checker for self-triage and diagnosis, and (ii) the enablers and barriers associated with using a symptom checker for self-triage and diagnosis.

Methods

Study Design and Aims

This qualitative analysis represents a subset of findings that emerged from a larger mixed-methods study that seeks to understand the factors associated with the behavioral intention of using symptom checkers for self-triage and self-diagnosis. A qualitative descriptive case study research design was used and is differentiated by other research study designs by its focus on a bounded system or case [18]. In this work, the behavioral intention of using symptom checkers is the phenomenon of interest; this phenomenon is bounded by the university campus and the selection of university students as participants. Although three notable researchers have previously described case study research, this work was conducted in line with Yin's interpretation, which focuses on methodology and adopts a postpositivist worldview [19-21]. This work is positioned in the postpositivist paradigm due to the use of theory, the collection of data to either support or refute this theory, and the changes and revisions made to the theory as findings emerged.

Recruitment

To allow for a broad range of perspectives to be gathered, university students between the ages of 18 and 34 years across faculties in all levels of education and year of study were eligible to participate. Following ethics approval from the Research Ethics Board at the University of Waterloo (41366), university students were notified of the study through emails from the administrative assistant of their faculty; as such, the number of students who received and opened the email is unknown. Interested individuals were asked to contact the principal investigator (SA) to schedule an interview. Participants were recruited between November 2019 and May 2020. A total of 24 participants were included in the study based on a first-come, first-served basis and time of data saturation. There were no dropouts in this study. All participants were provided with an information letter prior to the interview outlining the study objectives. Informed consent was obtained from all participants. One-on-one interviews took place on the university campus or virtually through a digital university-approved platform, Whereby. All participants were provided with a Can $10 (US $7.80) coffee shop gift card as a token of appreciation for taking the time to participate in the study.

Data Collection

The main sources of data were a preinterview questionnaire (Multimedia Appendix 1), semistructured interview protocol (Multimedia Appendix 2), protocol for the think-aloud exercise, and clinical vignette (Multimedia Appendix 3). The semistructured interview method was used because it offers flexibility to the interviewer in determining when it is appropriate to explore certain subjects in greater depth or to pose new questions that were not originally anticipated when the interview protocol was developed [22]. To provide contextual information on each participant, the preinterview questionnaire was comprised of questions related to demographics such as age and gender as well as self-perceived health [23] and four dimensions of health literacy [24], which are two validated tools that may influence participants' perspectives on the use of symptom checkers. Self-perceived health was measured using one question that has been previously validated [23]. Four dimensions (ie, feeling understood by health care providers, actively managing my health, ability to actively engage with health care providers, and ability to find good health information) from a total of nine of the Health Literacy Questionnaire (HLQ) questions developed by Osborne et al [24] were used; this approach has been permitted by the original authors and used in practice to reduce respondent burden. Two HLQ domains assessed in this study were measured using a 4-point Likert scale, whereas the other two were measured using a 5-point scale that ranges from "cannot or always difficult" to "always easy"; a higher number indicates a higher level of agreeableness and higher level of health literacy.

To ensure that all participants were familiar with symptom checkers, they were provided with a clinical vignette and, based on a draw, were asked to use one of two web-based symptom checkers: WebMD [25] or Babylon Health [26]. These two symptom checkers were chosen based on popularity and adoption by credible institutions such as the UK NHS, respectively. Both platforms are similar in terms of their objectives and process (eg, they both allow users to enter symptoms as free text and suggest symptoms from a drop-down list); however, there are key differences, including that Babylon Health requires the user's full name, email address, country of residence, and date of birth. Moreover, since Babylon Health probes for more information, it may take longer to complete.

After having read the clinical vignette and accessed the symptom checker, participants were guided by the first author (SA) to conduct the think-aloud exercise, which involved the participants thinking out loud while they performed a task without synthesizing or interpreting their thoughts [27]. Similar to another study [28], the clinical vignette used depicted symptoms of a disease (ie, scarlet fever) that is less common in young adults to avoid having participants rely on recent experiences during the exercise. The questions in the interview protocol were designed to answer the main objectives of this study, which included understanding how the use of symptom checkers is perceived as compared to using the Google search engine and the factors that facilitate or hinder the use of symptom checkers. The first author (SA) conducted all interviews. SA holds a Master of Science in health systems; is trained in qualitative research methods, including data collection and analysis; and was a PhD candidate in Public Health and Health Systems at the time of the study. Given that data collection and analysis were occurring concomitantly, it was possible to cease recruitment once data saturation was reached (ie, collecting more data would not reveal new information) [29], which occurred after the interview with the 20th participant.

Data Analysis

Data analysis was conducted independently by two authors (SA and SM) using the thematic analysis steps outlined by Castleberry and Nolen [30], which consist of compiling, disassembling, reassembling, and interpreting the data. The first step of compiling consisted of importing all transcribed interviews into the NVivo software program (version 12.6.0). To get a sense of the data as a whole, all transcripts were read in their entirety. To disassemble the data, a line-by-line coding approach was used to reduce the superimposition of preconceived notions on the data. This step generated descriptive codes [29], which were then used as a tag to retrieve and categorize similar data. Given the limited literature on this topic, the coding process was highly inductive, and a codebook was developed throughout the coding process. The codebook contains all generated codes with an indication of when they should be used.

The third step of reassembling consisted of grouping the codes into main themes; in NVivo, this consists of creating nodes (themes) and child nodes (codes under those themes)—the hierarchy can contain many levels depending on the level of detail required. In this work, a hierarchy was used to represent how themes are subordinate or superordinate to each other [30]. The final step of the analysis consisted of interpreting the analyzed data as they related to the study's overarching aim and objectives. By interpreting the data at a higher level than themes, it was possible to answer the research question and objectives. Throughout the coding process, SA and SM discussed the identified themes and resolved any discrepancies.

Results

Participant Information

Most of the participants had a high score on the four health domains measured (see Table 1) with the exception of two, one, and two participants who had a low score on the following health domains: feeling understood by health care providers, actively managing my health, and ability to find good health information, respectively. The think-aloud exercise took approximately 10 to 15 minutes to complete, with those who had used the Babylon Health platform taking a longer time to complete the task due to the higher number of questions asked. A total of 11 participants were familiar with symptom checkers prior to the interview, 2 of whom had used a symptom checker for the first time to assess COVID-19–related symptoms. Participants who had previously used a symptom checker learned about the platform through word of mouth or a Google search.

Table 1. Participant information (N=24).

Characteristics	Value
Gender, n (%)	
Female	14 (58)
Male	9 (38)
Nonbinary	1 (4)
Racial group, n (%)	
White	9 (38)
Asian	6 (25)
Chinese	3 (13)
Arab	2 (8)
Indian	2 (8)
Black	2 (8)
Highest level of education, n (%)	
High school	2 (8)
Undergraduate degree	14 (58)
Master’s degree	8 (33)
Faculty, n (%)	
Engineering	8 (33)
Sciences	6 (25)
Applied health sciences	3 (13)
Environment	3 (13)
Arts	3 (13)
Mathematics	1 (4)
Self-perceived health, n (%)	
Excellent	2 (8)
Very good	13 (54)
Good	5 (21)
Fair	4 (17)
Poor	0 (0)
Health literacy, mean (range)	
Feeling understood by health care providers[a]	2.92 (1.5-4.0)
Actively managing my health[a]	3.05 (1.8-4.0)
Ability to actively engage with health care providers[b]	3.64 (2.6-4.6)
Ability to find good health information[b]	3.81 (1.8-5.0)
Symptom checker used, n (%)	
WebMD	11 (46)
Babylon Health	13 (54)

[a]Maximum possible average is 4.

[b]Maximum possible average is 5.

Themes Related to Using a Google Search Engine Versus a Symptom Checker

Theme Classification

Data related to the use of a search engine or symptom checker were grouped into positive or negative themes. Positive themes suggest a desirable attribute, function, or experience related to a platform, whereas negative themes encompass themes that suggest the opposite. An overview of these themes is provided in Textbox 1. These themes were further grouped into four main themes, which are supported by participant quotes. For example, the description under the subsection "Symptom Severity and Input" includes the main findings related to themes—both positive and negative—that pertain to how symptoms may influence participant perspectives related to the Google search engine and symptom checkers.

Textbox 1. Overview of themes related to using a Google search engine vs a symptom checker.

Positive themes

Google search engine

- Provides information without claiming a diagnosis
- More customizable
- Allows entry of all symptoms in the search engine

Symptom checkers

- More personalized
- More interactive due to chatbot feature
- Good for those who do not know how to use Google
- Straightforward design
- Easy to use
- Real-time output
- Makes the correlation between symptoms and potential conditions
- More intuitive
- More reliable
- More specific
- More structured

Negative themes

Google search engine

- Absence of chatbot feature

Symptom checkers

- Accuracy is questionable
- Limits the number of symptoms that can be inputted
- Not widely known
- Thought process of the platform is unclear
- User more vulnerable when using this platform

Both Google search engine and symptom checkers

- Text input is insufficient
- Suboptimal reliability

Symptom Severity and Input

Participants perceived the Google search engine and symptom checkers to be useful for mild symptoms; however, some perceived that using the Google search engine was faster than having to answer questions in a symptom checker: "If you're Googling something quick then it's easy, quick, and straightforward, you don't have to take 10 minutes to answer all these questions…" [P2].

Positive themes related to the use of the internet mainly pertained to the perspective that a Google search engine allows users to input as many symptoms as needed, enabling a more comprehensive search of potential conditions that may be relevant to their health context. Some users also mentioned that they prefer that the platform does not claim that this is the condition they may have.

> *On the other hand, I think it may be easier to get accurate results on symptoms through a Google search because I can type multiple symptoms and see how they fit, I may get more garbage results but I can use my judgment to decide what is true and not true. Whereas the symptom checker has only one piece of information which is fever. The symptom checker did not give me the opportunity to put in more from what I can recall.* [P11]

Nonetheless, some participants mentioned that the absence of a chatbot feature in the Google search engine limits the platform's ability to ask follow-up questions based on symptoms inputted. As such, some users who many not be able to identify all symptoms experienced may omit certain symptoms or may not elaborate on symptoms, which hinders the quality and comprehensiveness of results.

> *…if I were to Google my symptom, I would just put in a fever and rashes that could be a million things. But with a symptom checker, I would put in fever and it asked me for a specific temperature and other specific questions which I would not know to search on my own.* [P21]

Perceived Characteristics of Symptom Checkers

Symptom checkers were perceived by some participants to be a good option for individuals who are less proficient in using the internet for information retrieval. Some had a positive attitude toward the symptom checkers because the platform asked questions regarding age and gender, giving the impression that it is more personalized and in turn, in their perception, more accurate.

> *…surfing through the internet and coming through a particular diagnosis takes a lot of time although it might give you more information about other diseases that have similar symptoms, but this is not what I am looking for, I am looking for what I am suffering from. So, for which, I think a personalized software is helpful.* [P2]

Some participants believed that the symptom checker "had more structure," "provided a greater level of detail," "was more interactive," and "was more reliable" than using the Google search engine.

> *So I think having that more structured approach to inputting symptoms and figuring out what is likely wrong with you would be a lot nicer for the user and the user would have more faith in the result rather than just going on Google that brings up a whole bunch of results and the user thinking that they could have anything.* [P4]

Symptom Checker Limitations

Although having a more structured approach to symptom input was favored, some participants were unable to enter all symptoms in the platform, which led them to question the accuracy and reliability of the platform; this also hindered trust toward the platform.

> *I feel like I don't like the symptom checker as much because it limits the number of symptoms. I did not have the chance to mention the thing with the red bumps; it just asked me a lot of questions about the one "symptom that was bothering me the most."* [P20]

There was also a sense that participants would feel more vulnerable using a symptom checker due to the more personalized nature of the questions asked. Interestingly, some participants believed that their judgment and thought process to identify potential diagnoses was superior than using a symptom checker due to lack of knowledge about how symptom checkers work.

> *It feels more vulnerable and personal to put my symptoms into a list or generator of some kind. It feels like I am just looking at a series of articles I feel there's more of a distance…. If I am typing in a symptom checker and it comes back at me with answers, I don't know how it came to that conclusion and I don't know what the process was to decide that "yes, this is what you have," whereas if I am the one doing the analysis through a bunch of articles that I deem legitimate—whether or not they truly are legitimate—at least I know what the thought process was and I feel like I can trust that.* [P6]

Despite various shortcomings that were mentioned related to the use of symptom checkers, some participants believed that an important issue is the lack of awareness about the existence of the platform: "But the issue is that we don't know about symptom checkers so making them widely available would be super helpful." [P13]

Accessing Health Services as a Preferred Option

In addition, there was a consensus that consulting a primary care provider or nurse was superior than searching the internet or using a symptom checker to assess the severity of symptoms; this was especially the case when certain symptoms required a physical examination and text input was insufficient. Reliability of the Google search engine and symptom checker was also questionable and was perceived negatively by some participants.

> *I think seeing a provider face to face is better than both options. I feel that you can't accurately portray all your symptoms and general health by text input. You need someone looking at you and take measurements and touch injured areas, I think that's far superior.* [P1]
>
> *I think Google is a very wide platform so it's very hard to analyze the reliability or the source. In this case, it depends on the reliability of the symptom checker as well.* [P19]

Themes Related to Barriers and Enablers of Using Symptom Checkers

Classification of Themes

Factors that would hinder the use of a symptom checker were identified as barriers, whereas factors that would enable an individual to use a symptom checker were identified as enablers. Participants enumerated many enablers and barriers for using symptom checkers, which were mainly related to the (1) individual, (2) disease, (3) health care system, or (4) symptom checker. An overview of all identified barriers and enablers is provided in Textbox 2. Example quotes for barriers and enablers are provided in Multimedia Appendix 4.

Textbox 2. Overview of enablers and barriers for using symptom checkers.

Individual-level factors

Enablers:

- Internet access
- Low health literacy
- Trust in the platform
- Younger age
- Lack of time
- Convenience
- Lack of trust in doctors
- Curiosity
- Embarrassing topic
- Increase empowerment
- Aversion to medical professionals
- Having pre-existing conditions
- Unable to discuss the topic with a health provider
- Uncertain about care required
- Worried about health of oneself

Barriers:

- Lack of internet access
- Low health literacy
- Lack of trust in the platform
- Low technology literacy
- Older age
- Social influence
- Not wanting to know
- Previous bad experience

Disease-level factors

Enablers:

- Mild symptoms
- A “broad category of illness”
- Symptoms can be easily described

Barriers:

- Severe condition
- Need for a physical examination

Health system–level factors

Enablers:

- Approved by doctors
- Lack of access to health services
- Cost of health services
- Public education
- Increased awareness

- Long wait times for health services
- Reputable organizations recommend it

Barriers:

- Authoritarianism in health care

Characteristics of symptom checkers

Enablers:

- Increased advertisement
- Easy interface
- Data privacy
- Free of charge
- Short to complete
- Precision
- Use of artificial intelligence
- Gamification
- Good source of information
- Integrated with an electronic health record
- Useful in identifying potential conditions
- Information about the creators of the platform
- Interactive platform
- Reliability

Barriers:

- Lack of awareness
- Poor design
- Asking identifiable questions
- Cost of the platform
- Time to complete
- Lack of inclusivity measures
- Lack of language options
- Lack of credibility
- Lack of human interaction
- Disclaimer
- Inability to obtain elaboration on a question
- Liability
- Concerns about using data for profits

Individual-Level Factors

Internet access, health literacy, trust toward the platform, technology literacy, and age were factors mentioned to be either enablers or barriers for using a symptom checker for self-triage and self-diagnosis. Younger age, internet access, high technology literacy, and trust toward the platform were perceived to enable the use of symptom checkers. Low health literacy was perceived to be both an enabler and barrier for using a symptom checker. Although some participants believed that individuals with low health literacy are more likely to use a symptom checker because they may be less critical, others perceived low health literacy to be a barrier due to inability to understand and input symptoms into the platform.

> *…maybe if they just did not know a ton about health in general maybe they would be less critical than me.* [P11]
>
> *Also, sometimes it's hard to articulate to have the proper term of how you feel. For example, in the fever*

> *or the lymph node, you don't know of things like that unless you have specific knowledge about it. So, it is hard for someone who does not have medical terminology to input what they have in there.* [P13]

Disease-Level Factors

Given that a disease does not define the individual, disease-related factors were considered separately. Having a "broad category of illness," an embarrassing issue, or an issue perceived to be mild were perceived to be enablers, whereas experiencing severe symptoms and needing a physical examination were mentioned to be barriers. Participants seemed to be more willing to use a symptom checker if they were experiencing nonspecific symptoms (eg, fatigue) due to the perceived notion that a symptom checker would allow the user to narrow down on a health condition. They are also more willing to use symptom checkers for issues perceived to be "embarrassing" such as conditions related to mental health.

> *If something was serious, people would not want to use it, they would want to go to a doctor. Not just physically but also emotionally, I could see them go to the doctor right away.* [P13]

Health System–Level Factors

Lack of access to and cost of health services were perceived to enable the use of a symptom checker. Having the platform approved by reputable organizations and approved by doctors were also mentioned to be important factors for enabling individuals to use the platform. In contrast, some believed that some primary care providers may not be accepting of the technology, thus limiting its use.

> *First of all, they have to somehow not only advertise but maybe if the website is promoted by the health care organization that is reliable for people then I can make sure that the platform is trusted by an authentic organization so for sure I would use it.* [P24]

An important factor related to the symptom checker was advertisement. More than half of participants were not aware of symptom checkers, thus limiting their use: "But the issue is that we don't know about symptom checkers so making them widely available would be super helpful." [P13]

Characteristics of Symptom Checkers

Developing an easy interface, guaranteeing data privacy, offering the platform free of charge, and ensuring that the platform's questionnaire is short to complete were all mentioned as potential enablers for using a symptom checker.

> *It sounds very interesting and it is very easy to use. Definitely I will use it again, I had a good user experience.* [P24]

> *And if it was short—I think if there were options "hey, do you want to take the shorter version and it might not be as accurate or do you want to take the longer one that will take more time but will be more accurate." I think people want something quick but quick won't be as accurate.* [P9]

The data suggest that the main barriers for using symptom checkers are the lack of transparency on how the data collected are used; some participants mentioned that they would not have an issue with the data being used by governmental institutions to improve health services but did not want their data to be used to generate profits. Although most of the participants understood the medical terms that were used by the digital platform, some believed that the average person may not understand some of the questions asked. Providing a brief description of medical terms would allow users to interact with the platform in a more informed manner.

> *I would not want my data to be used to anything that would harm me. I don't know what it could be used for but if it is being used to find out the prevalence of a certain disease or whatever that is helpful for the health care system, I am fine with that but anything that would encourage the business part of it or pharmaceutical side of it or anything that is business related or goes back to making money, I would not like it.* [P3]

Participants also stressed the importance for the digital platform to elaborate on why certain questions were being asked. In contrast to seeing a health professional, users are unable to interject and ask the platform questions for further elaboration. Moreover, most platforms use a disclaimer that they do not provide medical advice, which undermines the platform's credibility.

> *If they know not to take it seriously, they won't feel encouraged to do the test at all. If the disclaimer says this is not really a diagnosis, then what am I doing? I should just go to the doctor.* [P10]

The platforms that were used during this study were in English; however, some mentioned the importance of having these platforms available in various languages to ensure that they are accessible to those who are less proficient in English. Lack of inclusivity measures does not allow persons with disabilities to use the platform and was also mentioned as a barrier to use: "…or various disabilities being able to use the screen or use computers or any type of access issues would be a problem." [P6]

Discussion

Principal Results

Approximately half of the participants were not aware about the existence of symptom checkers until their participation in this study. Most of the participants preferred consulting a health professional to address their health needs rather than researching using the internet's search engine or using a symptom checker. Nonetheless, symptom checkers were generally preferred over the internet's search engine due to their personalized approach; however, some perceived that the latter is faster to use than having to answer questions. There was also an acknowledgment that the results provided by symptom checkers can only be as good as the data that informed them.

In sum, it was perceived that individuals who are younger, have low health literacy, and high technology literacy were more likely to use a symptom checker. Lack of time, convenience, having symptoms that were perceived to be minor or "embarrassing" were other factors that would result in the use of a symptom checker. Enablers that were related to the health system were lack of access to care, having the technology approved by credible associations, and having symptom checkers integrated as part of the public health system. Nonetheless, participants mentioned many improvements that would have to be made to the symptom checker to enable its use, including improving accuracy, ensuring that the platform is freely accessible, and ensuring privacy and anonymity. Although the participants appreciated the personalized approach of the platform, they would not want to use a symptom checker that asks too many questions or that takes too long to complete. Barriers for using symptom checkers included lack of access to the internet, medical jargon, and lack of trust. Despite the barriers and shortcomings of the platform, participants believed that symptom checkers would be useful if they were tested and validated. Some believed that these platforms must have a positive influence on their health due to the perception that these platforms are designed by medical doctors.

Limitations

Although this study has some strengths, various limitations warrant mention. First, given that all participants had previously used the Google search engine for seeking health information, we did not ask participants to use the Google search engine during the interview, meaning that they had to rely on their previous Google search experiences to answer questions. Second, we asked the participants if they had previously used a symptom checker, but we did not ask about the frequency of use, which limited our ability to assess whether responses differed based on this potentially important factor. Third, we did not distinguish responses based on the digital platform used, as the main focus of this work was to understand perspectives on the use of symptom checkers in general; however, these perspectives may have differed if participants used another symptom checker than those used in this study (ie, WebMD or Babylon Health). In line with this, we did not examine whether participants chose the correct diagnosis based on the clinical vignette as the focus of the study was on the process of obtaining the list of diagnoses and getting participants familiar with the platform. Last but not least, the sample was comprised of highly educated individuals who were perceived to have a good health status; as such, this may limit the transferability of findings to other populations. Future studies should explore perspectives of other user groups.

Comparison With Prior Work

Findings from this study are in line with the literature, which suggests that using the Google search engine for health information has many limitations, including the vast amount of information available and lack of quality requirements for publishing content [5,6]. Many of these limitations could be addressed by symptom checkers as these platforms are typically developed through a collaboration between developers and medical experts. As found in other studies, symptom checkers were perceived to be a useful tool for self-assessing the severity of symptoms [14,15]; however, this was mostly the case for symptoms that are perceived to be mild and for which text input is perceived to be sufficient. Most of the participants in this study were more accepting of using symptom checkers for self-triage than for self-diagnosis. Moreover, consulting a primary care provider was the favored option over using the Google search engine or a symptom checker. This finding is in line with results from a Canadian national study, which showed that while the public supports investments in artificial intelligence and technology, they do not want to see these investments occur at the expense of the health workforce [31].

This study highlights various factors associated with the use of symptom checkers that could be used as a starting point for future investigations studying the acceptance of such technology in other population groups. Lack of time and convenience were important enabling factors for using symptom checkers; these factors also explain the use of the internet's search engine for health information [32]. An important factor that seemed to hinder the perceived credibility of symptom checkers is that most of these platforms include a disclaimer that they are not providing medical advice. Although there are legitimate and legal reasons related to this practice, it may make people wonder why they would spend time using the platform in the first place. Ensuring that health professionals are working in conjunction with the platform has been proposed previously and may be an approach to address this issue [15].

Importantly, despite the participants reporting positive aspects of the question-answer format used by the symptom checker, most would have favored a more interactive platform that provides more information to the user regarding why certain questions are being asked. Participants also mentioned the importance of being able to ask questions to the platform—this is something that could be easily done during a conversation with a medical provider or a Google search; however, given the more rigid nature of most symptom checkers, this feature is not yet readily available. This is important for symptom checker developers to consider as patient-centered communication has been shown to be important for patient outcomes [33]. Moreover, participants mentioned that they would be more trusting of this platform if it provided them with a diagnosis that they thought they had, which is in line with another study [34], indicating the presence of confirmation bias.

Lack of internet access is also a critical element that hinders access to any web-based information platform. Although lack of internet access is more prevalent in developing countries [35], it remains an issue for certain remote and rural regions in Canada. However, efforts to address this issue have been outlined in the 2019 Canadian budget in which the government announced its commitment of reaching a target of 95% of Canadian homes and businesses having access to internet speeds of at least 50/10 Mbps (50 Mbps downloading speed and 10 Mbps uploading speed) by 2026 and 100% access by 2030 [36]. Interestingly, the lack of connectivity is related to poor literacy and digital skills rather than to lack of affordability [35]. This highlights the importance of ensuring that populations have the means of accessing these platforms. Failing to do so will undermine the purpose and mission of many of these technology companies that aim to reach those that are disadvantaged and living in developing countries [37]. Participants also mentioned that lack of human interaction may be a potential barrier for older adults, but they did not believe it to be a barrier for them. The importance of integrating human support in technology has been recommended for improving adherence, communication with care teams, and improving the quality of tool use [38].

In line with perspectives from experts in the field [39], symptom checkers have the potential to improve quality in health care; however, various barriers should be addressed to improve acceptance and use of the platform by end users. Given the wide array of factors elucidated in this study, future studies should focus on understanding the relative importance of these factors as they relate to the acceptance and use of symptom checkers. For example, participants can be asked to rank the barriers and enablers for using symptom checkers in order of their perceived importance. Importantly, to ensure client-centric product development, companies or governmental institutions developing these platforms should include end users in the process. Similarly, seeking health care provider perspectives should be prioritized to inform how symptom checkers should be utilized by the health care system to maximize its benefits while ensuring that they meet user needs.

Conclusions

Symptom checkers are promising tools and seem to be more accepted for self-triage rather than for self-diagnosis. To maximize acceptance and use among young adults, it is important to address the various barriers identified in this study, including those that seek to improve the user experience. Importantly, awareness related to the existence of symptom checkers and their integration into the health care system are required to maximize the benefits related to these platforms. Future studies targeting other group segments are needed to understand perspectives of symptom checker use among the wider population.

Acknowledgments

The authors would like to thank the university students who agreed to participate in the study as well as the administration staff at the University of Waterloo for aiding with participant recruitment.

Authors' Contributions

SA conceptualized the study and collected the data. SA and SM analyzed the data. SM, JW, and AC contributed to the interpretation of data. SA wrote the initial draft of the manuscript. SM, JW, and AC reviewed the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preinterview questionnaire.
[DOCX File , 17 KB - publichealth_v7i1e22637_app1.docx]

Multimedia Appendix 2
Interview questionnaire.
[DOCX File , 15 KB - publichealth_v7i1e22637_app2.docx]

Multimedia Appendix 3
Think-aloud exercise protocol and clinical vignette.
[DOCX File , 13 KB - publichealth_v7i1e22637_app3.docx]

Multimedia Appendix 4
Example quotes of enablers and barriers for using symptom checkers.
[DOCX File , 22 KB - publichealth_v7i1e22637_app4.docx]

References

1. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: A literature review. Health Informatics J 2016 Dec 26;22(4):992-1016 [FREE Full text] [doi: 10.1177/1460458215602939] [Medline: 26377952]
2. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
3. White RE, Horvitz E. Web to world: predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. AMIA Annu Symp Proc 2010 Nov 13;2010:882-886 [FREE Full text] [Medline: 21347105]

4. Karnam S, Raghavendra P. Hybrid Doctors: The Need Risen From Informed Patients. J Clin Diagn Res 2017 Feb;11(2):ZI01-ZI04 [FREE Full text] [doi: 10.7860/JCDR/2017/23163.9200] [Medline: 28384991]
5. Cline R, Haynes K. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec 16;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
6. Metzger MJ, Flanagin AJ. Using Web 2.0 technologies to enhance evidence-based medical information. J Health Commun 2011 Jul 29;16(Suppl 1):45-58. [doi: 10.1080/10810730.2011.589881] [Medline: 21843095]
7. Addressing the 18 million health worker shortfall - 35 concrete actions and 6 key messages. World Health Organization. 2019 May 18. URL: https://www.who.int/hrh/news/2019/addressing-18million-hw-shortfall-6-key-messages/en/ [accessed 2020-01-31]
8. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
9. Our NHS Services. Babylon GP at hand. 2019. URL: https://www.gpathand.nhs.uk/our-nhs-service [accessed 2020-01-05]
10. Healthdirect Symptom Checker. Government of Australia. 2019. URL: https://www.healthdirect.gov.au/symptom-checker [accessed 2020-01-06]
11. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The Use of Artificially Intelligent Self-Diagnosing Digital Platforms by the General Public: Scoping Review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
12. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269. [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
13. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun 11;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
14. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient Perspectives on the Usefulness of an Artificial Intelligence-Assisted Symptom Checker: Cross-Sectional Survey Study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
15. Using Technology to Ease The Burden on Primary Care. Healthwatch Enfield. 2019. URL: https://www.healthwatch.co.uk/reports-library/using-technology-ease-burden-primary-care [accessed 2020-06-23]
16. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
17. Venkatesh V, Thong J, Xu X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. J Assoc Inf Syst 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
18. Creswell J. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, CA: Sage Publications; 2014.
19. Merriam S. Qualitative research and case study applications in education. San Francisco: Jossey-Bass; 1998.
20. Stake R. The art of case study research. Thousand Oaks, CA: Sage Publications; 1995.
21. Yin RK. Applications of case study research. Thousand Oaks, CA: Sage Publications; 2012.
22. Patton M. Qualitative research and evaluation methods. Thousand Oaks, CA: Sage Publications; 2002.
23. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. A meta-analysis. J Gen Intern Med 2006 Mar;21(3):267-275 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00291.x] [Medline: 16336622]
24. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13(1):658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
25. WebMD Symptom Checker. WebMD. URL: https://symptoms.webmd.com/ [accessed 2020-12-12]
26. Babylon Chat. Babylon. URL: https://www.babylonhealth.com/ask-babylon-chat [accessed 2020-12-12]
27. Ericsson KH. Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press; 1993.
28. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
29. Charmaz K. Constructing grounded theory. Thousand Oaks, CA: Sage Publications; 2006.
30. Castleberry A, Nolen A. Thematic analysis of qualitative research data: Is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
31. Shaping the Future of Health and Medicine. Canadian Medical Association. 2018 Aug 14. URL: https://www.cma.ca/sites/default/files/pdf/Media-Releases/Shaping%20the%20Future%20of%20Health%20and%20Medicine.pdf [accessed 2020-05-13]
32. Osei Asibey B, Agyemang S, Boakye Dankwah A. The Internet Use for Health Information Seeking among Ghanaian University Students: A Cross-Sectional Study. Int J Telemed Appl 2017 Dec 06;2017(2):1-9. [doi: 10.1155/2017/1756473]
33. King AR, Hoppe RB. "Best practice" for patient-centered communication: a narrative review. J Grad Med Educ 2013 Sep;5(3):385-393 [FREE Full text] [doi: 10.4300/JGME-D-13-00072.1] [Medline: 24404300]

34. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
35. United NE, ScientificCultural O. New report on global broadband access underscores urgent need to reach the half of the world still unconnected. United Nations Educational, Scientific and Cultural Organization. 2019 Sep 23. URL: https://en.unesco.org/news/new-report-global-broadband-access-underscores-urgent-need-reach-half-world-still-unconnected [accessed 2020-01-07]
36. Investing in the Middle Class: Budget 2019. Government of Canada. 2019 Mar 19. URL: https://www.budget.gc.ca/2019/docs/plan/budget-2019-en.pdf [accessed 2020-01-07]
37. Morita T, Rahman A, Hasegawa T, Ozaki A, Tanimoto T. The Potential Possibility of Symptom Checker. Int J Health Policy Manag 2017 Oct 01;6(10):615-616 [FREE Full text] [doi: 10.15171/ijhpm.2017.41] [Medline: 28949479]
38. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: Barriers to the use of apps for depression. Digit Health 2017 Jun 08;3:2055207617713827 [FREE Full text] [doi: 10.1177/2055207617713827] [Medline: 29942605]
39. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Nov 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]

Abbreviations

HLQ: Health Literacy Questionnaire
NHS: National Health Service

Edited by T Sanchez; submitted 03.08.20; peer-reviewed by K Morse, A Radomski; comments to author 09.10.20; revised version received 26.10.20; accepted 27.11.20; published 06.01.21.

Please cite as:
Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A
Young Adults' Perspectives on the Use of Symptom Checkers for Self-Triage and Self-Diagnosis: Qualitative Study
JMIR Public Health Surveill 2021;7(1):e22637
URL: https://publichealth.jmir.org/2021/1/e22637
doi:10.2196/22637
PMID:33404515

Original Paper

Prevalence and Predictors of Health-Related Internet and Digital Device Use in a Sample of South Asian Adults in Edmonton, Alberta, Canada: Results From a 2014 Community-Based Survey

Mark J Makowsky[1], BSP, PharmD; Charlotte A Jones[2], FRCP(C), PhD, MD; Shahnaz Davachi[3], RNutr, RD, PhD

[1]Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada

[2]Faculty of Medicine, Southern Medical Program, University of British Columbia Okanagan Campus, Kelowna, BC, Canada

[3]Primary Health Care, Alberta Health Services, Calgary, AB, Canada

Corresponding Author:
Mark J Makowsky, BSP, PharmD
Faculty of Pharmacy and Pharmaceutical Sciences
University of Alberta
2-142E Katz Group - Rexall Centre for Pharmacy & Health Research
11361 87 Avenue
Edmonton, AB, T6G 2E1
Canada
Phone: 1 7804921735
Fax: 1 7804921217
Email: makowsky@ualberta.ca

Abstract

Background: South Asian Canadians are at high risk of developing cardiovascular disease and diabetes. Consumer-oriented health information technology may help mitigate lifestyle risk factors and improve chronic disease self-management.

Objective: This study aims to explore the prevalence, patterns, and predictors of the use of the internet, digital devices, and apps for health purposes as well as preferences for future use of eHealth support in South Asian Canadians.

Methods: We conducted a cross-sectional, mixed-mode survey in a convenience sample of 831 South Asian adults recruited at faith-based gathering places, health care settings, and community events in Edmonton, Alberta, in 2014. The 706 responders (mean age 47.1, SD 17.6 years; n=356, 50.4% female; n=509, 72.1% Sikh) who provided complete sociodemographic information were included in the analysis, and the denominators varied based on the completeness of responses to each question. Multivariate logistic regression was used to determine sociodemographic and health status predictors of internet use, being a web-based health information seeker, smartphone or tablet ownership, health app use, and willingness to use various modes of eHealth support.

Results: Of all respondents, 74.6% (527/706) were internet users and 47.8% (336/703) were web-based health information seekers. In addition, 74.9% (527/704) of respondents owned a smartphone or tablet and 30.7% (159/518) of these had a health and fitness app. Most internet users (441/527, 83.7%) expressed interest in using ≥1 mode of eHealth support. Older age, being female, having less than high school education, preferring written health information in languages other than English, and lacking confidence in completing medical forms predicted lack of internet use. Among internet users, factors that predicted web-based health information seeking were being female, use of the internet several times per day, being confident in completing medical forms, and preferring health information in English. Predictors of not owning a smartphone or tablet were being older, preferring health information in languages other than English, having less than high school education, living in Canada for <5 years, having a chronic health condition, and having diabetes. Increasing age was associated with lower odds of having a health app. Preferring health information in languages other than English consistently predicted lower interest in all modes of eHealth support.

Conclusions: eHealth-based chronic disease prevention and management interventions are feasible for South Asian adults, but digital divides exist according to language preference, education, age, sex, confidence in completing medical forms, and number of years lived in Canada. Community-based, culturally tailored strategies targeting these factors are required to address existing divides and increase the uptake of credible web-based and app-based resources for health purposes.

(JMIR Public Health Surveill 2021;7(1):e20671) doi:10.2196/20671

KEYWORDS

consumer health information; cardiovascular disease; type 2 diabetes; eHealth; mobile phone; ethnicity; cross-sectional survey; Canada

Introduction

South Asians originating from India, Pakistan, Bangladesh, and Sri Lanka are among the fastest growing and largest visible minority groups in Canada [1]. Cardiovascular disease (CVD) and diabetes are among the most prevalent health problems facing South Asians regardless of whether they live in their country of origin or abroad [2]. Recent reviews have highlighted that South Asian migrants in Canada have 1.5 to 2 times the prevalence of coronary artery disease compared with age- and sex-adjusted Whites of European ancestry [2-4]. New cases of CVD disproportionally affect younger South Asian individuals. This was demonstrated in a large, international case-control study where the median age of first myocardial infarction in South Asians (53 years) was 6 to 10 years younger than those in North America or Western Europe [5].

The increased risk of coronary artery disease is primarily driven by a higher incidence of known atherosclerotic CVD risk factors, particularly type 2 diabetes and impaired glucose tolerance [6]. Both biological and nonbiological mechanisms are implicated in the increased risk of coronary artery disease and diabetes. For example, a recent meta-analysis found that South Asian Canadians had a higher prevalence of type 2 diabetes, hypertension, abdominal obesity, percentage body fat, increased carbohydrate intake, and sedentary lifestyle [3]. Individual studies have shown that South Asians are 2 to 3 times more likely to develop type 2 diabetes compared with other populations and develop diabetes at a younger age; approximately 4.6 years younger than Chinese or White Canadians [7-10]. Differences in genetic factors may explain some of the increased rates of CVD risk factors, but existing evidence suggests that the biology of CVD is no different in South Asians compared with other ethnic groups [6]. Nonbiologic mechanisms, including acculturation, a shift from traditional dietary habits, physical inactivity, other environmental factors (eg, psychosocial stress, social support), and access to health services, have all been implicated in the increased risk of CVD, diabetes, and other CVD-related risk factors [6,11,12].

Clinical practice guidelines recommend lifestyle management focusing on diet and physical activity, pharmacologic therapy, and self-management education in the primary prevention and management of CVD and diabetes and their associated risk factors [6,13,14]. Despite these recommendations, evidence suggests that risk factors and diabetes control are suboptimal in South Asian individuals [15]. Canadian data suggest that 55% of South Asian patients are above-recommended blood glucose A_{1c} targets, 36% exceed blood pressure targets, and 58% exceed lipid level targets [15]. Language barriers, sociocultural factors, limited diabetes and CVD awareness, lack of access to culturally tailored diet counseling, misconceptions around diet, perceptions around physical activity, and lower compliance with pharmacotherapy may contribute to the increased risk [2,16-18].

There has been large growth in consumer-oriented health information technology, such as Web 2.0, and app-based interventions supporting healthy lifestyles and the management of chronic health conditions [19]. Emerging evidence suggests that mobile health (mHealth), internet, and social media–based interventions may improve the prevention and management of chronic health conditions [20], cardiovascular risk factors including unhealthy diet and physical inactivity [21,22], and diabetes [23-25]. Several successful culturally tailored programs targeting diabetes and cardiovascular risk have been developed in Canada, but accessing these programs can be challenging [26-30]. The use of credible consumer-oriented eHealth resources by the South Asian community in Canada could increase access to and efficiency in the delivery of culturally tailored chronic disease self-management programs, which may further assist in the prevention and management of CVD and type 2 diabetes and their common risk factors and complications in this high-risk population.

Large, nationally representative surveys suggest high levels of digital device ownership [31], uptake of the internet [32], and web-based health information seeking in North America [33,34]. However, digital divides in internet use for health information related to sociodemographic factors and ethnicity [35] exist in the United States. There is limited information on use patterns and predictors of web-based health information–seeking behaviors and use of digital devices for health purposes among English- and Punjabi-speaking South Asian Canadians. This information is important and could be used to justify and inform the development of tailored consumer-oriented eHealth interventions. Such interventions may help to overcome identified gaps in the knowledge and skills needed to effectively apply high-quality web-based and mobile phone–based resources for the prevention and management of chronic conditions. This information could also be used to inform and assist clinicians on how to optimally engage individuals with existing web-based health information resources.

The objective of this study is to describe prevalence, patterns, and predictors of internet use for health purposes, ownership of digital devices, use of health and fitness apps, and preferences for different eHealth-based support tools in a sample of English- or Punjabi-speaking South Asian adults recruited from Edmonton, Alberta. Specifically, we explore the extent to which these variables are influenced by sociodemographic, health status, and technology use factors, including age, gender, education, health literacy, language preferences, and the presence of chronic health conditions.

Methods

Study Design

We used a community-based approach and worked in partnership with 13 faith-based, cultural, community, and health care organizations in a major metropolitan Canadian city, Edmonton, Alberta. We conducted a descriptive cross-sectional, mixed-mode anonymous survey. The survey was primarily delivered via a computer-assisted personal interview using the Qualtrics (Qualtrics Corporation) web-based survey platform. One-on-one interviews using paper-based surveys and an optional web-based version were also used.

Participants, Recruitment, and Survey Administration

Participant recruitment occurred at 4 gurdwaras, 2 temples, 1 community pharmacy, 1 medical clinic, 2 community centers, and 2 large South Asian community events between May 18 and August 31, 2014. English- or Punjabi-speaking community members were eligible to participate if they were aged older than 18 years, self-identified their ethnic origins in India, Pakistan, Bangladesh, Nepal, or Sri Lanka, and were currently living in Alberta.

At community events, potential respondents were notified of the presence of the research team via announcements and posters. Potential respondents were then approached by community volunteers and presented with the survey information letter and asked if they would like to participate. If the potential participant agreed to participate, consent was implied and the survey was administered. Bilingual, trained community volunteers administered the survey in English or Punjabi according to respondents' preference. Participants who felt comfortable using tablet computers self-administered the survey.

Potential participants who were unwilling to complete the in-person survey were invited to complete the survey on the web, which was also advertised using posters in the community, via social media and word of mouth. At selected survey locations, including the participating community pharmacy and family physician clinic, we attempted to recruit consecutive attendees. At these locations, the survey was conducted while waiting to have prescriptions filled or awaiting assessment. Respondents who completed the survey in person were offered a reusable shopping bag as an incentive and the opportunity to enter a draw for a tablet computer or various gift cards, whereas those who completed the survey on the web were only eligible to enter the draw.

Survey Instrument

The e-Patient Project Survey evaluated the levels of digital device ownership, internet use, health information–seeking behaviors, health and fitness app use, levels of eHealth literacy, and preferences for participation in different modes of eHealth support (Multimedia Appendix 1). The research team developed the survey in 3 stages: literature review, key informant interviews with 16 individuals from the target communities, and a pilot test with 19 other individuals from the target communities. Most of the items were adopted from existing instruments, including the Pew Research Centre's Internet & American Life Project 2012 Health survey [34,36,37], the 2012 Statistics Canada Canadian Internet Use Survey [33], the eHealth Literacy Scale [38], and a health literacy screening questionnaire [39].

The survey was translated into Punjabi according to the World Health Organization guidance for translation and adaptation of instruments [40]. One translator with a medical background who was fluent in both Punjabi and English conducted forward translation from English to Punjabi. Emphasis was placed on conceptual rather than literal translations. A panel of 2 bilingual community member reviewers further identified and reviewed inadequate expressions and concepts in the translated version. The back translation was conducted by a separate translator who was fluent in both English and Punjabi. Translation discrepancies were discussed and addressed by the project team.

Measurement of Outcome Variables

We reported technology use outcomes as dichotomous variables. Individuals who answered affirmatively to either "Do you go online at least occasionally?" or "Do you send or receive email at least occasionally?" were characterized as internet users. Web-based health information seekers were those who indicated getting information about health on the internet or on the web. Individuals who answered affirmatively to "Is your cellphone a smartphone such as an iPhone, Android, Blackberry or Windows phone?" or "Do you own an iPad or other tablet computer such as an Android tablet, Microsoft surface or Kindle Fire?" were considered owners of a smartphone or tablet device. Owners of digital devices were asked about the use of health and fitness apps using one question, "On your smartphone or tablet, do you happen to have any health or fitness software apps (eg, track your food intake, weight, physical activity, or keep track of your medications)." We explored internet users' preferences for the use of 6 different modes of eHealth support in the future, including (1) accessing a webpage includes a forum where you could connect with others like you, (2) accessing a YouTube channel for people with your condition(s) that has experts talking about best management, (3) using a smartphone app or wearable device that can monitor your condition, track your progress on your health goals, and/or provide reminders about when to take your medications, (4) following a specific Twitter account for your condition(s), (5) signing up for personalized text messages providing health updates or reminders for your condition(s), or (6) using a web-based education program. We adjusted the response options to those who indicated having at least one chronic condition, those who indicated they had diabetes, or those without a chronic condition.

Measurement of Sociodemographic and Health Status Predictor Variables

Demographic factors included age, sex, education, marital status, duration of time lived in Canada, and the South Asian community with which respondents identified. Self-rated health status was assessed using a single question from the 36-item Short Form survey. The presence of 6 chronic health conditions was assessed by asking, "Have you ever been told by a doctor, nurse or other health care professional that you have, followed by the response options (eg, 'diabetes or sugar disease')." Language preference was assessed by asking, "In what language would you prefer to receive written health information?" and categories were collapsed into *includes English* or *does not include English*. One question "How confident are you filling out medical forms by yourself" estimated health literacy [39].

Analysis

We limited the analysis to individuals who provided complete information on sociodemographic variables, language preference, health literacy, health status, and diabetes status variables. Surveys with missing data for other items were included in the analysis. Descriptive statistics were tabulated and depicted as the proportion of valid cases where incomplete responses for each outcome variable and *choose to not answer* or *don't know* responses were considered missing. Descriptive data were analyzed using SPSS version 23 for Mac (IBM Corporation). Logistic regression was performed using R 3.1.3 (The R Foundation) to assess the effect of demographic and health factors on the dichotomous outcome variables. Variables shown to be statistically ($P<.05$) and clinically significant in the descriptive and univariate level analyses were selected to be included in the models. Self-rated health status was dropped for models that could not include all the factors. This was based on the widely used rule of thumb that there should be at least 10 events per parameter. This factor was dropped, as it was thought to be the least important. Other models included all variables. Multicollinearity was assessed by variance inflation factor (VIF), and VIF coefficients >10 were considered as high multicollinearity.

Goodness-of-fit, measuring the discrepancy between observed values and the expected value under the model, was assessed by using Craig and Uhler Pseudo R-square, Hosmer and Lemeshow goodness-of-fit test, and area under the curve (AUC). A P value <.05 indicated statistical significance. All the models fit reasonably well, as multicollinearity was not present in any model, all P values for the Hosmer-Lemeshow goodness-of-fit were >.05 (indicating no evidence of poor fit), and all AUC scores were greater than 0.7. However, most Craig and Uhler Pseudo R-square values were low (<0.5).

Ethics Approval

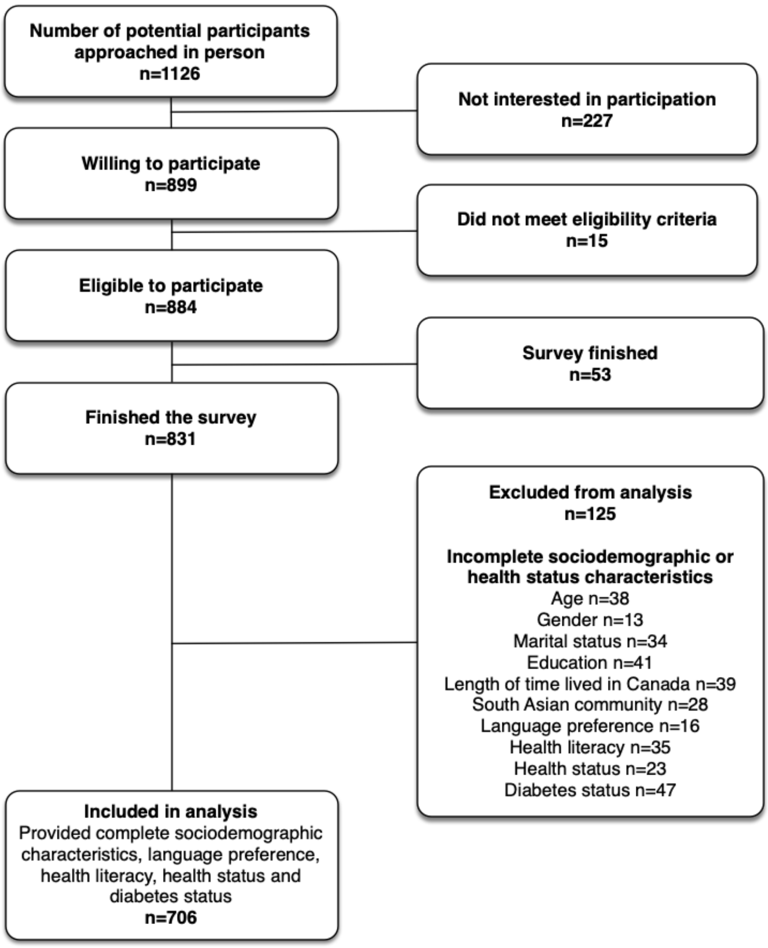
The Health Research Ethics Board at the University of Alberta (Pro00038210) approved this study.

Results

Participant Flow

We approached 1126 potential participants for face-to-face surveys at community events and 831 agreed to complete the survey. A total of 706 individuals (706/831, 85.0%) provided complete sociodemographic and health status information (Figure 1).

Figure 1. Flow diagram of the e-Patient Project Survey, Edmonton, Alberta, in 2014.



Participant Characteristics

The characteristics of the 706 study participants are shown in Table 1. Overall, the mean age was 47.1 (SD 17.6) years, and 50.4% (356/706) were female, 64.6% (456/706) had college or university education, and 72.1% (509/706) self-identified as Sikh. A total of 25.4% (179/706) of the participants lived in Canada for <5 years, and 31.0% (219/706) preferred written health information in a language other than English. Overall, 53.4% (377/706) of the participants self-reported at least one chronic health condition and 19.8% (140/706) reported diabetes. Most respondents rated their own health in the past 4 weeks as good (283/706, 40.1%) or very good (169/706, 23.9%), whereas 15.6% (110/706) and 3.7% (26/706) rated their health status as fair and poor, respectively. In addition, 11% (78/706) of the participants indicated being not at all confident in filling out medical forms on their own. Most respondents (397/706, 56.2%) were recruited at places of faith-based gathering and community events, whereas a minority (41/706, 5.8%) completed the survey on their own on the web.

Table 1. Demographic characteristics of the study sample.

Characteristics	Respondents	Missing data, n (%)
Age (years;n=706)		N/A[a]
Mean (SD)	47.1 (17.6)	
Median (IQR)	45 (32-63)	
Age group (years; n=706), n (%)		N/A
18-34	213 (30.2)	
35-49	181 (25.6)	
50-64	156 (22.1)	
≥65	156 (22.1)	
Sex (n=706), n (%)		N/A
Male	350 (49.6)	
Female	356 (50.4)	
Marital status (n=706), n (%)		N/A
Not married	140 (19.8)	
Married	566 (80.2)	
Education (n=706), n (%)		N/A
Less than high school	63 (8.9)	
High school	187 (26.3)	
College, university, or higher	456 (64.6)	
Lived in Canada (years; n=706), n (%)		N/A
>5	527 (72.1)	
0-5	179 (25.4)	
Community (n=706), n (%)		N/A
Sikh	509 (72.1)	
Hindu	134 (19.0)	
Other	63 (8.9)	
Language preference (n=706), n (%)		N/A
English	487 (69.0)	
Not English	219 (31.0)	
Confidence in filling out medical forms (n=706), n (%)		N/A
Greater than not at all	628 (89.0)	
Not at all	78 (11.0)	
Health status (n=706), n (%)		N/A
Excellent	118 (16.7)	
Very good	169 (23.9)	
Good	283 (40.1)	
Fair	110 (15.6)	
Poor	26 (3.7)	
Medical conditions, n (%)[b]		
Diabetes or sugar disease (n=706)	140 (19.8)	N/A
High blood pressure (n=706)	178 (25.2)	N/A
Heart disease (eg, angina, heart attack, or stroke; n=689)	48 (7.1)	17 (2.4)

Characteristics	Respondents	Missing data, n (%)
Lung conditions (eg, asthma or bronchitis; n=687)	41 (5.8)	19 (2.7)
Arthritis (n=688)	124 (17.6)	18 (2.5)
Cancer (n=684)	23 (3.4)	22 (3.1)
Other chronic condition treated with daily medication (n=688)	110 (16.2)	18 (2.5)
High cholesterol (n=608)[c]	144 (23.7)	98 (13.9)
≥1 condition (n=706), n (%)		N/A
No	329 (46.6)	
Yes	377 (53.4)	
Location of recruitment (n=706), n (%)		N/A
Community setting	397 (56.2)	
Health setting	268 (38.0)	
On the web	41 (5.8)	

[a]N/A: not applicable.

[b]Data are n (%) out of 706 respondents unless there were missing data, in which case the n (%) of valid cases is reported.

[c]High cholesterol was unintentionally omitted from the paper version of the survey administered at the first community event.

Internet Use, Sources of Health Information, and Web-Based Health Information–Seeking Behavior in Internet Users

Overall, 74.6% (527/706) of respondents were classified as internet users, whereas 25.4% (179/706) were nonusers (Table 2). Respondents used a median of 3 (IQR 2-5) different sources of health information, most commonly their doctor or health care provider (656/704, 93.2%) and family (398/702, 56.7%). Overall, 47.8% (336/703) of all respondents, or 63.4% (332/524) of internet users, used the internet for health information (Table 2). When asked how important it is to find health information tailored to their specific needs as a person of a South Asian background, 73.8% (513/695) indicated it was *very* or *extremely* important.

Patterns of use among the 527 internet users are shown in Multimedia Appendix 2. Most internet users (373/517, 72%) were on the web several times per day and most watched videos on YouTube, used social media sites, or made video calls. The most commonly reported web-based health information–seeking tasks were looking for information on healthy lifestyles (354/524, 67.6%) and on a specific disease or medical condition (248/460, 53.9%) and symptoms they were experiencing (222/523, 42.4%).

Regarding the use of Web 2.0 for health, just less than half of internet users (240/523, 45.9%) watched a web-based video about health or medical issues, 42.4% (222/523) read about someone else's experience about health or medical issues in a blog, newsgroup, or website, and 29.5% (153/518) reported going on the web to find others who might have similar health concerns**.** Although there were significant missing data because of a problem with the printed version of the survey, just more than one-fourth of respondents indicated that web-based health information they found or someone else found for them affected a treatment decision (94/343, 27.4%), whereas more participants responded that it had led them to ask their doctor new questions, go to see their doctor, or change the way they maintain their health (Multimedia Appendix 2).

Table 2. Internet user status, sources of health information, digital device ownership, and health and fitness apps.

Characteristics	Overall, n (%)	Missing data, n (%)
Internet use (n=706)		
Internet user	527 (74.6)	N/A[a]
Noninternet user	179 (25.4)	N/A
Where do you get information about health questions that you have?		
Doctor or health care provider (n=704)[b]	656 (93.2)	2 (0.3)
Family (n=702)	398 (56.7)	4 (0.6)
Internet (n=703)	336 (47.8)	3 (0.4)
Print (n=704)	309 (43.9)	2 (0.3)
Friends (n=703)	285 (40.5)	3 (0.4)
TV or radio (n=705)	283 (40.1)	2 (0.3)
Others with the same condition (n=639)	66 (10.3)	67 (9.5)
Never looked (n=618)	14 (2.3)	88 (12.5)
How important is it for you to find health information tailored to your needs as someone of South Asian background? (n=695)		
Extremely or very important	513 (73.8)	11 (1.6)
Device ownership (Do you own…)		
A desktop or laptop computer at home connected to the internet? (n=701)	615 (87.7)	5 (0.7)
A cellphone, iPhone, Blackberry, or other device that is a cellphone? (n=706)	571 (80.9)	N/A
Is your cellphone a smartphone? (n=571)	443 (77.6)	N/A
An iPad or other tablet computer (n=704)	376 (53.4)	2 (0.3)
Smartphone or tablet (n=704)	527 (74.9)	2 (0.3)
Device use (Do you use your smartphone or tablet to…)		
Send or receive text messages (n=527)	432 (82.0)	N/A
On your smartphone or tablet, do you happen to have any health or fitness apps? (n=518)	159 (30.7)	9 (1.7)
What type of health and fitness apps are you currently using? (n=138)		
Tracking food, diet, or calorie intake	88 (63.8)	21 (13.2)
Monitoring weight	49 (35.5)	21 (13.2)
Physical activity tracking	65 (47.1)	21 (13.2)
Track runs that you take	20 (14.5)	21 (13.2)
Mobile pedometer	32 (23.2)	21 (13.2)
Research or diagnose medical conditions	9 (6.5)	21 (13.2)
Keep track of medications	9 (6.5)	21 (13.2)
Stress management	17 (12.3)	21 (13.2)
Communicate with your doctor or health care provider	10 (7.2)	21 (13.2)
Monitor sleep cycle	12 (8.7)	21 (13.2)
Record your blood pressure	13 (9.4)	21 (13.2)
Record your blood sugar or diabetes	8 (5.8)	21 (13.2)
Other	6 (4.3)	21 (13.2)

[a]N/A: not applicable.

[b]Data are n (%) out of 706 respondents unless otherwise specified. When there were missing data, the n (%) of valid cases was reported.

Digital Device Ownership and Health and Fitness Apps in Smartphone or Tablet Owners

Overall, 62.8% (443/705) of respondents owned a smartphone, 53.4% (376/704) owned a tablet computer, and 74.9% (527/704) owned either a smartphone or tablet (Table 2). Most smartphone or tablet owners (432/527, 82%) reported sending or receiving text messages. Just less than one-third of the smartphone or tablet owners surveyed (159/518, 30.7%) indicated that they had a health and fitness app on their mobile device (Table 2). The most commonly used apps included those designed to track food, diet, or calorie intake (88/138, 63.8%), track physical activity (65/138, 47.1%), and monitor weight (49/138, 35.5%).

Preferences for Future eHealth Interventions in Internet Users

Most internet users (441/527, 83.7%) responded that they were likely or very likely to use at least one of the 6 presented eHealth tools to address a health issue in the next 12 months (Multimedia Appendix 2). Although there were some systematic issues with missing information regarding YouTube, Twitter, and a web-based education program, most respondents favored accessing a YouTube channel (330/425, 77.6%) followed by using a webpage with peer-to-peer support (353/500, 70.6%), using an app or a wearable device (316/493, 64.1%), or receiving personalized text messages (282/483, 58.4%).

Barriers in Nonusers of the Internet

The 179 respondents who were not internet users reported several barriers, the most common being *lack of skills* (114/177, 64.4) and *no interest* (72/177, 40.7%; Table 3). One-third of nonusers (57/178, 32%) said they were planning to get on the web in the future. Of these, 67.3% (35/52) indicated being *likely* or *very likely* to attend a hands-on workshop, whereas 72.2% (38/52) were *likely* or *very likely* to learn from a friend or family member. Most internet nonusers (136/175, 77.7%) reported knowing someone who could help them get on the web.

Table 3. Barriers to nonusers of the internet.

Characteristics	Noninternet users, n (%)	Missing data, n (%)
What are the reasons you do not go on web?[a]		
Lack of skills (n=177)	114 (64.4)	2 (1.1)
No interest (n=177)	72 (40.7)	2 (1.1)
Too late to learn (n=176)	38 (21.1)	3 (1.7)
Limited access to a computer (n=176)	9 (5.1)	3 (1.7)
Uncomfortable using a computer (n=175)	7 (4.0)	4 (2.2)
Privacy reasons (n=176)	7 (4.0)	3 (1.7)
Fear of technology (n=175)	5 (2.9)	4 (2.2)
Because of disability (n=176)	4 (2.3)	3 (1.7)
Cost (n=178)	3 (1.7)	1 (0.6)
Are you likely to start going on web in the future?		
Yes, within 6 months (n=178)	23 (12.9)	1 (0.6)
Yes, within 6 to 12 months (n=179)	16 (9.0)	N/A[b]
Yes, in more than a year (n=179)	18 (10.1)	N/A
Not likely (n=179)	61 (34.3)	N/A
Never (n=179)	60 (33.7)	N/A
Likeliness to use the following strategies to improve their ability to go on the web (n=57; who answered yes)		
Likely or very likely to attend a hands-on workshop (n=52)	35 (67.3)	5 (8.8)
Likely or very likely to talk with a friend or family member (n=52)	38 (72.2)	5 (8.8)
Know someone who could help them, if they needed to go on web to do something (n=175)	136 (77.7)	4 (2.2)

[a]Data are n (%) out of 179 noninternet users unless otherwise specified. When there were missing data, the n (%) of valid cases was reported.
[b]N/A: not applicable.

Predictors of Internet Use and Web-Based Health Information Seeking

As shown in Multimedia Appendix 3, 5 of the predictor variables were associated with internet use in the multiple logistic regression analysis, including all 706 respondents. Preferring written information in a language other than English (odds ratio [OR] 0.21, 95% CI 0.12-0.36), lacking confidence in filling out medical forms (OR 0.27, 95% CI 0.11-0.65), being female (OR 0.47, 95% CI 0.26-0.85), and increasing age (OR 0.92, 95% CI 0.90-0.94) were predictive of lower internet use, whereas educational achievement (OR 4.00, 95% CI 1.52-11.07 for college university or higher) predicted greater odds of internet use.

There were 4 independent predictor variables of the use of the internet for health information in the 514 internet users. Females (OR 2.34, 95% CI 1.49-3.71) and people who used the internet several times per day (OR 3.83, 95% CI 2.36-6.30) were more likely to be web-based health information seekers, whereas those lacking confidence in filling out medical forms (OR 0.24, 95% CI 0.07-0.72) and those expressing a preference for written health information in languages other than English (OR 0.53, 95% CI 0.30-0.94) were less likely to be web-based health information seekers (Multimedia Appendix 3).

Predictors of Digital Device Ownership and Having a Health and Fitness App

A total of 6 predictor variables were associated with ownership of a smartphone or tablet when the whole group was assessed (n=704): educational achievement (college or university or higher: OR 5.44, 95% CI 2.36-12.96) was associated with higher odds of device ownership, while living in Canada for <5 years (OR 0.47, 95% CI 0.27-0.81), preferring written information in languages other than English (OR 0.51, 95% CI 0.31-0.86), having a chronic health condition (OR 0.53, 95% CI 0.31-0.90), having diabetes (OR 0.50, 95% CI 0.28-0.87), and increasing age (OR 0.94, 95% CI 0.92-0.95) were associated with lower odds of device ownership (Multimedia Appendix 3). In the subgroup of smartphone or tablet owners (n=521), only increasing age was associated with lower odds of having downloaded a health and fitness app (OR 0.97, 95% CI 0.95-0.99; Multimedia Appendix 3).

Predictors of Preferences for Future eHealth Interventions in Internet Users

The multivariable analysis shown in Multimedia Appendix 4 indicated that individuals who preferred written health information in a language other than English were less interested in all modes of eHealth-based support. Those who reported watching YouTube videos were more likely to be interested in a YouTube channel for health issues. Those who are married were more interested in a website with peer support. Interest in app-based interventions was higher in those who own a smartphone or tablet but lower in participants with diabetes. Interest in text message–based interventions was higher in older individuals, those who already send text messages, use the internet several times per day, or are married (Multimedia Appendix 4).

Discussion

Principal Findings

Among our sample of primarily Sikh South Asian adults recruited from community and health care settings, we found a high prevalence of internet users and ownership of smart digital devices that allow the use of apps. Health care providers were the most common source of health information, and only less than half of all respondents reported using the internet as a source of health information. Although most smartphone or tablet owners indicated that they used texting, only one-third reported having a health and fitness app on their device. The most commonly used apps were food, diet, or calorie intake trackers. Most internet users indicated that they were likely or very likely to use at least one of the eHealth tools to address a health issue in the next 12 months, and many preferred YouTube videos, a peer-to-peer support website, or smartphone app. Among internet nonusers, lack of technological skills and interest were cited as the most common barriers, and only one-third of these respondents indicated they were likely to get on the web in the future. However, just more than three-fourth of nonusers indicated that they had access to someone who could help them use the internet.

Language preferences, higher educational attainment, and age were common factors driving a digital divide in internet use and digital device ownership in our sample of South Asian adults. Being female, frequent internet use, being confident in filling out medical forms, and preferring written information in English were all positive independent predictors of using the internet for health information purposes in internet users. Age independently influenced whether participants reported having downloaded a health and fitness app. Those who preferred written information in languages other than English showed less interest in all modes of future eHealth support.

Comparison With Previous Research in the South Asian Community

At the time the survey was conducted in 2014, our study was unique in that it was the first to use a community-based approach, where we mobilized community resources in health, faith gathering, and other settings to explore ownership of digital devices, internet use, and willingness to use eHealth tools specifically among members of the South Asian community in Canada. Furthermore, to our knowledge, it remains to be the only study to explore the predictors of these outcomes in this growing ethnocultural minority group. Data on the use and uptake of technology to address health needs in South Asians in India [41], Sri Lanka [42], and Pakistan [43] have suggested highly variable rates of web-based health information seeking among internet users (ie, 1%-75%). Our findings are comparable with data reported from a 2009-2010 survey of 709 South Asian adults living in the metropolitan Washington DC region, which found that the internet was the leading source of health information (76.9%) [44]. They also found that older participants and those who were US born were more likely to obtain health information from physicians rather than the internet, whereas those who rarely or never speak English at home are more likely to cite friends as a source of health information rather than the internet. We also found that language preference and age were predictors of web-based health information seeking, whereas the duration of time lived in Canada was not a predictor.

Comparison With Previous Research in the General Population

Previous research exploring variables influencing internet use has identified age [32,45], education [46-49], English language proficiency [47,50], and health literacy [51] as predictors of internet use. Similarly, we found education to be a strong predictor of internet use, whereas preference for South Asian languages (rather than English) predicted lower odds of internet use. We did not find a relationship between internet use and recent immigration [52] or the presence of chronic conditions [53]. An analysis of the 2010 Canadian internet use survey documented that recent immigrants to Canada have lower rates of internet access but that recent immigrants who are on the web have significantly higher levels of web-based activity than Canadian-born residents and earlier immigrants [52]. Although recent studies do not suggest differential internet uptake between males and females [46], in our sample, we found that being female was independently associated with a lower likelihood of internet use. Although gender inequality has existed in South Asian culture [54,55] and may contribute to this difference, males and females in our study gave similar reasons for not being internet users.

Although several theories have been used to explain health information seeking on the internet, the most recent reviews of methods and measures for health information do not provide insight into the factors predicting the uptake of these behaviors [56,57]. Most studies investigating predictors of using the internet for health information purposes identify age, female sex, and educational attainment as independent sociodemographic predictors, whereas other studies have also identified other demographic factors (race, income, and employment), health status, health care access, and digital literacy factors (eg, internet usage intensity) as mediators of web-based health information seeking [57-64]. Although inconsistent, most studies have found that age is a significant predictor of web-based health information–seeking behaviors [58,59,61]. Generally, as age increases, web-based health information seeking decreases; however, the relationship is complex. For example, Veenhof et al [48] documented that Canadians aged 16 to 25 years were significantly less likely to go on the web to search for health-related information than Canadians aged 26 to 65 years. Interestingly, among all respondents in our study, increasing age was a negative predictor of web-based information seeking; however, it was not a significant predictor among internet users. Our finding that female internet users are more likely to be web-based health information seekers is consistent with that reported by others [58-60,62].

Similar to others, we found that smartphone or tablet owners were more likely to be younger, affluent, and highly educated than nonowners [31]. Our finding that 30% of smartphone or tablet owners used health apps is consistent with the range of 19% to 58% reported in studies of racially diverse Americans [36,65,66]. Our finding that younger individuals were more likely to use health apps is consistent with a national survey of 1604 American mobile phone users that found individuals who were younger, had higher income, were more educated, were Latino or Hispanic, and had a BMI in the obese range were more likely to use health apps [66].

Finally, several studies have explored willingness to get on the web and future use of eHealth tools [67-73]. Our finding that 32% of noninternet users thought they would go on the web is higher than the 8% who said they would like to start using the internet or email in a recent Pew Research Centre survey [67]. Encouragingly, 67% of noninternet users reported that they would likely go on the web in the future, indicating that they would be willing to take a workshop or learn from a friend or family member (72%). Other surveys have reported varying levels of interest in specific eHealth interventions, from highs of 83% of women willing to participate in an internet-based postpartum weight loss intervention [68] to lows where only 18% preferred to learn health, wellness, and lifestyle information from a mobile app [70]. Recently, a qualitative study of British South Asians suggested that short text messages to support medication adherence for type 2 diabetes would be acceptable and relevant [74]. Although limited research exists, language preferences and age have been found to predict willingness to use internet or smartphone app–based interventions for health [71,73], consistent with our finding that increasing age is a negative predictor of app use.

Clinical Implications

First, our results suggest that community-based, culturally tailored strategies would be welcomed and are required to reduce identified digital divides and increase the uptake and use of credible web-based and app-based resources for health purposes among South Asian adults who are current internet users and nonusers. This is particularly timely, given the need to consider and increase remote delivery of health care based on social distancing as a result of the COVID-19 pandemic. Although eHealth and mHealth interventions appear to be more likely to reach certain subgroups of individuals, such as those that are younger, English speaking, and with high educational achievement, particular attention must be paid not to exacerbate health inequities based on these digital divides. Although most internet users were interested in YouTube or web-based peer support interventions, a range of eHealth interventions will be needed to meet the needs of various subgroups within the community. Device and internet training offered in a culturally relevant way for noninternet users who are interested in getting on the web may reduce identified divides, whereas different means, such as using friends or family as intermediaries, will be required to reach noninternet users, particularly those who have no intention to get on the web. Second, as web-based resources are not designed to replace health professional interactions [75], health care professionals and health organizations must play an important role in referring and supporting patients to access credible eHealth resources, including those that are tailored specifically to South Asian health needs.

Limitations

Our study has several limitations. First, as nonprobability (ie, convenience) sampling was used, selection bias and sampling error make generalizability to the larger South Asian population a concern. We did not translate our survey into other commonly spoken languages (eg, Hindi, Urdu), and our results primarily pertain to the English- and Punjabi-speaking Sikh community. Second, our data were collected in 2014 and are likely not reflective of evolution in the use or ownership patterns. Third, the survey tool itself is not validated; however, most questions were sourced from existing large national surveys or other validated surveys. We recognize that the question relating to language preference for written health information could have been improved by instead asking about the primary language spoken in the home and that the use of a single health literacy screening question, rather than a full health literacy questionnaire, is not optimal. Although smoking is a well-established risk factor for CVD [76], we did not ask about smoking status or use of web-based health information or apps for smoking cessation as part of our survey. This was based on evidence that South Asian Canadians are less likely to be current or former smokers than Canadians of European descent [77] and that smoking is very rare among South Asian Canadian women [78]. Interestingly, a survey conducted around the same time as ours in British Columbia, Canada, also suggested that smoking rates are considerably lower in South Asians than in the general population (never smokers: 87% vs 59%) [79], as does other Canadian research [3]. However, our exclusion of smoking status may be an oversight, as surveys may fail to accurately capture the use of culturally specific smokeless tobacco products [80]. Furthermore, although the survey was translated into Punjabi and formally pretested, community volunteers were trained to administer the survey in Punjabi and 2 volunteers administered just more than 50% of the surveys, there may have been issues with conceptual translation and variability in survey administration. We had some issues with the early version of the survey administered on paper, which resulted in missing data for certain items. Fourth, volunteer and social desirability bias may overinflate our estimates of device ownership, internet use, and willingness to use future eHealth tools. In addition, self-report may have introduced recall bias in outcome and demographic variables. Fifth, the questions relating to likeliness to use eHealth interventions in the future were hypothetical and therefore may overestimate actual willingness had we asked participants to participate in a pilot test of eHealth interventions. Finally, we did not explore differences in survey responses by survey mode or by recruitment location.

Conclusions

Our survey provides insights into digital divides according to language preferences, education, age, and sex in an ethnocultural minority community in Canada. The high overall rates of internet use for health information, digital device ownership, and interest in eHealth-based interventions in internet users along with high access to individuals who could help them use the internet among nonusers suggest that eHealth supports are feasible among segments of English- or Punjabi-speaking South Asian adults. There is an opportunity for health care providers and health organizations to enhance the use of reliable and culturally relevant eHealth resources to promote health, prevent chronic disease, and support self-management of chronic health conditions for South Asian adults.

Acknowledgments

This work was generously supported by the Lawson Foundation (grant number GRT 2012-057). They had no role in the design, collection, analysis, or interpretation of data, writing of the report, or decision to submit the paper for publication. The authors wish to acknowledge the members of the South Asian Communities in the Edmonton Area who participated in the e-Patient Project, e-Patient Project Advisory Committee Members, particularly Dr Sudheer Sharma, Mr Jesse Bhondi, Ms Nandini Desai, Mr Jasbir Bhui, and Ms Sanjeet Chattha. The authors thank those who provided access to the participating sites Medicine Shoppe Pharmacy #170, Meadowbrook Medical Clinic (Dr Narpinder Hans), Gurdwara Millwoods, Gurdwara Singh Sabha, Gurdwara Nanaksar, Sikh Society of Alberta, Hindu Society of Alberta, Bharthia Cultural Society, Millwoods Cultural Society of Retired & Semi Retired, Mill Woods Seniors Activity Centre, and organizers of community events (Nagar Kirtan and Tiyan Da Mela) where the survey was administered. The authors thank Ms. Kaitlin Laurisden and Dr Bikram Jammu who served as project coordinators and oversaw the coordination of the survey. The authors thank Ms Manjot Bhui, Ms Lakwinder Grewal, the numerous other community volunteers including those from Headway School and J Percy Page School for assisting with survey administration. They thank Mr Kirtmeet Singh Kohar at the Nawi Dunia Newspaper, Ms Nirlep Kaur Rai, and Mr Jagwinder Singh Sidhu for assistance in the survey translation. The authors thank the Health Sciences Research & Education Commons at the University of Alberta for loaning tablet computers to the research team for survey administration. They thank Dr Ken Cor for facilitating access to the Qualtrics web-based survey platform and helping in survey development. The authors acknowledge the EPICORE Center at the University of Alberta for conducting the multivariable regression analysis. Finally, the authors thank Dr Ravina Sanghera and Dr Lisa Guirguis for their thoughtful comments on this manuscript.

Authors' Contributions

All authors contributed to conception and design, critical revision of the manuscript for intellectual content, and final approval of the version to be published. MM contributed to acquisition and analysis and interpretation of data and prepared the first draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
e-Patient Project Survey.
[DOCX File , 49 KB - publichealth_v7i1e20671_app1.docx]

Multimedia Appendix 2

Web-based health information–seeking patterns and preferences for future eHealth interventions in internet users.
[DOCX File , 36 KB - publichealth_v7i1e20671_app2.docx]

Multimedia Appendix 3
Predictors of being an internet user, using the web for health information, smartphone or table ownership, and having health and fitness apps.
[DOCX File , 22 KB - publichealth_v7i1e20671_app3.docx]

Multimedia Appendix 4
Determinants associated with being likely or very likely to use different modes of eHealth support in the future in internet users.
[DOCX File , 37 KB - publichealth_v7i1e20671_app4.docx]

References

1. Visible Minority Population and Population Group Reference Guide, 2006 Census. Statistics Canada. 2006. URL: https://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/visible_minority-minorites_visibles-eng.cfm [accessed 2012-05-17]
2. Fernando E, Razak F, Lear SA, Anand SS. Cardiovascular disease in South Asian migrants. Can J Cardiol 2015 Sep;31(9):1139-1150. [doi: 10.1016/j.cjca.2015.06.008] [Medline: 26321436]
3. Rana A, de Souza RJ, Kandasamy S, Lear SA, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. CMAJ Open 2014 Jul;2(3):E183-E191 [FREE Full text] [doi: 10.9778/cmajo.20130064] [Medline: 25295238]
4. Zaman MJ, Philipson P, Chen R, Farag A, Shipley M, Marmot MG, et al. South Asians and coronary disease: is there discordance between effects on incidence and prognosis? Heart 2013 May;99(10):729-736 [FREE Full text] [doi: 10.1136/heartjnl-2012-302925] [Medline: 23406688]
5. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
6. Volgman AS, Palaniappan LS, Aggarwal NT, Gupta M, Khandelwal A, Krishnan AV, American Heart Association Council on EpidemiologyPrevention; Cardiovascular DiseaseStroke in WomenSpecial Populations Committee of the Council on Clinical Cardiology; Council on CardiovascularStroke Nursing; Council on Quality of CareOutcomes Research;Stroke Council. Atherosclerotic cardiovascular disease in South Asians in the United States: epidemiology, risk factors, and treatments: a scientific statement from the American heart association. Circulation 2018 Jul 3;138(1):e1-e34. [doi: 10.1161/CIR.0000000000000580] [Medline: 29794080]
7. Bhurji N, Javer J, Gasevic D, Khan NA. Improving management of type 2 diabetes in South Asian patients: a systematic review of intervention studies. BMJ Open 2016 Apr 20;6(4):e008986 [FREE Full text] [doi: 10.1136/bmjopen-2015-008986] [Medline: 27098819]
8. Khan NA, Wang H, Anand S, Jin Y, Campbell NR, Pilote L, et al. Ethnicity and sex affect diabetes incidence and outcomes. Diabetes Care 2011 Jan 1;34(1):96-101 [FREE Full text] [doi: 10.2337/dc10-0865] [Medline: 20978094]
9. Mohan V, Mathur P, Deepa R, Deepa M, Shukla DK, Menon GR, et al. Urban rural differences in prevalence of self-reported diabetes in India--the WHO-ICMR Indian NCD risk factor surveillance. Diabetes Res Clin Pract 2008 Apr;80(1):159-168. [doi: 10.1016/j.diabres.2007.11.018] [Medline: 18237817]
10. Creatore MI, Moineddin R, Booth G, Manuel DH, DesMeules M, McDermott S, et al. Age- and sex-related prevalence of diabetes mellitus among immigrants to Ontario, Canada. Can Med Assoc J 2010 May 18;182(8):781-789 [FREE Full text] [doi: 10.1503/cmaj.091551] [Medline: 20403889]
11. Gujral UP, Pradeepa R, Weber MB, Narayan KM, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white caucasian and other populations. Ann N Y Acad Sci 2013 Apr;1281:51-63 [FREE Full text] [doi: 10.1111/j.1749-6632.2012.06838.x] [Medline: 23317344]
12. Sattar N, Gill JM. Type 2 diabetes in migrant south Asians: mechanisms, mitigation, and management. Lancet Diabetes Endocrinol 2015 Dec;3(12):1004-1016. [doi: 10.1016/S2213-8587(15)00326-5] [Medline: 26489808]
13. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston MN, Hubbard VS, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. J Am Coll Cardiol 2014 Jul 1;63(25 Pt B):2960-2984 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.003] [Medline: 24239922]
14. Diabetes Canada Clinical Practice Guidelines Expert Committee, Prebtani AP, Bajaj HS, Goldenberg R, Mullan Y. Reducing the risk of developing diabetes. Can J Diabetes 2018 Apr;42(Suppl 1):S20-S26. [doi: 10.1016/j.jcjd.2017.10.033] [Medline: 29650097]

15. Shah BR, Cauch-Dudek K, Anand SS, Austin PC, Manuel DG, Hux JE. Absence of disparities in the quality of primary diabetes care for South Asians and Chinese in an urban Canadian setting. Diabetes Care 2012 Apr;35(4):794-796 [FREE Full text] [doi: 10.2337/dc11-1845] [Medline: 22323411]
16. Misra A, Ramchandran A, Jayawardena R, Shrivastava U, Snehalatha C. Diabetes in South Asians. Diabet Med 2014 Oct;31(10):1153-1162. [doi: 10.1111/dme.12540] [Medline: 24975549]
17. Sohal T, Sohal P, King-Shier KM, Khan NA. Barriers and facilitators for type-2 diabetes management in South Asians: a systematic review. PLoS One 2015;10(9):e0136202 [FREE Full text] [doi: 10.1371/journal.pone.0136202] [Medline: 26383535]
18. Shariff AI, Kumar N, Yancy WS, Corsino L. Type 2 diabetes and atherosclerotic cardiovascular disease in South Asians: a unique population with a growing challenge. Curr Diab Rep 2020 Jan 30;20(1):4. [doi: 10.1007/s11892-020-1291-6] [Medline: 32002674]
19. Meier CA, Fitzgerald MC, Smith JM. eHealth: extending, enhancing, and evolving health care. Annu Rev Biomed Eng 2013;15:359-382. [doi: 10.1146/annurev-bioeng-071812-152350] [Medline: 23683088]
20. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
21. Devi R, Singh SJ, Powell J, Fulton EA, Igbinedion E, Rees K. Internet-based interventions for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2015(12):CD009386. [doi: 10.1002/14651858.CD009386.pub2] [Medline: 26691216]
22. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
23. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013 Mar 28(3):CD008776 [FREE Full text] [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
24. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
25. Tao D, Wang T, Wang T, Liu S, Qu X. Effects of consumer-oriented health information technologies in diabetes management over time: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2017 Sep 1;24(5):1014-1023. [doi: 10.1093/jamia/ocx014] [Medline: 28340030]
26. Davachi S, Flynn M, Al E. A health region/community partnership for type 2 diabetes risk factor screening in Indo-Asian communities. Can Med Assoc J 2005;29(2):87-94.
27. van Draanen J, Shafique A, Farissi A, Wickramanayake D, Kuttaiya S, Oza S, et al. How to offer culturally relevant type 2 diabetes screening: lessons learned from the South Asian diabetes prevention program. Can J Diabetes 2014 Oct;38(5):329-333. [doi: 10.1016/j.jcjd.2013.11.008] [Medline: 24797496]
28. University of British Columbia. iCON: InterCultural Online Health Network. iCON. URL: http://iconproject.org [accessed 2017-07-17]
29. STOP Diabetes Foundation. Stop Diabetes Foundation. URL: http://www.stopdiabetesfoundation.com/index.html [accessed 2017-07-20]
30. DIL Walk. DIL Walk Foundation. URL: http://www.dilwalk.ca [accessed 2017-07-20]
31. Anderson M. Technology Device Ownership. Pew Research Centre. 2015 Oct. URL: http://www.pewinternet.org/2015/10/29/device-ownership-2015-about-this-report/ [accessed 2016-08-11]
32. Individual Internet use and E-commerce. Statistics Canada. 2013 Apr 15. URL: http://www.statcan.gc.ca/daily-quotidien/111012/dq111012a-eng.htm [accessed 2012-05-21]
33. Canadian Internet Use Survey, 2012. Statistics Canada. 2012. URL: http://www.statcan.gc.ca/daily-quotidien/131126/dq131126d-eng.pdf [accessed 2012-05-21]
34. Fox S, Duggan M. Health Online 2013. Pew Research Center Internet & Technology. 2013. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2016-08-13]
35. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
36. Fox S, Duggan M. Mobile Health 2012. Pew Research Center Internet & Technology. 2012. URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2016-08-20]
37. Fox S, Duggan M. Tracking for Health. Pew Research Center Internet & Technology. 2013. URL: http://www.pewinternet.org/2013/01/28/tracking-for-health/ [accessed 2016-08-23]
38. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
39. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]

40. Process of Translation and Adaptation of Instruments. World Health Organization. URL: http://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2014-04-07]
41. Akerkar SM, Kanitkar M, Bichile LS. Use of the Internet as a resource of health information by patients: a clinic-based study in the Indian population. J Postgrad Med 2005;51(2):116-118 [FREE Full text] [Medline: 16006703]
42. Kommalage M. Use of the internet by patients attending specialist clinics in Sri Lanka: a cross sectional study. BMC Med Inform Decis Mak 2009 Feb 12;9:12 [FREE Full text] [doi: 10.1186/1472-6947-9-12] [Medline: 19210796]
43. Shaikh IA, Shaikh MA, Kamal A, Masood S. Internet access and utilization for health information among university students in Islamabad. J Ayub Med Coll Abbottabad 2008;20(4):153-156. [Medline: 19999231]
44. Vyas AN, Chaudhary N, Ramiah K, Landry M. Addressing a growing community's health needs: Project SAHNA (South Asian health needs assessment). J Immigr Minor Health 2013 Jun;15(3):577-583. [doi: 10.1007/s10903-012-9655-x] [Medline: 22684910]
45. Smith A. Older Adults and Technology Use. Pew Research Center. 2014. URL: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [accessed 2016-08-11]
46. Perrin A, Duggan M. Americans' Internet Access: 2000-2015. Pew Research Center. 2015. URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [accessed 2016-08-10]
47. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
48. Veenhof B, Clermont Y, Sciadas G. Literacy and Digital Technologies: Linkages and Outcomes. Statistics Canada. URL: http://www.statcan.gc.ca/pub/56f0004m/56f0004m2005012-eng.htm [accessed 2016-08-19]
49. Chou WS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the health Information national trends survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
50. Zickuhr KS. Digital differences. Pew Research Center. 2012. URL: http://www.pewinternet.org/2012/04/13/digital-differences/ [accessed 2016-08-23]
51. Bailey SC, O'Conor R, Bojarski EA, Mullen R, Patzer RE, Vicencio D, et al. Literacy disparities in patient access and health-related use of Internet and mobile technologies. Health Expect 2015 Dec;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
52. Haight M, Quan-Haase A, Corbett BA. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. Inf Commun Soc 2014 Mar 7;17(4):503-519. [doi: 10.1080/1369118x.2014.891633]
53. The Diagnosis Difference. Pew Research Center Science & Society. 2013. URL: http://www.pewinternet.org/2013/11/26/the-diagnosis-difference/ [accessed 2016-08-15]
54. Bajaj S, Jawad F, Islam N, Mahtab H, Bhattarai J, Shrestha D, et al. South Asian women with diabetes: psychosocial challenges and management: Consensus statement. Indian J Endocrinol Metab 2013 Jul;17(4):548-562 [FREE Full text] [doi: 10.4103/2230-8210.113720] [Medline: 23961469]
55. Assanand S, Dias M, Richardson E, Chambers N, Waxler-Morrison N. People of South Asian descent. In: Cross-Cultural Caring, 2nd ed. Vancouver, BC: UBC Press; 2005:197-246.
56. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. Patient Educ Couns 2011 Mar;82(3):346-354. [doi: 10.1016/j.pec.2010.12.008] [Medline: 21239134]
57. Marton C, Choo CW. A review of theoretical models of health information seeking on the web. J Doc 2012 Apr 20;68(3):330-352. [doi: 10.1108/00220411211225575]
58. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016 Dec;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
59. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
60. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. J Am Med Assoc 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
61. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
62. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
63. 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]
64. Renahy E, Chauvin P. Internet uses for health information seeking. Revue d'Épidémiologie et de Santé Publique 2006 Jun 21;54(3):263-275. [doi: 10.1016/s0398-7620(06)76721-9]

65. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among caucasian, filipino, korean, and latino americans: the digital link to health survey. JMIR Mhealth Uhealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
66. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
67. Zickuhr K. Who's Not Online and Why? Pew Research Center. 2013. URL: http://www.pewinternet.org/2013/09/25/whos-not-online-and-why/ [accessed 2016-08-16]
68. Urrutia RP, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet use and access among pregnant women via computer and mobile phone: implications for delivery of perinatal care. JMIR Mhealth Uhealth 2015 Mar 30;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
69. Sarkar U, Piette JD, Gonzales R, Lessler D, Chew LD, Reilly B, et al. Preferences for self-management support: findings from a survey of diabetes patients in safety-net health systems. Patient Educ Couns 2008 Jan;70(1):102-110 [FREE Full text] [doi: 10.1016/j.pec.2007.09.008] [Medline: 17997264]
70. Comstock J. Survey: Only 30 Percent of Insured Consumers Want to Track Health on a Mobile Device. MobiHealthNews. 2015 Jun 8. URL: http://mobihealthnews.com/44159/survey-only-30-percent-of-insured-consumers-want-to-track-health-on-a-mobile-device [accessed 2016-08-22]
71. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 7;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
72. Hughes LD, Done J, Young A. Not 2 old 2 TXT: there is potential to use email and SMS text message healthcare reminders for rheumatology patients up to 65 years old. Health Informatics J 2011 Dec;17(4):266-276 [FREE Full text] [doi: 10.1177/1460458211422019] [Medline: 22193827]
73. James DC, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American Women. Health Educ Behav 2018 Jun;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
74. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British South Asian patients' perspectives on the relevance and acceptability of mobile health text messaging to support medication adherence for type 2 diabetes: qualitative study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15789 [FREE Full text] [doi: 10.2196/15789] [Medline: 32310150]
75. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
76. Ezzati M, Henley SJ, Thun MJ, Lopez AD. Role of smoking in global and regional cardiovascular mortality. Circulation 2005 Jul 26;112(4):489-497. [doi: 10.1161/CIRCULATIONAHA.104.521708] [Medline: 16027251]
77. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague PA, et al. Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the study of health assessment and risk in ethnic groups (SHARE). Lancet 2000 Jul 22;356(9226):279-284. [doi: 10.1016/s0140-6736(00)02502-2] [Medline: 11071182]
78. Jones CA, Nanji A, Mawani S, Davachi S, Ross L, Vollman A, et al. Feasibility of community-based screening for cardiovascular disease risk in an ethnic community: the South Asian cardiovascular health assessment and management program (SA-CHAMP). BMC Public Health 2013 Feb 21;13:160 [FREE Full text] [doi: 10.1186/1471-2458-13-160] [Medline: 23432996]
79. South Asian Health Report. Fraser Health Authority. 2015. URL: https://www.fraserhealth.ca/health-topics-a-to-z/south-asian-health/south-asian-health-institute#.XwixAy0ZNTY [accessed 2020-07-10]
80. Mukherjea A, Modayil MV. Culturally specific tobacco use and South Asians in the United States: a review of the literature and promising strategies for intervention. Health Promot Pract 2013 Sep;14(5 Suppl):48S-60S. [doi: 10.1177/1524839913485585] [Medline: 23690257]

Abbreviations

AUC: area under the curve
CVD: cardiovascular disease
mHealth: mobile health
OR: odds ratio
VIF: variance inflation factor

Edited by G Eysenbach; submitted 25.05.20; peer-reviewed by B Sainz-de-Abajo, A Budenz; comments to author 15.06.20; revised version received 20.07.20; accepted 29.10.20; published 08.01.21.

Please cite as:
Makowsky MJ, Jones CA, Davachi S
Prevalence and Predictors of Health-Related Internet and Digital Device Use in a Sample of South Asian Adults in Edmonton, Alberta, Canada: Results From a 2014 Community-Based Survey
JMIR Public Health Surveill 2021;7(1):e20671
URL: https://publichealth.jmir.org/2021/1/e20671
doi:10.2196/20671
PMID:33416506

Original Paper

Improving Detection of Disease Re-emergence Using a Web-Based Tool (RED Alert): Design and Case Analysis Study

Nidhi Parikh[1], PhD; Ashlynn R Daughton[1], PhD; William Earl Rosenberger[1], BS; Derek Jacob Aberle[2], BSc; Maneesha Elizabeth Chitanvis[3], MPH; Forest Michael Altherr[4], MPH; Nileena Velappan[5], MS; Geoffrey Fairchild[1], PhD; Alina Deshpande[5], PhD

[1]Information Systems and Modeling Group, Los Alamos National Laboratory, Los Alamos, NM, United States
[2]Neutron Science and Technology Group, Los Alamos National Laboratory, Los Alamos, NM, United States
[3]Weill Cornell Medical, Cornell University, New York, NY, United States
[4]BioFire Diagnostics, LLC, Salt Lake City, UT, United States
[5]Biosecurity and Public Health Group, Los Alamos National Laboratory, Los Alamos, NM, United States

Corresponding Author:
Alina Deshpande, PhD
Biosecurity and Public Health Group
Los Alamos National Laboratory
TA-43, Building 1, MS - M888
Los Alamos, NM, 87545
United States
Phone: 1 505 667 9938
Email: deshpande_a@lanl.gov

Abstract

Background: Currently, the identification of infectious disease re-emergence is performed without describing specific quantitative criteria that can be used to identify re-emergence events consistently. This practice may lead to ineffective mitigation. In addition, identification of factors contributing to local disease re-emergence and assessment of global disease re-emergence require access to data about disease incidence and a large number of factors at the local level for the entire world. This paper presents Re-emerging Disease Alert (RED Alert), a web-based tool designed to help public health officials detect and understand infectious disease re-emergence.

Objective: Our objective is to bring together a variety of disease-related data and analytics needed to help public health analysts answer the following 3 primary questions for detecting and understanding disease re-emergence: Is there a potential disease re-emergence at the local (country) level? What are the potential contributing factors for this re-emergence? Is there a potential for global re-emergence?

Methods: We collected and cleaned disease-related data (eg, case counts, vaccination rates, and indicators related to disease transmission) from several data sources including the World Health Organization (WHO), Pan American Health Organization (PAHO), World Bank, and Gideon. We combined these data with machine learning and visual analytics into a tool called RED Alert to detect re-emergence for the following 4 diseases: measles, cholera, dengue, and yellow fever. We evaluated the performance of the machine learning models for re-emergence detection and reviewed the output of the tool through a number of case studies.

Results: Our supervised learning models were able to identify 82%-90% of the local re-emergence events, although with 18%-31% (except 46% for dengue) false positives. This is consistent with our goal of identifying all possible re-emergences while allowing some false positives. The review of the web-based tool through case studies showed that local re-emergence detection was possible and that the tool provided actionable information about potential factors contributing to the local disease re-emergence and trends in global disease re-emergence.

Conclusions: To the best of our knowledge, this is the first tool that focuses specifically on disease re-emergence and addresses the important challenges mentioned above.

(JMIR Public Health Surveill 2021;7(1):e24132) doi:10.2196/24132

KEYWORDS

disease re-emergence; infectious disease; supervised learning; random forest; visual analytics; surveillance

Introduction

Infectious diseases remain a leading cause of death, contributing to millions of deaths each year [1]. The current COVID-19 pandemic demonstrates the speed with which an infectious disease can travel from one location to another including new locations, and in turn become a global health threat in today's world of increased travel and globalization. COVID-19 is an infectious disease caused by a newly discovered coronavirus called SARS-CoV-2. In addition to such newly emerging diseases, some diseases that were considered controlled or eliminated are also re-emerging. The past few decades have seen the re-emergence of dengue in Brazil [2], measles in France [3], and yellow fever in Angola [4]. A re-emerging infectious disease is a disease that was a major health problem historically in a location, saw a persistent decline in its incidence, and then saw its incidence increase again. Many factors such as ecological disruptions, changing environment, urbanization and human behaviors, international travel and commerce, and war and civil unrest contribute to the re-emergence of infectious diseases [2,5-8].

Early detection and understanding of disease re-emergence is important for better response and mitigation of these events. However, there are several challenges: The definition of disease re-emergence merely suggests an up–down–up incidence pattern and does not offer any guidance on quantitative measures by which such patterns can consistently identify re-emergence. The current practice of identifying disease re-emergence relies on the knowledge and experience of public health analysts rather than specific criteria, which can lead to inconsistent identification of re-emergence [9]. While high-level factors (such as those mentioned above) contribute to the re-emergence of infectious diseases, it is difficult to identify specific factors contributing to a local disease re-emergence and requires a systematic analysis of a number of factors. Local public health analysts may not have this kind of information readily available. Currently, the recognition and understanding of global disease re-emergence relies on analysis of data about historical outbreaks at the country level around the world [10-13]. Again, such data may not be easily available for the entire world and even if available, retrospective analysis is a time-consuming process. Better methods and data are thus essential to address this challenge.

In the last few years, a number of web-based analytics, tools, and databases have been developed to collect data from multiple sources to monitor disease-related activities [14-16], provide situational awareness [17], or now-cast infectious diseases [18]. While there are currently no tools focused on detecting re-emergence, this presents an opportunity for developing new analytics.

Machine learning algorithms use observation data to identify trends and patterns that can help make better decisions. Supervised algorithms identify patterns from the data that are useful in predicting specific outcomes while unsupervised algorithms extract trends and patterns from the data without relating them to any outcomes. Both supervised and unsupervised methods are used extensively in public health. Unsupervised machine learning is used to understand spatial dynamics of an epidemic [19], extract meaningful structure in electronic health records [20], and identify subgroups among home health patients with heart failure [21]. Supervised machine learning is used for disease forecasting [22,23], mortality risk score prediction in an elderly population [24], predicting blood pressure based on health behaviors [25], and assessing vaccination sentiments [25,26]. Recently, our team developed supervised machine learning models to detect potential infectious disease re-emergence for 4 infectious diseases: measles, cholera, yellow fever, and dengue [9]. Combining such an algorithm with visual analytics could provide a rapid, easy to use, and easy to interpret tool for detecting potential re-emergence.

Visual analysis is a technique that utilizes interactive visualizations to support analytical reasoning [27]. It can help with investigative analysis and hypothesis generation [28] and is especially useful for analyzing large data sets by reducing the load on working memory, offering cognitive support, and utilizing the power of human perception [29]. Recently, visual analytics are increasingly used to analyze data in public health and health care, including human emergency room and veterinary hospital data [30]; relationships between chronic conditions, demographics, behavioral and metal health, preventative health, overarching conditions [31]; and tracking symptom evolution during disease progression [32]. We have also developed a web-based visual analytic for the investigation of infectious disease outbreaks [17].

This paper details Re-emerging Infectious Disease Alert (RED Alert), a web-based tool [33] that integrates our supervised machine learning models [9] with visual analytics to help detect/warn and understand potential re-emergence at both local and global levels for 4 diseases: measles, cholera, dengue, and yellow fever. The diseases were selected in consultation with subject matter experts (SMEs) at the World Health Organization (WHO) as diseases of concern for re-emergence. These diseases also show diversity in transmission and disease burden, allowing us to show transferability of our approach. RED Alert combines disease-related data and analytics needed to help the public health community answer the following questions for detecting and understanding disease re-emergence: Is there a potential disease re-emergence at the local (country) level? What are the potential contributing factors for this re-emergence? Is there a potential for global re-emergence?

This publication describes the methods used to answer these questions and evaluation of machine learning classifiers to detect disease re-emergence and the tool through case studies.

Methods

Data

Historical case count data, together with disease subcategories such as severe dengue and deaths, were obtained from the WHO [34-36], Gideon [37], and the Pan American Health Organization (PAHO) [38]. Population data were obtained from 2 data sets: LandScan [39] and the World Bank population data [40]. Rates for measles-containing vaccine first dose and second dose were obtained from the WHO [41] together with the WHO region membership information for each country [42]. The host, pathogen, and environment represent the traditional epidemiological triad [43] and can provide information about the potential causes of re-emergence. For indicators that can be a proxy for re-emergence causes, public health indicator data were obtained from the World Bank [44] using their application programming interface (API) [45]. Detailed information about these data sources can be found in Multimedia Appendix 2.

Development of RED Alert

RED Alert was developed for application to 4 primary diseases of concern: cholera, measles, dengue, and yellow fever. The visual analytic was developed to have a web application as a front end to the data and analysis. A web API was developed to be used by any program to access the analysis results and underlying data. The back end was developed as a Django-based application. The front end uses JavaScript to read from these API endpoints and dynamically build the corresponding visualizations.

Detection of Potential Disease Re-emergence

We integrated previously developed supervised machine learning classifiers to detect potential disease re-emergence for a given location and year [9] into RED Alert. Classifiers are supervised learning algorithms that use a set of labeled data (known observation–class pairs, eg, samples of re-emergence and non-re-emergence events [or outbreaks]) and extract patterns that help predict class (eg, re-emergence or not). These patterns can then be used to map a new observation (eg, outbreak) to a class (eg, re-emergence or not re-emergence).

We used yearly disease data at the country level to train disease-specific classifiers for the 4 diseases: measles, cholera, dengue, and yellow fever. For creating the labeled data set for each disease, the SMEs in our team were given data for 100 countries selected at random (and anonymized), and they labeled each location–year pair as a re-emergence or not. A systematic approach was followed to label the training data. For each disease, SMEs developed a re-emergence schema described in detail by Chitanvis et al [9] that takes into account general disease incidence and trend information (eg, raw incidence, case counts, change in incidence from last few years, or percentile rank) and relevant disease-specific information (eg, vaccination coverage for measles and information on severe dengue cases and death due to dengue) that can help detect potential re-emergence. These factors were organized in a decision tree format to guide the labeling process.

Selection of the Classifier

We compared 2 classifiers, decision tree and random forest, using scikit-learn, a free machine learning Python library [46]. See tables 1a-b in [9] for features used for training the classifiers and imputation methods for missing data. For both decision tree and random forest, we explored the following parameter values: (1) Split criteria: gini and entropy; (2) The number of minimum samples required at leaf nodes: 1 to 10; and (3) The number of trees for random forest: 20 to 100.

Precision, recall, and F_1 are widely used metrics to evaluate the performance of classification and can be calculated as follows:

Precision = True positives/(True positives + False positives)

Recall = True positives/(True positives + False negatives)

F_1 = 2 × (Precision × Recall)/(Precision + Recall)

As our goal was to identify all potential cases of disease re-emergence while allowing some false positives, we used F_2 to evaluate the performance of the classifiers. F_2 takes into account both precision and recall but recall is given more weightage. It can be calculated as follows:

F_2 = 5 × (Precision × Recall)/([4 × Precision] + Recall)

We evaluated classifiers on held-out test data using nested cross-validation [47], where the inner cross-validation is used to choose the optimal parameters, and the outer cross-validation is used to evaluate the performance of the model with the optimal parameters on a held-out data set to test for overfitting or generalization error. Overfitting occurs when the model learns the structure of the given data set instead of the underlying data-generating phenomenon, so it performs well on the given training data set but fails to perform well on additional data or new observations. We used leave-one-out or 1000-fold cross-validation (whichever is lower) for the inner cross-validation and 10-fold cross-validation for the outer cross-validation.

Identifying Potential Contributing Factors for Re-emergence

We developed a re-emergence causal wheel for each disease in RED Alert; an example can be seen in Figure 1B. The causal wheel was modeled on the epidemiological triad [43]: host, pathogen, and environment. In the causal wheel, these categories were further divided into subcategories based on disease-specific factors that contribute to re-emergence identified from the literature. We thus created multiple rings around the primary inner ring of the epidemiological triad in our visual display for this information in RED Alert. For example, for cholera, the broad category of environment was divided into socioeconomic and natural factors affecting the environment which included natural environment, population density, public health infrastructure, and human behavior. The natural environment was further divided into weather patterns, climate change, and natural disasters. Natural disasters were further divided into floods, typhoon/hurricane, earthquakes, and drought. This causal wheel is displayed on the web application when a user selects a disease of interest, providing general information about the component causes of re-emergence for the disease. We also added links to detailed information about a component cause to facilitate access.

Figure 1. RED Alert's input form and causal wheel. Panel A shows RED Alert's first tab, which provides descriptions about the application's core functionalities and allows user to select the disease of interest. Panel B shows the causal wheel for measles. It includes factors known to have contribute to resurgence of measles in previous scenarios. Factors at the center of circle represent components of epidemiological triad and expanding distance from the center of the circle correspond to increased specificity of factors contributing to component causes. When the user clicks on a component cause, detailed information about the component cause is provided at the bottom of the chart along with the associated references.

These component causes were mapped to 1 or more indicator variables (obtained from the World Bank), which served as the proxy measurement for the corresponding component cause. Assessment of these component causes and their interactions can help guide effective intervention strategies. We developed a table for visualization of disease-specific indicators that allows comparison of the values for the user's location and year of interest to the historical range, so that the user can determine which re-emergence cause and indicator might be contributing to country's potential re-emergence.

Component causes and corresponding related indicators for a given disease and location from 2000 to the year of interest are shown in the table. If data were not available for the year of interest, the indicator value for the most recent year when the data are available was displayed. We also identify indicators where the values for the year of interest are outside the 25th and 75th percentile or 10th and 90th percentile, as these indicators show relatively extreme values for the re-emergence year as compared to the historical values for the location of interest and hence may be potential contributing factors for the disease re-emergence. These indicators and components are displayed to the user in a form of table along with the associated value for the year of interest and statistics for historically observed values (eg, median and 25th and 75th percentiles). Indicators with values outside the 10th and 90th percentiles are highlighted in dark red or dark blue colors if they are potential risk or protective factors, respectively. Similarly, indicators with less extreme values (ie, values outside the 25th and 75th percentiles) are highlighted in light red or light blue colors if they are potential risk or protective factors, respectively.

Understanding Potential for Global Disease Re-emergence

To help identify the potential for global re-emergence, we developed a visual summary in the form of a map showing the spatial distribution of national re-emergence events (identified through the machine learning classifier described above) worldwide within recent history (last 10 years). The map is time enabled, allowing the user to scroll dynamically through the last 10 years of historic data. Re-emergence events in the year selected by the slider are colored in red, whereas re-emergence events identified in previous years are identified by black points. The size of the points represents the number of historic re-emergence events in the last 10 years. Multiple re-emergence events across different countries or continents may suggest potential for global re-emergence and require further investigation by the user.

Additional Visual Analytics

In addition to features developed to answer the 3 primary objectives described above, we developed visual analytics that could help deepen the understanding of potential contributing factors for the re-emergence and global re-emergence assessment. RED Alert visual analytics were developed to illustrate the relationship between potential contributing factors (eg, sanitation facilities, urbanization, or vaccinated percentage of the population) and disease re-emergence. We also developed visual analytics to compare locations with similar disease incidence (ie, locations with incidence within 50%-150% of user-specified data). These additional analytics were provided in a second tab of the RED Alert output.

To help assess the global disease re-emergence situation, we organized different types of global data in a third tab of RED Alert. This includes information about disease incidence globally for the last 10 years from the year of interest input by a user and recent reports of disease occurrence on FluTrackers [48], an online disease community bulletin board. We also provided the following questions on this tab that guide users through the data and facilitate hypothesis generation:

- Are the highest 2 quantiles of disease incidence dispersed over multiple continents?
- Has disease incidence intensified, across geographic areas, over time?
- Are the most recent FluTrackers community posts dispersed over multiple continents?

Evaluating RED Alert Through Case Studies

To evaluate the performance of the fully developed RED Alert analytic, we used case studies for each of the 4 diseases (measles, dengue, cholera, and yellow fever). Specific inputs were identified based on the outbreak selected, and we evaluated the output with respect to its utility in addressing the 3 main objectives that the visual analytic was developed for: (1) Can we identify potential disease re-emergence in a country? (2) What might be the contributing factors to re-emergence in that location? and (3) Are there indications of a global re-emergence based on the input situation? Using the same case studies, we also evaluated the utility of the additional visual analytics that guide hypothesis generation and provide actionable information to the user. We identified the scope of use and the type of actionable information that can be obtained from RED Alert by defining specific work roles to also understand the broad utility and diversity of information that can be used.

Case studies were selected from the 2015 to 2019 timeframe to best illustrate every feature of the analytic. One of the primary challenges is the availability of updated global data. As RED Alert is dependent upon the updating cycle of data sources used (World Bank and WHO), it is often difficult to examine all the features using the current year. Complete, global data sets for public health indicators and infectious disease case counts are currently available up to 2017 or 2018. However, we believe this is still a reasonable representation of situations that occurred in 2019/2020 and the near future of about 5 years, as the natural and built environments are not expected to significantly change in such a short timeframe.

Results

Detecting Potential Disease Re-emergence

We selected random forest as the classifier to integrate into RED Alert because it outperformed the decision tree classifier in terms of the F_2 score for the re-emergence class for all diseases. Table 1 shows the performance of random forest classifiers in terms of average and SD of precision, recall, F_1, and F_2 measures over 10 nested cross-validations. For the specific diseases in RED Alert, the models were able to identify 82%-90% of all potential re-emergence events as potential re-emergence cases. Of all instances classified as potential re-emergence, about 19% to 31% (except 46% for dengue) were false positives. Our models identified most of the country-level re-emergence events identified in the literature while missing a few events that were restricted to smaller geographic areas and did not contribute enough disease cases to affect disease incidence at the country level. In some cases, our models also identified earlier disease re-emergence events as compared to the literature, underscoring the utility of our models for early detection and warning.

Table 1. Random forest performance over 10 nested cross-validation.[a]

Measure and class	Measles	Cholera	Dengue	Yellow fever
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Precision				
RED	0.7100 (0.1015)	0.8100 (0.1197)	0.5411 (0.0436)	0.6914 (0.1270)
Not RED	0.9925 (0.0057)	0.9913 (0.0063)	0.9883 (0.0040)	0.9964 (0.0036)
Recall				
RED	0.9064 (0.0736)	0.8236 (0.1267)	0.8421 (0.0554)	0.8856 (0.1130)
Not RED	0.9689 (0.0147)	0.9889 (0.0098)	0.9480 (0.0117)	0.9857 (0.0087)
F_1				
RED	0.7909 (0.0688)	0.8051 (0.0814)	0.6567 (0.0439)	0.7631 (0.0752)
Not RED	0.9805 (0.0075)	0.9901 (0.0046)	0.9677 (0.0060)	0.9910 (0.0037)
F_2				
RED	0.8546 (0.0601)	0.8129 (0.0971)	0.7557 (0.0425)	0.8278 (0.0781)
Not RED	0.9735 (0.0118)	0.9893 (0.0074)	0.9558 (0.0094)	0.9878 (0.0066)

[a]RED and not RED represent re-emergence and non-re-emergence classes, respectively.

Evaluation of RED Alert Through Case Studies

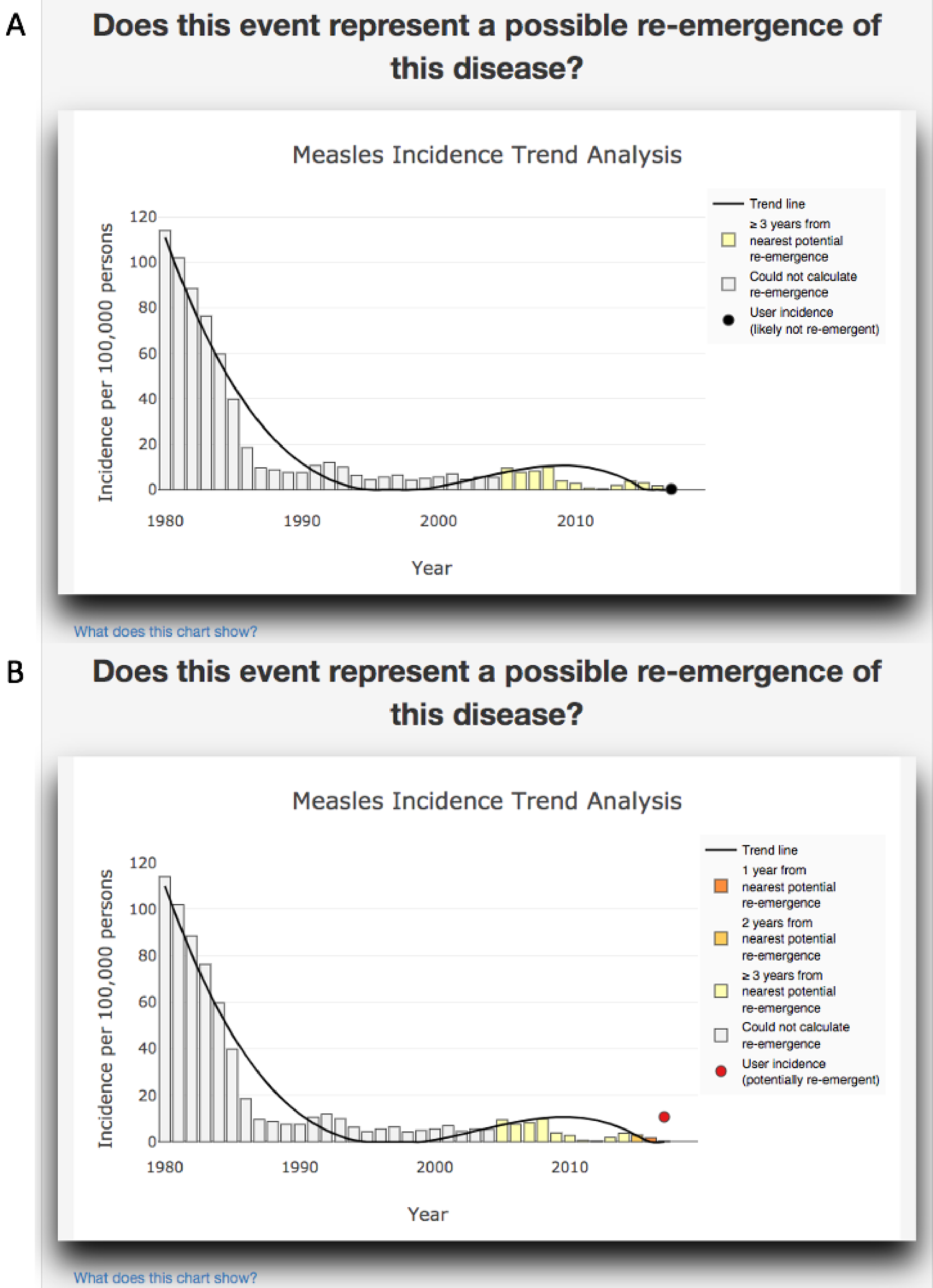
RED Alert features 2 primary modes for users to engage with the application: cumulative and historical analysis. The modes depend on the user's access to data and the user's willingness to upload data into the application. It is important to note that any data the user inputs in the form is not stored by the application at any point. The lowest burden mode to the user is the historical mode. This mode displays all historic data as calculated incidence for the user's defined location. The cumulative mode is of moderate complexity and is the most frequently utilized option in RED Alert. This mode requires that the users know the year they are interested in analyzing as well as the corresponding case counts. For each disease, the analytic provides the most appropriate data source depending on the location. A user selects the cumulative mode if he/she intends to utilize the tool to explore how the data relate to the historic collection of case counts. We describe the results of using RED Alert for a case study for measles. We describe additional case studies for yellow fever, cholera, and dengue in Multimedia Appendix 1. The tool is very rich in information and data, and wherever possible, we have tried to evaluate how the different facets of the analytic could support different types of analysis.

For the measles case study, we specified a public health analyst as the work role and identified the following task for the analyst: Determine the historical profile of measles in China over the past several decades to review the natural temporal fluctuations in measles, and determine if the reported case count for China in 2017 is indicative of a re-emergence. Following the selection of measles from the drop-down menu on the first tab (Figure 1A), the first image seen was a sunburst chart (Figure 1B) that provided the user information on the various causes of re-emergence of measles. The causes were broadly categorized into host, pathogen, and environmental causes, and the user could obtain further detailed information for each of these causes. For example, one of the pathogen-specific factors leading to re-emergence is a new measles type introduced into an endemic country. The following case study inputs were used to generate answers to the 3 main questions used for evaluating the tool: Location—"China," population data source—Default World Bank), mode—cumulative, year of interest—2017, number of cases—3940.

The output was seen in a tabbed format, with the first tab "RED summary" showing the answers to the 3 primary questions:

Q1, "Does this event represent a possible re-emergence of this disease?": The time series (Figure 2A) showed a dip in incidence in 2011 and 2012 followed by a slight rise in cases in 2013 and 2014 and a steady decrease since then. The legend on the chart also indicated that the input data did not reflect a potential re-emergence. When the case count was changed to 150,000, the chart did change (Figure 2B) and the legend on the chart indicated a potential re-emergence together with a red dot on the chart.

Figure 2. Measles incidence trend analysis and re-emergence detection for China. Panel A shows the case corresponding to 3940 measles cases in China in 2017 (as in the case study). The model does not predict re-emergence for this case. Panel B shows the case when the number of measles cases in China are increased to 150,000 in 2017. As this represents a large change in incidence, the model predicts potential re-emergence.



Q2, "What are potential contributing factors?": A summary table (Figure 3) showed the range of factors that potentially contribute to re-emergence, including the values of public health indicators that map to causes of RED for measles, for both the user input year and the median for the recent history (2000 to present). Harmful or protective values were colored red or blue.

Figure 3. Potential contributing factors for measles in China.

What are potential contributing factors?

Indicator Summary Table

Indicator	User value	Median (25%-75% quantile) (2000-2017)	Related to
Current health expenditure per capita (current US$)	398.33 (in 2016)	132.79 (63.70 - 283.52)	Access to healthcare
Improved sanitation facilities (% of population with access)	76.50 (in 2015)	67.80 (63.38 - 72.20)	Sanitation
Immunization, measles (% of children ages 12-23 months)	99.00	99.00 (86.00 - 99.00)	Individuals are too young for vaccinations
Adults (ages 15+) living with HIV	No data	No data	No Data
Incidence of tuberculosis (per 100,000 people)	63.00	82.50 (71.00 - 94.25)	Immune system strength

Table key:

Protective, above 90%/ below 10%

Protective, above 75%/ below 25%

Harmful, above 90%/ below 10%

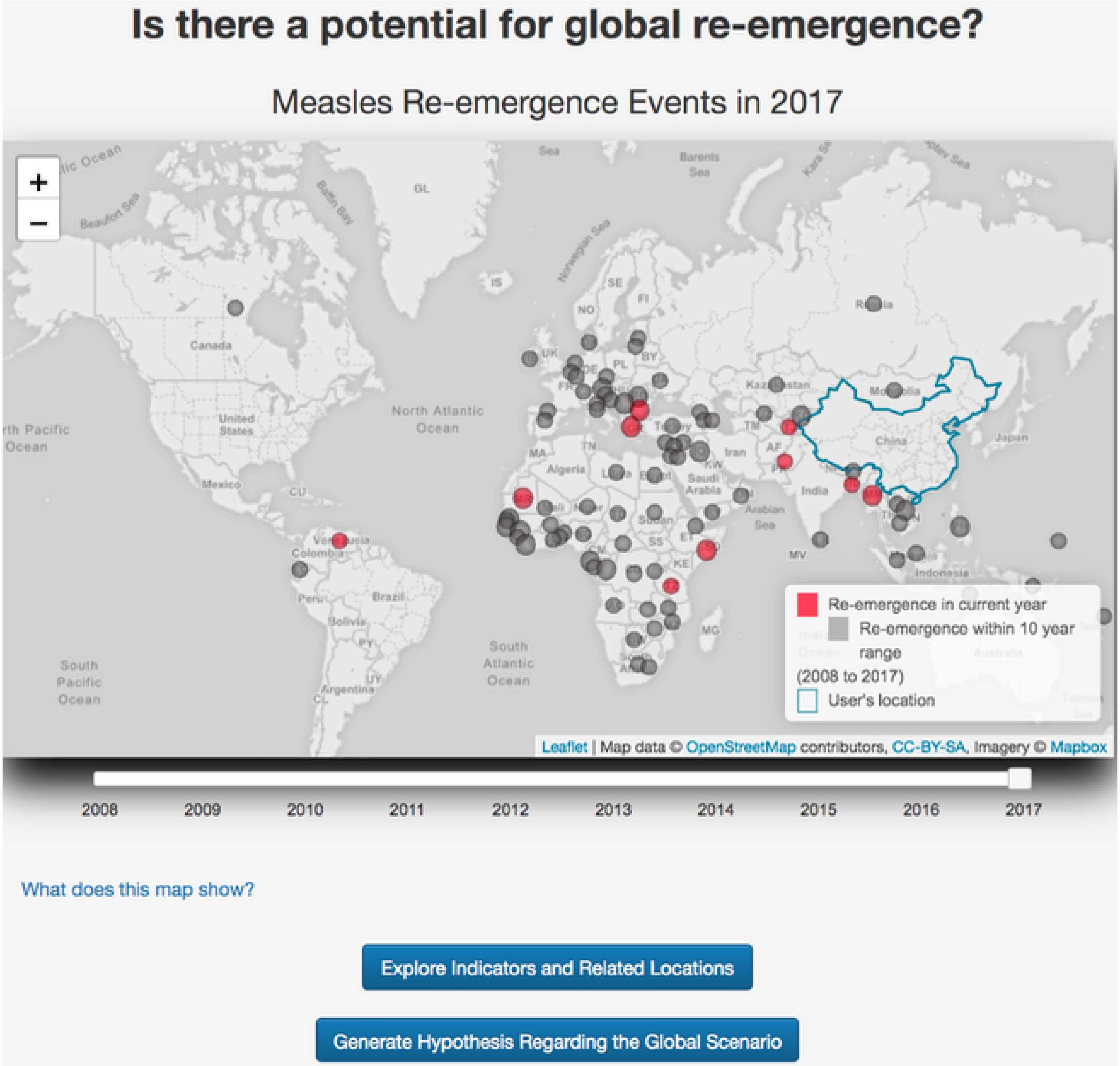
Harmful, above 75%/ below 25%

Normal data, within 25%-75%

What does this table show?

Q3, "Is there a potential for global re-emergence?": A dynamic review of the past 10 years from input year (Figure 4) showed that global re-emergence likely began around the 2008-2009 timeframe. Interestingly, most experts identified global re-emergence around the 2011-2012 timeframe, indicating that RED Alert could have provided earlier warning. Within the past 5 years from 2017, several countries showed re-emergence of measles, but the geographic distribution was concentrated in Eastern Europe and Africa. Myanmar and Bangladesh, which border China, experienced potential re-emergence of measles in 2017, but the disease did not travel across the border to elicit a similar disease event in China.

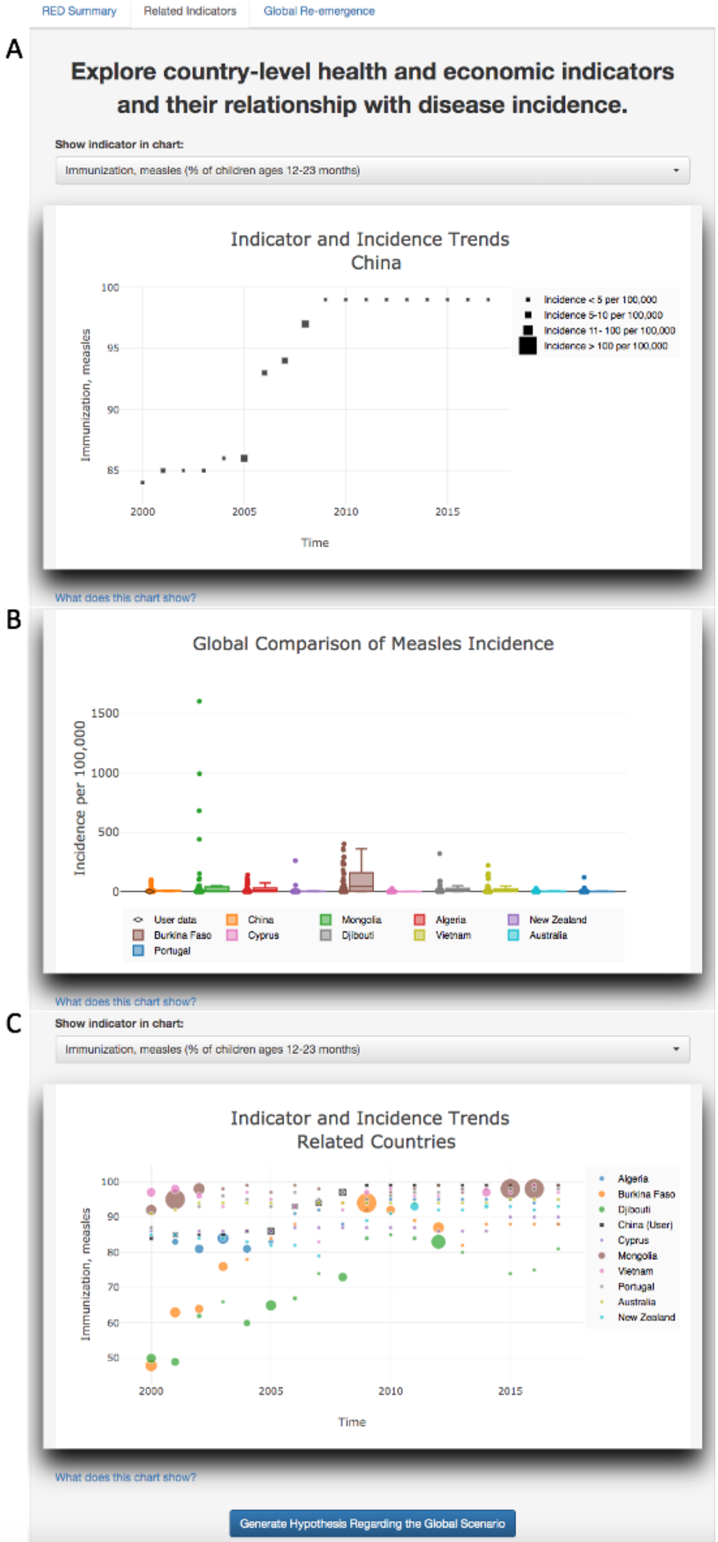
Figure 4. Distribution of national measles re-emergence events worldwide.



Thus, RED Alert was able to successfully address the primary objectives for which it was developed, and provide actionable information.

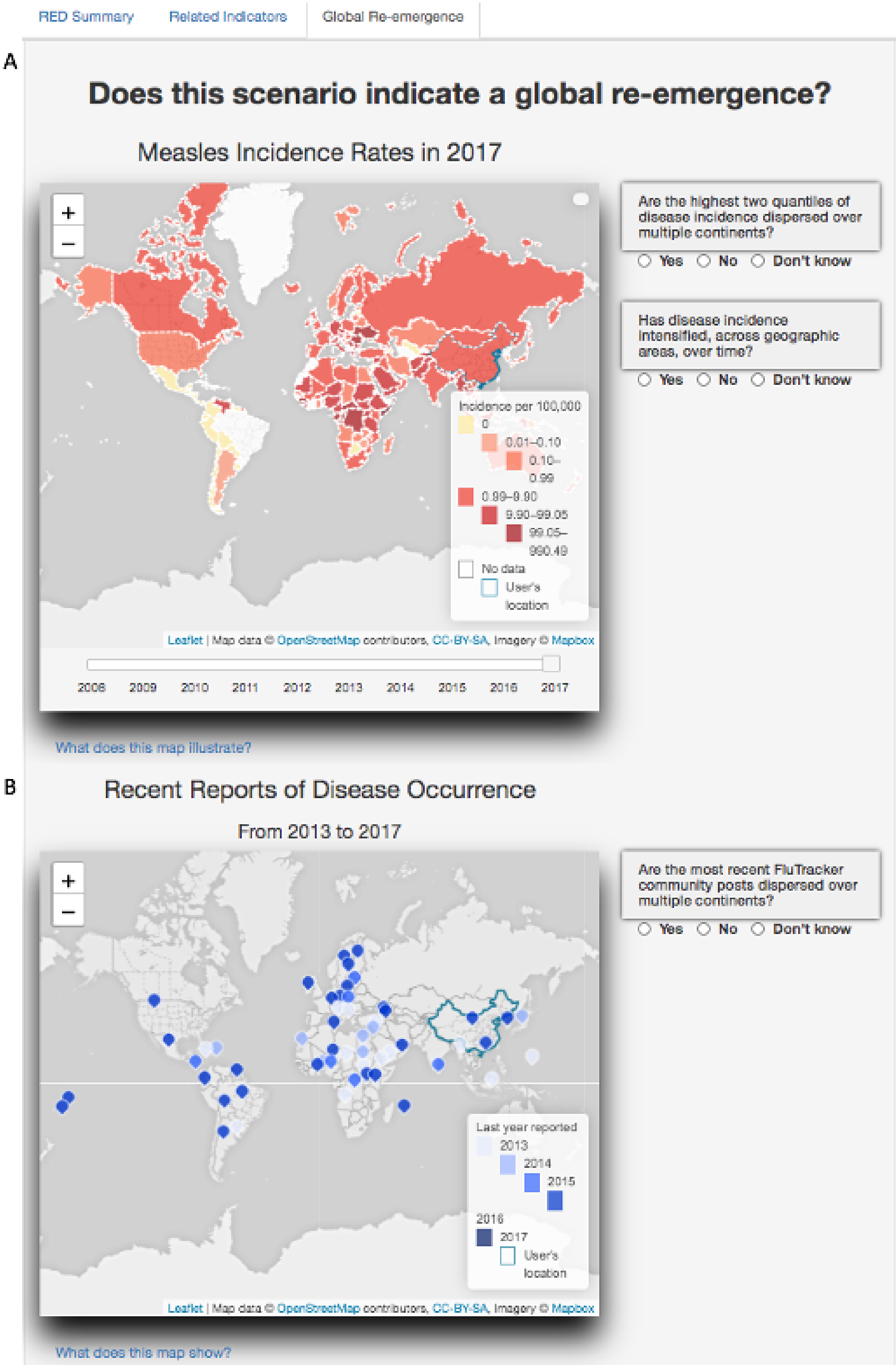
The output of the additional analytics was examined on the “related indicators” tab. Selection of “Immunization, measles (% of children ages 12-23 months)” for the first plot (Figure 5A) on this tab showed that the measles immunization rate exceeded 90% and was maintained above the 90% threshold since 2006, offering a potential reason as to why re-emergence was not identified in China. The comparative boxplot (Figure 5B) showed the countries that had an incidence between 50% and 150% of China’s incidence in 2017, offering a global context. For example, the chart showed that New Zealand and China had very similar incidence perhaps due to similar vaccination rates. This hypothesis could be validated by the selection of “Immunization, measles (% of children ages 12-23 months)” above the third plot (Figure 5C), which showed incidence rates and vaccination coverage to be similar within the past 5 years for New Zealand and China.

Figure 5. RED Alert's related indicators tab. Panel A shows relationship between measles incidence and measles vaccination for China. Panel B shows global comparison of measles incidence. Panel C shows indicator and measles incidence trend for related countries.



Finally, the utility of visual analytics to understand the global scenario of re-emergence was examined on the "Global Re-emergence" tab. The first global map (Figure 6A) showed the incidence in 2017 and the highest incidence values in Africa. A dynamic review of the past 10 years showed that the incidence was globally higher 5 years before 2017. The second global map showed that measles had been discussed on the international disease bulletin website FluTrackers across all continents within the past 2 years (Figure 6B). These maps provided a context to the 2017 China situation and indicated that global re-emergence of measles has occurred much earlier.

Figure 6. RED Alert's global re-emergence tab. Panel A shows measles incidence worldwide. Panel B shows recent reports of disease occurrence for measles based on a disease bulletin.



Discussion

Principal Findings

In this paper, we presented RED Alert, a web-based tool that can provide early warning and detection of infectious disease re-emergence (not disease emergence). It is designed to help public health analysts detect and understand disease re-emergence at both the local (ie, country level) and the global scale through contextual data analysis. It uses supervised machine learning models to detect local disease re-emergence and visual data analytics to help identify and explore potential factors contributing to this re-emergence and assess global situation for potential disease re-emergence. Consistent with our goal of identifying all potential cases of disease re-emergence events while allowing some false positives, our supervised learning models were able to classify 82%-90% of the re-emergence cases, however, with 19% to 31% (except 46% for dengue) false positives. A detailed evaluation of the models used for re-emergence detection is described in [9]. We have also evaluated the utility of the tool through a number of case studies. RED Alert contains all the relevant information to not only provide early warning for potential re-emergence of disease locally and globally, but also offers causes for it. Through the diverse visual presentations and data at their fingertips, RED Alert allows users to verify their hypotheses about local and global re-emergence, and thus facilitates decision making in real-time. A user can access this tool as a one-stop shop for both data and relevant analyses and write a complete report.

While there are a number of online tools for disease surveillance [15,16,18], to the best of our knowledge, this is the first tool that is designed specifically for re-emerging diseases and focuses on detection of potential re-emergence at both local and global level as well as identification of potential contributing factors for the local re-emergence event.

Prior work in disease re-emergence has focused on the contributing factors of re-emergence. In particular, recent work has focused on the tremendous impact of climate change [49,50]. Changes to the climate impact almost every facet of disease transmission from increasing the habitat of disease vectors [51] to increasing the threat of civil unrest and violence [52], which in turn destabilizes infrastructure necessary for resiliency to re-emergence. To complicate this, it is clear that human factors such as urbanization and international travel also impact disease re-emergence [5-8]. However, despite the fact that the literature is clear that there is a complex system at work, the authors have not been able to find any other work in the data fusion or visualization space to allow public health experts to actually interact with the components necessary. Indeed, it is because of this complex milieu that RED Alert is necessary.

Our hope is that RED Alert can provide actionable information to public health analysts and decision makers that can be used for planning purposes. Our tool can provide indications that disease re-emergence may be occurring in a given region (or globally) and also help inform the user of possible contributing factors. This information may be useful in helping better understand the situation, as well as helping determine possible mitigations.

Currently, the tool has data for 4 diseases at the country level and yearly time scale. However, our methodology is applicable to other diseases, as well as other spatial and temporal scales. In addition, although the current application is designed for use on a laptop or desktop computer, we are currently also developing a mobile app for this tool.

Limitations

RED Alert is the first tool designed for detecting and understanding disease re-emergence and provides novel analysis. However, it relies on the availability and quality of data, which depends upon the public health infrastructure of the country. Under-reporting is common in biosurveillance systems [53]. While there are some missing data, historical data collected by the tool are relatively complete. By contrast, there is often some delay in reporting case counts data to agencies such as the WHO or PAHO or data collection companies such as Gideon. Similarly, there is also some delay in the estimation and availability of population and related indicators on the World Bank website. This often leads to missing data for many countries for a couple of recent years. To deal with this, we allow users to input recent case counts data and use values from the latest available years for population and related indicators for analysis purpose. We believe this is reasonable, as this information is less likely to change significantly in a short period. However, discrepancies in the data may affect our analysis.

While it is common in machine learning applications for humans to label data, due to the lack of the concrete definition of re-emergence, labeling is a subjective assessment. It may be possible that the SMEs in our team mislabeled data in some cases. Further, due to the lack of a concrete quantitative definition of re-emergence, it is difficult to fully validate our analysis.

Future Directions

There are many opportunities for future work including adding more diseases to the tool based on their likelihood of re-emerging. Currently, the ability to perform the same analysis at a subnational level is mainly restricted by data availability. We are working on obtaining data at subnational levels for a few diseases and countries and plan to make this functionality available through a mobile app. Re-emergence detection models can also be improved by using other disease-related factors such as weather or climate data for mosquito-borne diseases, as mosquito density depends upon temperature and humidity.

Acknowledgments

This work was supported by the U.S. Department of Energy through Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy (Contract No. 89233218CNA000001). This work was funded by the Defense Threat Reduction Agency (DTRA) (grant #10027). Dr Ramesh Krishnamurthy, along with other team members of the WHO's Department of Information, Evidence, and Research in the Health Systems and Innovation Cluster, provided subject matter expertise on disease re-emergence as a global phenomenon, as well as detailed insight into the formative stages of the study. Drs Bryan Lewis and James Schlitt, Biocomplexity Institute, University of Virginia, were important contributors to understanding the spatial components of disease re-emergence.

Authors' Contributions

NP, ARD, WR, MC, FA, NV, and GF collected the data. NP, ARD, WR, and GF cleaned and ingested the data. WR and GF designed the back end and ARD, WR, DA, and GF developed the front end. All authors helped design the front end/visualizations. ARD, MC, FA, and NV labeled the data for classification. NP and GF trained the classifiers and integrated them with the tool. AD conceived and led the project. NP and AD wrote the first version of the manuscript and all authors reviewed it.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional case studies for cholera, dengue, and yellow fever to show utility of RED Alert in detecting and understanding disease re-emergence.
[DOCX File , 25478 KB - publichealth_v7i1e24132_app1.docx]

Multimedia Appendix 2
Data Sources for RED Alert.
[DOCX File , 16 KB - publichealth_v7i1e24132_app2.docx]

References

1. The top 10 causes of death. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2020-12-17]
2. Teixeira MG, Costa MDCN, Barreto F, Barreto ML. Dengue: twenty-five years since reemergence in Brazil. Cad Saude Publica 2009;25 Suppl 1:S7-18 [FREE Full text] [doi: 10.1590/s0102-311x2009001300002] [Medline: 19287868]
3. Antona D, Lévy-Bruhl D, Baudon C, Freymuth F, Lamy M, Maine C, et al. Measles elimination efforts and 2008-2011 outbreak, France. Emerg Infect Dis 2013 Mar;19(3):357-364 [FREE Full text] [doi: 10.3201/eid1903.121360] [Medline: 23618523]
4. Woodall J, Yuill T. Why is the yellow fever outbreak in Angola a 'threat to the entire world'? Int J Infect Dis 2016 Jul;48:96-97 [FREE Full text] [doi: 10.1016/j.ijid.2016.05.001] [Medline: 27163382]
5. Barrett R, Kuzawa CW, McDade T, Armelagos GJ. Emerging and Re-emerging Infectious Diseases: The Third Epidemiologic Transition. Annu. Rev. Anthropol 1998 Oct 21;27(1):247-271. [doi: 10.1146/annurev.anthro.27.1.247]
6. Fauci AS. Emerging and reemerging infectious diseases: the perpetual challenge. Acad Med 2005 Dec;80(12):1079-1085. [doi: 10.1097/00001888-200512000-00002] [Medline: 16306276]
7. Harrus S, Baneth G. Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases. Int J Parasitol 2005 Oct;35(11-12):1309-1318. [doi: 10.1016/j.ijpara.2005.06.005] [Medline: 16126213]
8. Plans P, Torner N, Godoy P, Jané M. Lack of herd immunity against measles in individuals aged <35 years could explain re-emergence of measles in Catalonia (Spain). Int J Infect Dis 2014 Jan;18:81-83 [FREE Full text] [doi: 10.1016/j.ijid.2013.09.015] [Medline: 24211476]
9. Chitanvis M, Daughton AR, Altherr F, Parikh N, Fairchild G, Rosenberger W, et al. Development of a Supervised Learning Algorithm for Detection of Potential Disease Reemergence: A Proof of Concept. Health Secur 2019;17(4):255-267. [doi: 10.1089/hs.2019.0020] [Medline: 31433278]
10. Staples J, Breiman R, Powers A. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. Clin Infect Dis 2009 Sep 15;49(6):942-948. [doi: 10.1086/605496] [Medline: 19663604]
11. Hartskeerl R, Collares-Pereira M, Ellis W. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. Clin Microbiol Infect 2011 Apr;17(4):494-501 [FREE Full text] [doi: 10.1111/j.1469-0691.2011.03474.x] [Medline: 21414083]
12. Seleem M, Boyle S, Sriranganathan N. Brucellosis: a re-emerging zoonosis. Vet Microbiol 2010 Jan 27;140(3-4):392-398. [doi: 10.1016/j.vetmic.2009.06.021] [Medline: 19604656]
13. Gubler DJ. The Global Threat of Emergent/Re-emergent Vector-Borne Diseases. In: Atkinson PW, editor. Vector Biology, Ecology and Control. Dordrecht, The Netherlands: Springer; 2010:39-62.
14. Generous N, Fairchild G, Khalsa H, Tasseff B, Arnold J. Epi Archive: automated data collection of notifiable disease data. OJPHI 2017;9(1). [doi: 10.5210/ojphi.v9i1.7615]
15. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. J Am Med Inform Assoc 2008;15(2):150-157 [FREE Full text] [doi: 10.1197/jamia.M2544] [Medline: 18096908]
16. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, et al. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 2008 Dec 15;24(24):2940-2941 [FREE Full text] [doi: 10.1093/bioinformatics/btn534] [Medline: 18922806]

17. Velappan N, Daughton AR, Fairchild G, Rosenberger WE, Generous N, Chitanvis ME, et al. Analytics for Investigation of Disease Outbreaks: Web-Based Analytics Facilitating Situational Awareness in Unfolding Disease Outbreaks. JMIR Public Health Surveill 2019 Feb 25;5(1):e12032 [FREE Full text] [doi: 10.2196/12032] [Medline: 30801254]
18. Codeço CT, Cruz OG, Riback TI, Degener CM, Gomes MF, Villela D, et al. InfoDengue: a nowcasting system for the surveillance of dengue fever transmission. bioRxiv. 2016. URL: https://www.biorxiv.org/content/10.1101/046193v1.full.pdf [accessed 2019-09-30]
19. Wang J, McMichael AJ, Meng B, Becker NG, Han W, Glass K, et al. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. Bull World Health Organ 2006 Dec;84(12):965-968 [FREE Full text] [doi: 10.2471/blt.06.030247] [Medline: 17242832]
20. Beaulieu-Jones BK, Greene CS. Semi-supervised learning of the electronic health record for phenotype stratification. J Biomed Inform 2016 Dec;64:168-178 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.007] [Medline: 27744022]
21. Bose E, Radhakrishnan K. Using Unsupervised Machine Learning to Identify Subgroups Among Home Health Patients With Heart Failure Using Telehealth. Comput Inform Nurs 2018 May;36(5):242-248. [doi: 10.1097/CIN.0000000000000423] [Medline: 29494361]
22. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2) [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
23. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis 2017 Oct;11(10) [FREE Full text] [doi: 10.1371/journal.pntd.0005973] [Medline: 29036169]
24. Rose S. Mortality risk score prediction in an elderly population using machine learning. Am J Epidemiol 2013 Mar 01;177(5):443-452. [doi: 10.1093/aje/kws241] [Medline: 23364879]
25. Chiang P, Dey S. Personalized Effect of Health Behavior on Blood Pressure: Machine Learning Based Prediction and Recommendation. 2018 Presented at: IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom); 2018; Ostrava, Czech Republic p. 1-6. [doi: 10.1109/healthcom.2018.8531109]
26. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
27. Thomas J, Cook K. Illuminating the Path: The Research and Development Agenda for Visual Analytics. United States: National Visualization and Analytics Ctr; 2005.
28. Youn-Ah Kang, Görg C, Stasko J. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. IEEE Transactions on Visualization and Computer Graphics 2011 May;17(5):570-583. [doi: 10.1109/TVCG.2010.84] [Medline: 20530814]
29. Simpao AF, Ahumada LM, Rehman MA. Big data and visual analytics in anaesthesia and health care. Br J Anaesth 2015 Sep;115(3):350-356 [FREE Full text] [doi: 10.1093/bja/aeu552] [Medline: 25627395]
30. Maciejewski R, Tyner B, Jang Y, Zheng C, Nehme R, Ebert D, et al. LAHVA: Linked Animal-Human Health Visual Analytics. 2007 Presented at: IEEE Symposium on Visual Analytics Science and Technology; 2007; Sacramento, CA, p. 27-34. [doi: 10.1109/VAST.2007.4388993]
31. Raghupathi W, Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. Int J Environ Res Public Health 2018 Mar 01;15(3) [FREE Full text] [doi: 10.3390/ijerph15030431] [Medline: 29494555]
32. Perer A, Sun J. MatrixFlow: Temporal Network Visual Analytics to Track Symptom Evolution during Disease Progression. AMIA Annu Symp Proc 2012:716-725. [Medline: 2330434]
33. RED Alert, Re-emerging Infectious Disease Alert. URL: https://redalert.bsvgateway.org/ [accessed 2020-12-20]
34. WHO vaccine-preventable diseases: monitoring system. 2020 global summary.: World Health Organization URL: http://apps.who.int/immunization_monitoring/globalsummary [accessed 2020-12-19]
35. Global Health Observatory data repository; Number of reported cases Data by country.: World Health Organization URL: http://apps.who.int/gho/data/node.main.175?lang=en [accessed 2019-09-29]
36. DengueNet; Welcome to the DengueNet database and geographic information system.: World Health Organization URL: http://apps.who.int/globalatlas/default.asp [accessed 2019-09-29]
37. Gideon. Gideon URL: https://www.gideononline.com/ [accessed 2020-12-20]
38. Dengue and Severe Dengue Cases and Deaths for countries and territories of the Americas.: Pan American Health Organization (PAHO) URL: https://www.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/257-dengue-casos-muertes-pais-ano-en.html [accessed 2019-09-29]
39. LandScan.: Oak Ridge National Laboratory URL: https://landscan.ornl.gov/ [accessed 2020-12-20]
40. DataBank; Population estimates and projections.: The World Bank URL: http://databank.worldbank.org/data/reports.aspx?source=population-estimates-and-projections [accessed 2019-09-29]
41. Measles-containing vaccine.: World Health Organization URL: http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tscoveragemcv1.html [accessed 2019-09-29]
42. Working with the regions.: World Health Organization URL: https://www.who.int/chp/about/regions/en/ [accessed 2020-12-22]

43. Dicker R, Coronado F, Koo D, Parrish RG. Introduction to Epidemiology. In: Principles of Epidemiology in Public Health Practice. Atlanta, GA: Centers for Disease Control and Prevention; 2006.
44. Indicators.: The World Bank URL: https://data.worldbank.org/indicator [accessed 2019-09-29]
45. About the Indicators API Documentation.: The World Bank URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/889392-apidocumentation [accessed 2019-09-29]
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825-2830.
47. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006 Feb 23;7 [FREE Full text] [doi: 10.1186/1471-2105-7-91] [Medline: 16504092]
48. FluTrackers.com, Tracking Infectious Diseases since 2006. URL: https://flutrackers.com/forum/ [accessed 2019-09-29]
49. Zell R. Global climate change and the emergence/re-emergence of infectious diseases. International Journal of Medical Microbiology Supplements 2004 Apr;293:16-26. [doi: 10.1016/s1433-1128(04)80005-6]
50. El-Sayed A, Kamel M. Climatic changes and their role in emergence and re-emergence of diseases. Environ Sci Pollut Res Int 2020 Jun;27(18):22336-22352 [FREE Full text] [doi: 10.1007/s11356-020-08896-w] [Medline: 32347486]
51. Rochlin I, Ninivaggi DV, Hutchinson ML, Farajollahi A. Climate Change and Range Expansion of the Asian Tiger Mosquito (Aedes albopictus) in Northeastern USA: Implications for Public Health Practitioners. PLoS ONE 2013 Apr 2;8(4). [doi: 10.1371/journal.pone.0060874]
52. Sofuoğlu E, Ay A. The relationship between climate change and political instability: the case of MENA countries (1985:01–2016:12). Environmental Science and Pollution Research 2020 Feb 8;27:14033-14043. [doi: 10.1007/s11356-020-07937-8]
53. Gibbons CL, Mangen MJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health 2014 Feb 11;14 [FREE Full text] [doi: 10.1186/1471-2458-14-147] [Medline: 24517715]

Abbreviations

API: application programming interface
PAHO: Pan American Health Organization
SME: subject matter expert
WHO: World Health Organization

Edited by T Sanchez; submitted 04.09.20; peer-reviewed by H Cousins, H Kadir; comments to author 25.09.20; revised version received 26.10.20; accepted 14.12.20; published 07.01.21.

Please cite as:
Parikh N, Daughton AR, Rosenberger WE, Aberle DJ, Chitanvis ME, Altherr FM, Velappan N, Fairchild G, Deshpande A
Improving Detection of Disease Re-emergence Using a Web-Based Tool (RED Alert): Design and Case Analysis Study
JMIR Public Health Surveill 2021;7(1):e24132
URL: http://publichealth.jmir.org/2021/1/e24132/
doi:10.2196/24132
PMID:33316766

Original Paper

How Health Care Workers Wield Influence Through Twitter Hashtags: Retrospective Cross-sectional Study of the Gun Violence and COVID-19 Public Health Crises

Ayotomiwa Ojo[1*], BS; Sharath Chandra Guntuku[2,3,4*], PhD; Margaret Zheng[4,5], BA; Rinad S Beidas[5], PhD; Megan L Ranney[6], MD, MPH

[1]Harvard Medical School, Boston, MA, United States
[2]Penn Medicine Center for Digital Health, Philadelphia, PA, United States
[3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States
[4]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States
[5]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States
[6]Brown-Lifespan Center for Digital Health, Brown University, Providence, RI, United States
[*]these authors contributed equally

Corresponding Author:
Megan L Ranney, MD, MPH
Brown-Lifespan Center for Digital Health
Brown University
139 Point St
Providence, RI, 02903
United States
Phone: 1 646 644 3053
Email: megan_ranney@brown.edu

Abstract

Background: Twitter has emerged as a novel way for physicians to share ideas and advocate for policy change. #ThisIsOurLane (firearm injury) and #GetUsPPE (COVID-19) are examples of nationwide health care–led Twitter campaigns that went viral. Health care–initiated Twitter hashtags regarding major public health topics have gained national attention, but their content has not been systematically examined.

Objective: We hypothesized that Twitter discourse on two epidemics (firearm injury and COVID-19) would differ between tweets with health care–initiated hashtags (#ThisIsOurLane and #GetUsPPE) versus those with non–health care–initiated hashtags (#GunViolence and #COVID19).

Methods: Using natural language processing, we compared content, affect, and authorship of a random 1% of tweets using #ThisIsOurLane (Nov 2018-Oct 2019) and #GetUsPPE (March-May 2020), compared to #GunViolence and #COVID19 tweets, respectively. We extracted the relative frequency of single words and phrases and created two sets of features: (1) an open-vocabulary feature set to create 50 data-driven–determined word clusters to evaluate the content of tweets; and (2) a closed-vocabulary feature for psycholinguistic categorization among case and comparator tweets. In accordance with conventional linguistic analysis, we used a $P<.001$, after adjusting for multiple comparisons using the Bonferroni correction, to identify potentially meaningful correlations between language features and outcomes.

Results: In total, 67% (n=4828) of #ThisIsOurLane tweets and 36.6% (n=7907) of #GetUsPPE tweets were authored by health care professionals, compared to 16% (n=1152) of #GunViolence and 9.8% (n=2117) of #COVID19 tweets. Tweets using #ThisIsOurLane and #GetUsPPE were more likely to contain health care–specific language; more language denoting positive emotions, affiliation, and group identity; and more action-oriented content compared to tweets with #GunViolence or #COVID19, respectively.

Conclusions: Tweets with health care–led hashtags expressed more positivity and more action-oriented language than the comparison hashtags. As social media is increasingly used for news discourse, public education, and grassroots organizing, the public health community can take advantage of social media's broad reach to amplify truthful, actionable messages around public health issues.

(JMIR Public Health Surveill 2021;7(1):e24562) doi:10.2196/24562

KEYWORDS

COVID-19; firearm injury; social media; online advocacy; Twitter; infodemiology; infoveillance; tweet; campaign; health care worker; influence; public health; crisis; policy

Introduction

Twitter has emerged as a novel way for physicians to organize and advocate for policy change, and combat misinformation amid national health crises. One in 5 adults in the United States uses Twitter, and 75% report using this platform as a news outlet [1]. When Twitter advocacy campaigns brand their movement with a hashtag, tagged tweets are easily archived and found, opening up discussions to users who do not have any personal connection to the authors.

#ThisIsOurLane and #GetUsPPE are examples of health care–initiated Twitter movements that went viral. In November 2018, in response to the National Rifle Association's tweet asserting that "Someone should tell self-important anti-gun doctors to stay in their lane…," Dr Michael Gonzalez coined #ThisIsOurLane to describe why health care professionals are involved in firearm injury prevention and treatment [2]. During the COVID-19 pandemic, Dr Esther Choo initiated #GetMePPE, later expanded to #GetUsPPE, to raise awareness about critical personal protective equipment (PPE) shortages [3]. Anecdotes suggest #ThisIsOurLane influenced societal perceptions of health care professionals' role in firearm injury [4], and #GetUsPPE galvanized attention to hospitals' unmet PPE needs [5,6].

Whether online discussions with health care–initiated hashtags actually differ from contemporaneous discussions of the firearm injury and COVID-19 epidemics has not been quantified. Nor, to our knowledge, has the involvement of Twitter users outside of health care been examined. Understanding the content and voice of health professionals on social media during public health crises is essential. Rampant misinformation about health care online has led to international debates about how best to change public knowledge and conversations. At the same time, some experts are bemoaning "infodemics," in which people are so overwhelmed by contradictory facts that they become unable to act to protect themselves and their families [7]. Examining the content, tone, and types of tweeters involved in health care–led social media campaigns could inform future efforts related to data dissemination by the medical and nonmedical community [8].

To examine the characteristics of these online discussions, we compared psycholinguistic characteristics (ie, content and affect) of tweets among two cohorts: contemporaneous tweets regarding gun violence (comparing tweets with #ThisIsOurLane vs #GunViolence) and contemporaneous tweets regarding the COVID-19 pandemic (#GetUsPPE/#GetMePPE vs #COVID19). We hypothesized that messages using health care–led hashtags would be more negative in tone (reflecting frustration and negative directives) but also more actionable in content (providing solutions) compared with non–health care–related hashtags, given health care professionals' personal stake and proximity to these issues.

Methods

This retrospective cross-sectional study selected a random 1% sample of publicly available Twitter data containing specific hashtags from across the United States.

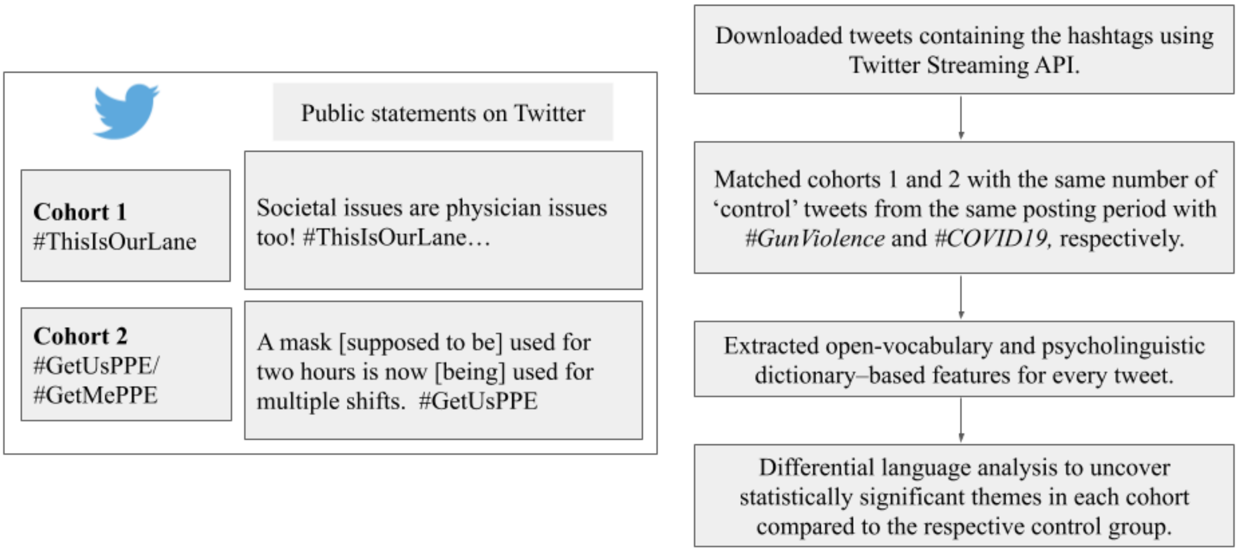
Data

For cohort 1, we identified tweets containing #ThisIsOurLane (n=38,774) or #GunViolence (n=52,183) between November 7, 2018, and October 13, 2019, given multiple episodes of gun violence with national attention during this time period. For cohort 2, we identified tweets with #GetUsPPE or #GetMePPE (n=39,658) or #COVID19 (n=200,000) between March 17, 2020, and May 20, 2020, which reflects the duration of the campaign at the time of the analysis. Both study periods began when the hashtag was introduced. After discarding retweets and tweets containing only hashtags and user mentions (without any other words), 7201 #ThisIsOurLane tweets and 21,605 #GetUsPPE/#GetMePPE tweets remained as "cases". Tweets containing both case and control hashtags were preserved as cases in the analysis. A random sample of 7201 of the remaining #GunViolence-only tweets and 21,605 of the #COVID19-only tweets were selected as comparators for two separate analyses (Figure 1). Although tweets about gun violence and COVID-19 used other hashtags, these were identified as trending and potentially the most common around the study period and were used as comparators.

We used the Python package TwitterMySQL [9], which utilizes the official Twitter application programming interface (API), to collect tweets matching at least one of the keywords described above in real time. We note that the Twitter API limits such streams to 1% of the total Twitter volume at any given moment. Similar methods have been used in prior work studying health-related tweets [10-14].

We obtained Twitter profile descriptions of the users in our data set using the Twitter API and searched for words indicating health care professional status using regular expressions (eg, "doc*," "medic*," "surg*"). When processing tweets for this analysis, we only utilized information publicly available on users' Twitter profiles. The University of Pennsylvania Institutional Review Board deemed the study exempt.

Figure 1. Study flowchart. API: application programming interface.



Extracting Language Features

After tokenizing the tweets [15], we extracted the relative frequency of single words and phrases and created two sets of features: (1) an open vocabulary feature set [16] defined using the MALLET (Machine Learning for Language Toolkit) implementation of latent Dirichlet allocation [17], an unsupervised clustering algorithm, to create 50 data-driven–determined word clusters; and (2) a closed vocabulary feature set defined as the normalized frequency of 71 psycholinguistic categories among case and comparator tweets, created with Linguistic Inquiry Word Count dictionary [18].

Statistical Analysis

Each feature set was input in a logistic regression model, with "case" (ie, #ThisIsOurLane or #GetUsPPE) as the dependent variable. In accordance with conventional linguistic analysis, we used a *P* value of <.001, after adjusting for multiple comparisons using the Bonferroni correction, to identify potentially meaningful correlations between language features and outcomes. We calculated regression coefficients with the #GunViolence and #COVID19 (comparator) groups as references.

Results

In total, 67% (n=4828) of #ThisIsMyLane tweets and 36.6% (n=7907) of #GetUsPPE tweets were authored by health care professionals, compared to 16% (n=1152) of #GunViolence and 9.8% (n=2117) of #COVID19 tweets.

The open-vocabulary feature set (ie, content) of #ThisIsOurLane and #GetUsPPE were more likely to contain language specific to health care than general tweets using hashtags #GunViolence and #COVID19 (Figures 2-5). Specifically, #ThisIsOurLane tweets discussed health care professionals' advocacy, research, or appreciation of colleagues, and were more likely to mention public health and community compared with #GunViolence tweets. #ThisIsOurLane tweets were less likely to mention political entities like #NRA and specific events such as #ElPaso. #GetUsPPE tweets described severe PPE shortages for health care workers, the need to support patient and staff safety, and referenced health care workers as heroes. Additionally, #GetUsPPE tweets included more action-oriented language (ie, deliver, sign, support) compared with #COVID19 tweets.

Analysis of closed-vocabulary associations (ie, psycholinguistic categories) demonstrated that tweets with #ThisIsOurLane or #GetUsPPE contained more language associated with health, positive emotions, affiliation, and group identity compared to tweets with #GunViolence or #COVID19, respectively (Figure 6). General tweets about gun violence and the COVID-19 pandemic contained more words associated with negative emotions or anger than tweets with health care–initiated hashtags.

Figure 2. Words associated with #ThisIsOurLane tweets compared to #GunViolence. Beta indicates the strength of association of each word with respective groups and color indicates frequency. All words are statistically significant at p<.05, Benjamin Hochberg correction.

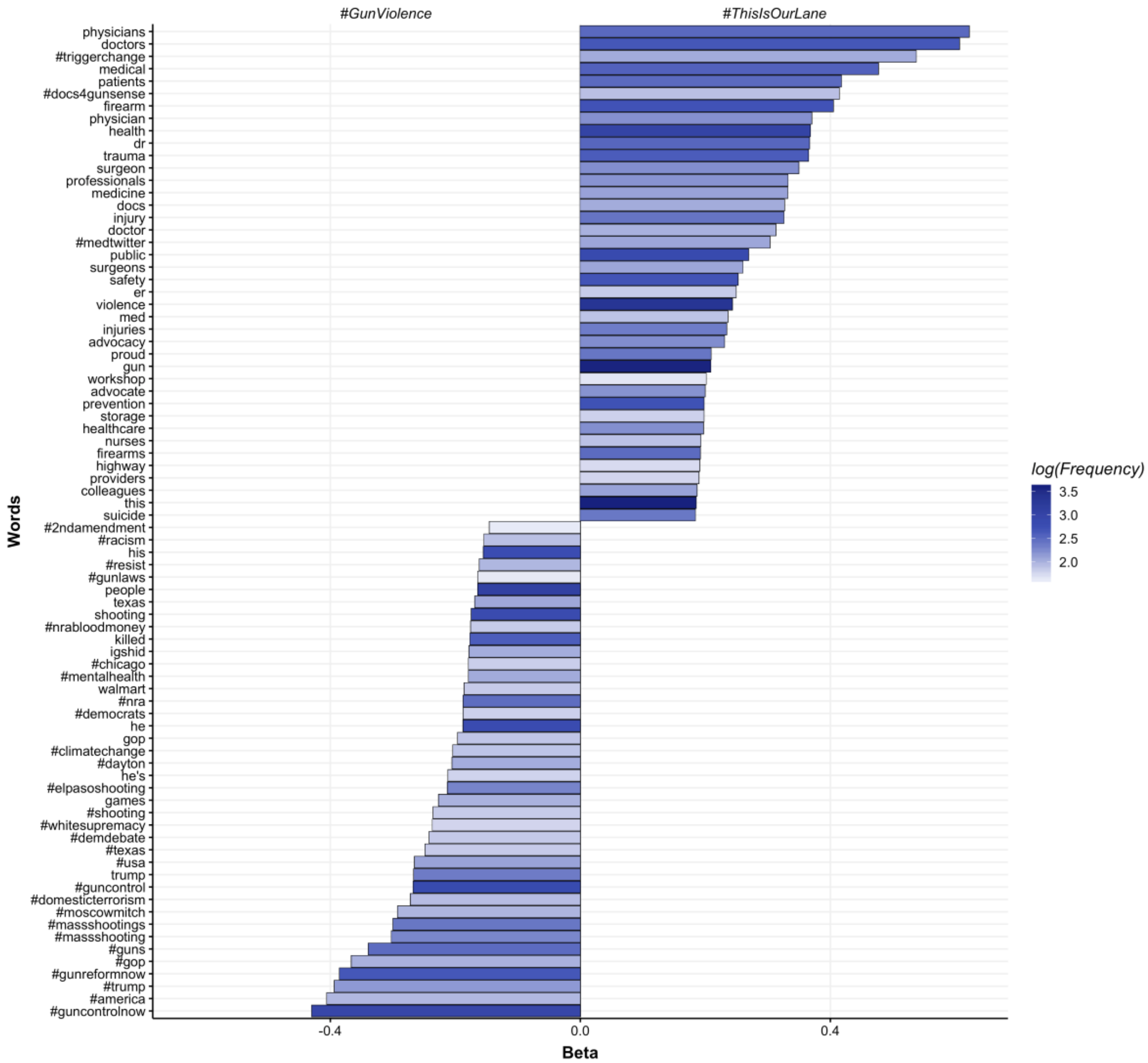


Figure 3. Words associated with #GetUsPPE tweets compared to #COVID19. Beta indicates the strength of association of each word with respective groups and color indicates frequency. All words are statistically significant at p<.05, Benjamin Hochberg correction.

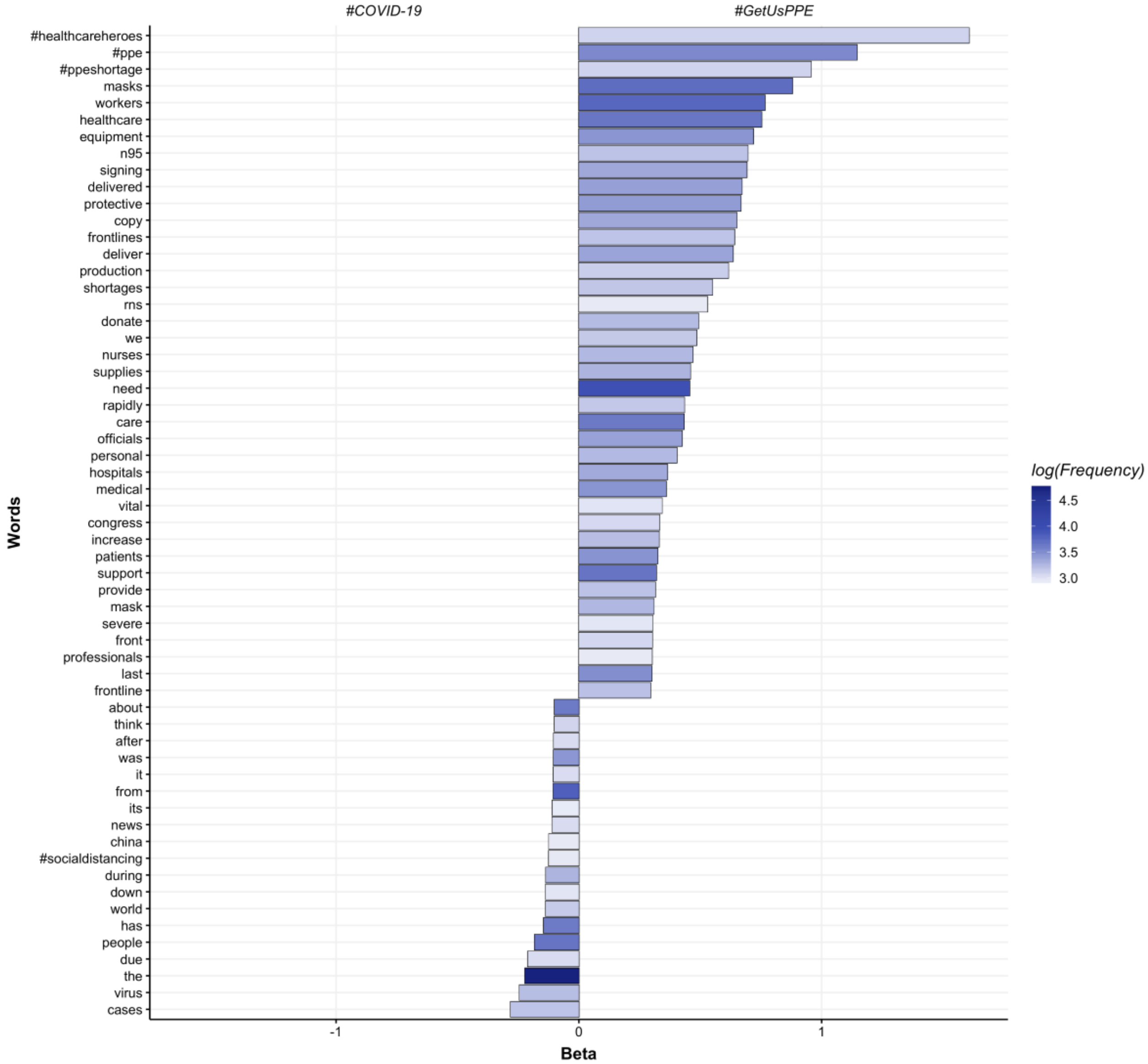


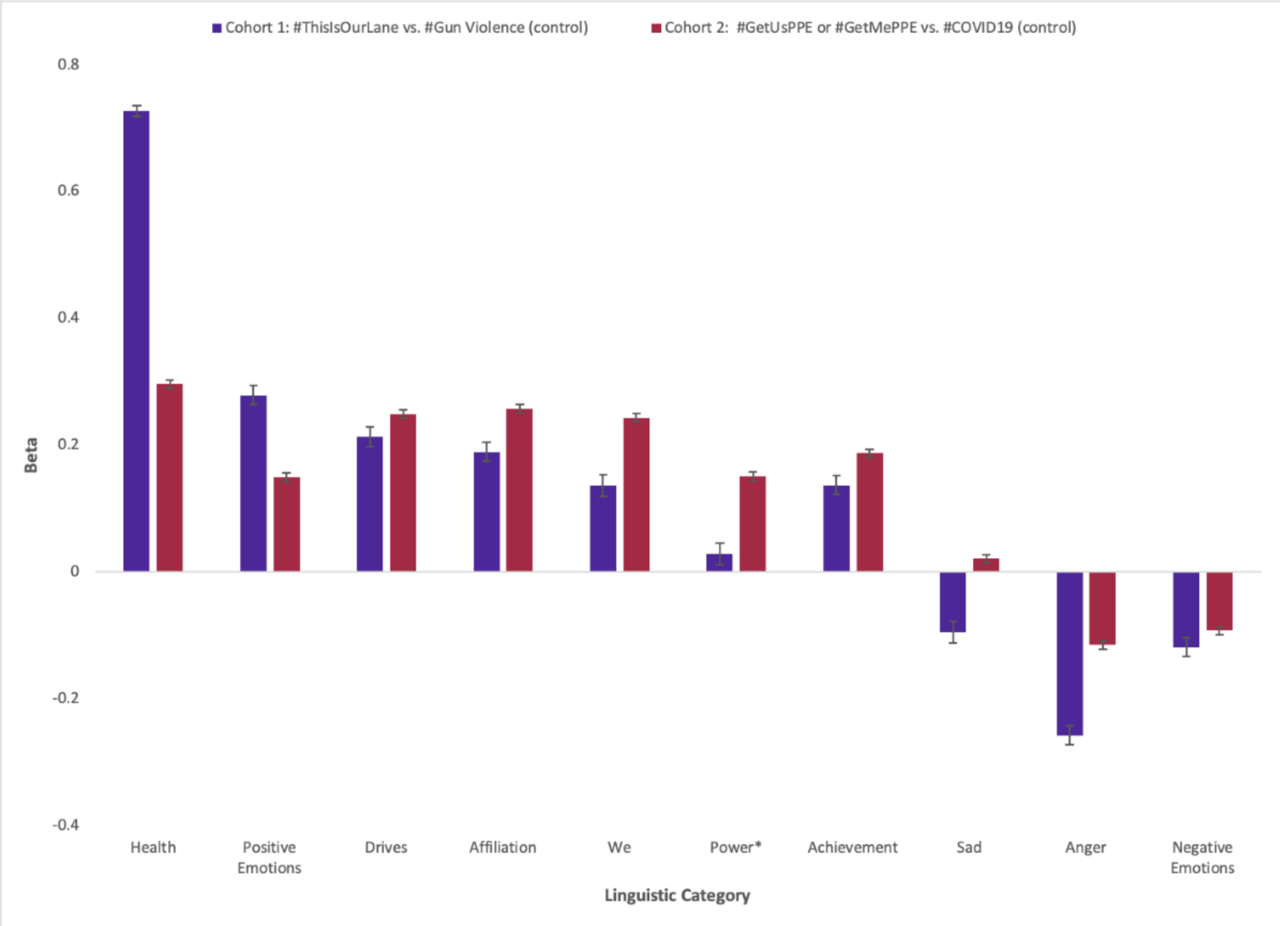
Figure 4. Highly correlated topics with mention of #ThisIsOurLane vs. #GunViolence. Beta indicates the strength of association of each topic. Top words and example paraphrased tweets for each topic are shown. Topics are statistically significant at p<.05, Benjamin Hochberg correction.

Topic Theme	Highly Correlated Words in the Topic	Beta (>0: #ThisIsOurLane; <0: #GunViolence)	Example Tweets
		#ThisIsOurLane	
Medical Profession	trauma, surgeon, surgeons, victims, gunshot, patients, doctors, hospital, wounds, dr	0.819	Emergency room doc sharing gut-wrenching stories of docs and their experiences dealing with gun violence. #ThisIsOurLane
Advocacy	health, patients, care, medical, physicians, #medtwitter, advocacy, medicine, advocate, social	0.773	"Doctors also have a #professional duty to advocate on broader issues affecting health, including tobacco, alcohol, poverty, and many other issues" #ThisIsOurLane #medtwitter <URL>
Prevention	firearm, injury, prevention, dr, research, death, safety, injuries, prevent, deaths	0.764	#ThisIsOURLane American College of Surgeons recommendations for reducing gun injury, death and disability. <URL>
Appreciation	proud, work, great, important, colleagues, stand, working, leadership, story, powerful	0.692	Proud of <USER> & other #emergencymedicine colleagues for representing #ThisIsOurLane #publichealth #BanAssaultWeapons <URL>
Research	research, violence, funding, safety, national, public, approach, data, policy, prevention	0.402	Yes. AND demand a long term comprehensive public health approach to reducing gun violence. Support the creation of a National Bureau for Gun Safety. #thisisourlane <USER> <URL>
		#GunViolence	
Specific Events	#guncontrolnow, #elpasoshooting, #elpaso, #gunreformnow, trump, america, #dayton, #daytonshooting, #gunreform, #whitesupremacistterrorism	-1.097	I am sick of hearing about thoughts and prayers. #gunreformnow #ElPasoShooting #GunReform #NRACarnage #NRA #NRAIsATerroristOrganization #walmartshooting #EPShooting #ElPaso #GunControlNow #GunViolence
	mass, shootings, shooting, paso, el, dayton, #massshooting, texas, 2019, #elpaso	-0.387	In addition to El Paso and Dayton, Virginia and Chicago also recorded mass shootings this weekend: GunDeaths ... #gunviolence
	shooting, people, killed, police, shot, dead, injured, chicago, city, man	-0.413	7 wounded, 1 killed Wednesday in shootings across Chicago <URL> ... #EnoughisEnough #StopChicagoShootings #gunviolence #BlackTwitterMovement
Political Affiliation	vote, #gunreformnow, put, care, nra, money, protect, #gunreform, republicans, americans	-0.685	I weep today for the victims lost due to #GunViolence I weep for the families. I weep for our country and for the soulless GOP that perfer to line their pockets with money rather than put forth #GunControlNow #mitchthemurderer #TrumpsTerrorists #massshootings
Advocacy	#guncontrol, #guncontrolnow, #guns, #gunsense, #nra, #gunreform, #2a, #gunreformnow, #gun, #usa	-1.14	>2.5 million lives are saved each year thru defensive gun use by citizens. #EnforceTheLaw #Walkaway #SecondAmendment #Selfdefense #Rights #2A #NotAboutGuns #GunReform #GunSense #GunControl #GunViolence #GunControlNow #GunBan #GOP #NRA #2AShallNotBeInfringed <URL>

Figure 5. Highly correlated topics with mention of #GetUsPPE vs. #COVID19. Beta indicates the strength of association of each topic. Top words and example paraphrased tweets for each topic are shown. Topics are statistically significant at p<.05, Benjamin Hochberg correction.

Topic Theme	Highly Correlated Words in the Topic	Beta (>0: #GetUsPPE; <0: #COVID19)	Example Tweets
		#GetUsPPE	
Advocacy	provide, work, patients, frontlines, care, congress, shortages, severe, vital, experiencing	1.603	RNs have been experiencing severe shortages in personal protective equipment (PPE) as they work on the frontlines to provide vital care to COVID-19 patients. Tell your member of Congress to do everything possible to #GetMePPE <URL>
	support, mass, delivered, deliver, ll, officials, increase, signing, copy, rapidly	1.068	Support <USER> by signing "We Need to Rapidly Increase PPE Mass Production"
Support	donate, supplies, medical, hospitals, masks, local, share, supply	1.35	Donate medical supplies.#GetMePPE <URL>
Patient/Healthcare Worker Safety	equipment, protective, personal, healthcare, lives	1.315	<USER> <USER> Our Health Care Workers Begging For Personal Protective Equipment!#GetMePPE <URL>
Supply Shortages	#ppe, #ppeshortage, make, masks, hospitals, supplies, n95	1.295	#GetUsPPE #PPEshortage #GetUsPPE #PPEshortage THESE HEROS NEED OUR HELP <URL>
		#COVID19	
Political Affiliation	trump, virus, response, china	-0.101	Trump is a Failure. #impeachTrumpnow #VoteDemocratic
General Discourse	news, good, great, world, due, read	-0.198	Some good news in the time of #CoronaVirus
	stop, spread, testing, virus, social, news	-0.229	Please stop coming into restaurants
	cases, testing, today, day, china	-0.319	Estimated to be over 100k cases and 5 confirmed cases of #coronavirus in #Ohio. Geez
Nation Association	virus, china, world, fight, due, it's, spread	-0.347	It originated in China. It's a China virus. Grow up!

Figure 6. Linguistic correlates of health care–led Twitter hashtag campaigns (#ThisIsOurLane and #GetUsPPE/#GetMePPE) compared with general ones (#GunViolence and #COVID19). Positive beta indicates a strong correlation of the linguistic category with the case compared to the control tweets. *"Power" was not significant at P<.001 for cohort 1.



Discussion

Principal Findings

This study demonstrates not just the reach but also the inclusiveness and uniqueness of tweets containing health care–led hashtags about commonly discussed health care epidemics. Consistent with our hypotheses, tweets containing health care–led hashtags differed qualitatively and quantitatively from other tweets on the same topic during the same time period, albeit not in the way that we predicted. Tweets with #ThisIsOurLane and #GetUsPPE expressed more positivity and a greater sense of group affiliation than comparison hashtags led by the general public. Both #ThisIsOurLane and #GetUsPPE tweets contained more actionable language such as "research," "prevention," and "support."

Social media's potential as a platform for enhancing health discussions is frequently discussed [19,20]. Some authors have even urged the use of social media to develop grassroots "new power" movements that can combat mistruths in science and public health [21]. Others have described the potential utility of specific health care–led tweets for disseminating factual information [22]. Our analysis supports that health care–led hashtags contribute unique, actionable content and tone to national discussions about health, and can create new, inclusive movements that provide opportunities for health care professionals and non–health care–based individuals. Although we did not examine the relative prevalence of facts versus misinformation between the two sets of hashtags, the results of our study offer further evidence of the value of using Twitter to shape and build support for public health movements.

Prior literature demonstrates social media's potential for reaching new groups regarding issues in medicine and public health. However, few previous studies have characterized whether the content of social media campaigns initiated by the health care community are truly unique. For example, TikTok videos about COVID-19 accumulated over 1 billion views; however, an analysis of these videos reports that only a small portion were led by health care professionals, and that few—even those developed by the World Health Organization—included actionable tools to prevent or handle the pandemic [23,24]. Another study reported that a Twitter campaign to raise skin cancer prevention awareness led to nearly 12 million impressions on social media [25] but did not examine content or tone of shared posts. Still, others have demonstrated that health-related content on social media reflects local public health concerns and sentiments but have not examined the relative contribution of health care– versus non–health care–led hashtags [26-28]. Our work is therefore unique in examining not only the number of posts but also what differentiated them from non–health care–led posts on the same topics at the same time.

A particularly noteworthy finding from our study is the positive tone and action-oriented content of tweets with health care–initiated hashtags. This finding differs from our expectations: we hypothesized that health care professionals would be sharing the truth about firearm injury and COVID-19, and that these realities would be negatively valenced. The finding of positive tone, even on difficult issues, may reflect societal expectations of professionalism from medical experts [29,30]. It may also reflect health care professionals' desire to motivate action in others: positive affect and positive tone both increase the acceptability and efficacy of behavioral interventions [31,32]. Indeed, some work has specifically provided guidance to health care and public health professionals on how to avoid or manage "trolls" [33]. Future work should examine whether successful hashtag campaigns are more positive than unsuccessful campaigns.

Establishing hashtags makes health care professionals' conversations more accessible to the nonmedical community and can be used to cultivate momentum around public health campaigns that carry educational and actionable content. Despite #ThisIsOurLane and #GetUsPPE being initiated and more commonly used by health care professionals, people outside of health care also commonly tweeted with these terms. Based on hashtag categories developed by Saxton et al [34], #ThisIsOurLane and #GetUsPPE are public education and call-to-action hashtags, which are most likely to be retweeted, and therefore most effective for online advocacy.

Future work should examine the characteristics of successful hashtag development and dissemination, as how to best create and shepherd these discussions is undetermined. Based on the origin story of #GetUsPPE and #ThisIsOurLane, a successful movement likely does not depend on derivation from a large company or influential organization. Instead, as Twitter increasingly serves as a news source for the general public [35], it offers a platform for average health care professionals to both spread facts and increase action on critical public health issues. Some works in the literature have developed best practices for successfully using health care hashtags to increase audience engagement [34]. Although the United States' Centers for Disease Control and Prevention has guidelines on Twitter use for health communication, initial analyses suggest mixed efficacy of their Twitter campaigns [36]. To inform others' work, future research should examine in more detail which characteristics of #ThisIsOurLane and #GetUsPPE enabled coalescence of a larger community.

Limitations

Limitations to this analysis include the correlational and noncausal nature of the results. This study cannot comment on whether health care–led hashtag campaigns introduced new thoughts on national health issues, as we did not review tweets from health care professionals about gun violence or the COVID-19 pandemic before the hashtags were introduced. Additionally, the magnitude of the influence of tweets with health care–led hashtags is not characterized. Finally, our analysis did not account for the voice of patients and survivors, who have previously been shown to have a powerful role on Twitter.

Conclusion

Historically, health care professionals play defining roles in social justice and public health movements. Health care–led hashtag campaigns are positive, actionable, and portray a united front in developing solutions to pressing public health issues. The #ThisIsOurLane and #GetUsPPE movements exemplify how online media can influence 21st-century social dialogues about disease, injury, and prevention.

Acknowledgments

RSB and MLR are both funded by R24 HD087149 (PI: Cunningham). SCG acknowledges the support from Google Cloud.

Conflicts of Interest

AO and MLR are volunteers with the organization GetUsPPE.org. RSB reports receiving grants from the National Institute of Mental Health, National Cancer Institute, National Institute on Aging, National Heart, Lung, and Blood Institute, National Institute of Nursing Research, National Institute of Allergy and Infectious Diseases, the National Psoriasis Foundation, Veterans Affairs Quality Enhancement Research Initiative, Patient Centered Outcomes Research Institute, and the Centers for Disease Control and Prevention; royalties from Oxford University Press; served as a consultant to Camden Coalition of Healthcare Providers; and receives an honorarium from Optum Behavioral Health Clinical Scientific Advisory Council. The funding organizations listed above are not related to this article and had no bearing on its outcome.

References

1. Hughes A, Wojcik S. 10 facts about Americans and Twitter. Pew Research Center. 2019 Aug 2. URL: https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/ [accessed 2020-05-26]
2. @Zindoctor. If your GUNS didn't kill & maim so many - men, women, children, of all shapes, sizes & colors - it wouldn't be in our lane. As an #EmergencyMedicine physician, I see and treat patients & families directly devastated by the very reason for your existence. #ThisISOurLane. Twitter. 2018 Nov 18. URL: https://twitter.com/Zindoctor/status/1060338793847418885 [accessed 2020-11-11]
3. @choo_ek. FRONTLINE HEALTH CARE WORKERS Share a pic of the PPE you're in that you need to stay safe. Tag your congresspeople and @VP. Use the hashtag #GetMePPE. Twitter. 2020 Mar 17. URL: https://twitter.com/choo_ek/status/1239790569510993920 [accessed 2020-11-11]

4. Rubin R. Physicians Are Steering the Conversation About Gun Violence. JAMA 2019 Jan 15;321(2):133-135. [doi: 10.1001/jama.2018.20385] [Medline: 30548081]
5. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages — The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/nejmp2006141]
6. Get Us PPE Mission, Vision, and Values. GetUsPPE.org. 2020. URL: https://getusppe.org/mission/ [accessed 2020-11-11]
7. Porat T, Nyrup R, Calvo RA, Paudyal P, Ford E. Public Health and Risk Communication During COVID-19-Enhancing Psychological Needs to Promote Sustainable Behavior Change. Front Public Health 2020 Oct 27;8:573397 [FREE Full text] [doi: 10.3389/fpubh.2020.573397] [Medline: 33194973]
8. Rochwerg B, Parke R, Murthy S, Fernando S, Leigh J, Marshall J, et al. Misinformation During the Coronavirus Disease 2019 Outbreak. Critical Care Explorations 2020;2(4):e0098. [doi: 10.1097/cce.0000000000000098]
9. Dlatk/TwitterMySQL. GitHub. URL: https://github.com/dlatk/TwitterMySQL [accessed 2020-11-11]
10. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking Mental Health and Symptom Mentions on Twitter During COVID-19. J Gen Intern Med 2020 Sep 07;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]
11. Yang Q, Tufts C, Ungar L, Guntuku S, Merchant R. To Retweet or Not to Retweet: Understanding What Features of Cardiovascular Tweets Influence Their Retransmission. J Health Commun 2018 Nov 07;23(12):1026-1035 [FREE Full text] [doi: 10.1080/10810730.2018.1540671] [Medline: 30404564]
12. Georgi S, Guntuku SC, Rahman M, Himelein-Wachowiak M, Kwarteng A, Curtis B. Twitter Corpus of the #BlackLivesMatter Movement And Counter Protests: 2013 to 2020. Arxiv. Preprint posted online September 28, 2020. URL: https://arxiv.org/abs/2009.00596
13. Guntuku SC, Schneider R, Pelullo A, Young J, Wong V, Ungar L, et al. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355 [FREE Full text] [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
14. Guntuku SC, Ramsay JR, Merchant RM, Ungar LH. Language of ADHD in Adults on Social Media. J Atten Disord 2019 Oct 08;23(12):1475-1485. [doi: 10.1177/1087054717738083] [Medline: 29115168]
15. dlatk / happierfuntokenizing. GitHub. URL: https://github.com/dlatk/happierfuntokenizing [accessed 2020-11-11]
16. Schwartz, HA, Giorgi S, Sap M, Crutchley P, Ungar L, et al. Dlatk: Differential language analysis toolkit. 2017 Presented at: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; Sept 2017; Copenhagen, Denmark p. 55-60. [doi: 10.18653/v1/d17-2010]
17. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
18. Pennebaker J, Boyd R, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin, TX: University of Texas at Austin; 2015.
19. Choo EK, Ranney ML, Chan TM, Trueger NS, Walsh AE, Tegtmeyer K, et al. Twitter as a tool for communication and knowledge exchange in academic medicine: A guide for skeptics and novices. Med Teach 2015 May;37(5):411-416. [doi: 10.3109/0142159X.2014.993371] [Medline: 25523012]
20. Park A, Conway M. Tracking Health Related Discussions on Reddit for Public Health Applications. AMIA Annu Symp Proc 2017;2017:1362-1371 [FREE Full text] [Medline: 29854205]
21. Perera K, Timms H, Heimans J. New power versus old: to beat antivaccination campaigners we need to learn from them-an essay by Kathryn Perera, Henry Timms, and Jeremy Heimans. BMJ 2019 Nov 21;367:l6447. [doi: 10.1136/bmj.l6447] [Medline: 31753812]
22. Rabarison KM, Croston MA, Englar NK, Bish CL, Flynn SM, Johnson CC. Measuring Audience Engagement for Public Health Twitter Chats: Insights From #LiveFitNOLA. JMIR Public Health Surveill 2017 Jun 08;3(2):e34 [FREE Full text] [doi: 10.2196/publichealth.7181] [Medline: 28596149]
23. Basch C, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10;19. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
24. Comp G, Dyer S, Gottlieb M. Is TikTok The Next Social Media Frontier for Medicine? AEM Education and Training 2020 Oct 21. [doi: 10.1002/aet2.10532]
25. Nguyen JL, Heckman C, Perna F. Analysis of the Twitter "Don't Fry Day" Campaign. JAMA Dermatol 2018 Aug 01;154(8):961-962 [FREE Full text] [doi: 10.1001/jamadermatol.2018.1481] [Medline: 29926080]
26. Paul M, Dredze M. You are what you tweet: Analyzing twitter for public health. 2011 Presented at: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Spain p. 265-272 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264
27. Ji X, Chun SA, Wei Z, Geller J. Twitter sentiment classification for measuring public health concerns. Soc Netw Anal Min 2015 May 12;5(1):13 [FREE Full text] [doi: 10.1007/s13278-015-0253-5] [Medline: 32226558]
28. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
29. Regis T, Steiner MJ, Ford CA, Byerley JS. Professionalism expectations seen through the eyes of resident physicians and patient families. Pediatrics 2011 Feb 17;127(2):317-324. [doi: 10.1542/peds.2010-2472] [Medline: 21242219]

30. Sawicki NN. Judging doctors-the person and the professional. Virtual Mentor 2011 Oct 01;13(10):718-722 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.10.msoc1-1110] [Medline: 23137893]
31. Lawton R, Conner M, McEachan R. Desire or reason: predicting health behaviors from affective and cognitive attitudes. Health Psychol 2009 Jan;28(1):56-65. [doi: 10.1037/a0013424] [Medline: 19210018]
32. Muench F, van Stolk-Cooke K, Morgenstern J, Kuerbis AN, Markle K. Understanding messaging preferences to inform development of mobile goal-directed behavioral interventions. J Med Internet Res 2014 Feb 05;16(2):e14 [FREE Full text] [doi: 10.2196/jmir.2945] [Medline: 24500775]
33. Jamison AM, Broniatowski DA, Quinn SC. Malicious Actors on Twitter: A Guide for Public Health Researchers. Am J Public Health 2019 May;109(5):688-692. [doi: 10.2105/AJPH.2019.304969] [Medline: 30896994]
34. Saxton GD, Niyirora JN, Guo C, Waters RD. #AdvocatingForChange: The Strategic Use of Hashtags in Social Media Advocacy. ASW 2015 Jul 27;16(1):154-169. [doi: 10.18060/17952]
35. News use across social media platforms 2018. Pew Research Center. 2018 Sep 10. URL: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [accessed 2020-11-11]
36. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of Health Agency Response and Public Engagement in Public Health Emergency: A Case Study of CDC Tweeting Patterns During the 2016 Zika Epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]

Abbreviations

API: application programming interface
MALLET: Machine Learning for Language Toolkit
PPE: personal protective equipment

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by T Chan, K Utter; comments to author 23.10.20; revised version received 12.11.20; accepted 12.12.20; published 06.01.21.

Please cite as:
Ojo A, Guntuku SC, Zheng M, Beidas RS, Ranney ML
How Health Care Workers Wield Influence Through Twitter Hashtags: Retrospective Cross-sectional Study of the Gun Violence and COVID-19 Public Health Crises
JMIR Public Health Surveill 2021;7(1):e24562
URL: https://publichealth.jmir.org/2021/1/e24562
doi:10.2196/24562
PMID:33315578

Original Paper

Using a Twitter Chat to Rapidly Identify Barriers and Policy Solutions for Metastatic Breast Cancer Care: Qualitative Study

Riti Shimkhada[1*], PhD; Deanna Attai[2*], MD; AJ Scheitler[1*], EdD; Susan Babey[1*], PhD; Beth Glenn[3*], PhD; Ninez Ponce[1,3*], PhD

[1]Center for Health Policy Research, University of California, Los Angeles, Los Angeles, CA, United States
[2]Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States
[3]Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States
[*]all authors contributed equally

Corresponding Author:
Riti Shimkhada, PhD
Center for Health Policy Research
University of California, Los Angeles
10960 Wilshire Blvd
Los Angeles, CA, 90024
United States
Phone: 1 310 7940909
Email: riti@ucla.edu

Abstract

Background: Real-time, rapid assessment of barriers to care experienced by patients can be used to inform relevant health care legislation. In recent years, online communities have become a source of support for patients as well as a vehicle for discussion and collaboration among patients, clinicians, advocates, and researchers. The Breast Cancer Social Media (#BCSM) community has hosted weekly Twitter chats since 2011. Topics vary each week, and chats draw a diverse group of participants. Partnering with the #BCSM community, we used Twitter to gather data on barriers to care for patients with metastatic breast cancer and potential policy solutions. Metastatic breast cancer survival rates are low and in large part conditioned by time-sensitive access to care factors that might be improved through policy changes.

Objective: This study was part of an assessment of the barriers to care for metastatic breast cancer with the goal of offering policy solutions for the legislative session in California.

Methods: We provided 5 questions for a chat specific to metastatic breast cancer care barriers and potential policy solutions. These were discussed during the course of a #BCSM chat on November 18, 2019. We used Symplur (Symplur LLC) analytics to generate a transcript of tweets and a profile of participants. Responses to the questions are presented in this paper.

Results: There were 288 tweets from 42 users, generating 2.1 million impressions during the 1-hour chat. Participants included 23 patient advocates (most of whom were patients themselves), 7 doctors, 6 researchers or academics, 3 health care providers (2 nurses, 1 clinical psychologist), and 2 advocacy organizations. Participants noted communication gaps between patient and provider especially as related to the need for individualized medication dosing to minimize side effects and maximize quality of life. Timeliness of insurance company response, for example, to authorize treatments, was also a concern. Chat participants noted that palliative care is not well integrated into metastatic breast cancer care and that insurance company denials of coverage for these services were common. Regarding financial challenges, chat participants mentioned unexpected copays, changes in insurance drug formularies that made it difficult to anticipate drug costs, and limits on the number of physical therapy visits covered by insurance. Last, on the topic of disability benefits, participants expressed frustration about how to access disability benefits. When prompted for input regarding what health system and policy changes are necessary, participants suggested a number of ideas, including expanding the availability of nurse navigation for metastatic breast cancer, developing and offering a guide for the range of treatment and support resources patients with metastatic breast cancer, and improving access to clinical trials.

Conclusions: Rapid assessments drawing from online community insights may be a critical source of data that can be used to ensure more responsive policy action to improve patient care.

(JMIR Public Health Surveill 2021;7(1):e23178) doi:10.2196/23178

KEYWORDS

metastatic breast cancer; Twitter; infodemiology; infoveillance; health care barriers; health care policy; social media; policy; breast cancer

Introduction

Web-based social media platforms have changed the face of support networks, breaking down the barriers of distance, time, and physical limitations to bring together diverse voices in a common cause. For social scientists interested in hearing current conversations from social networks, these platforms offer access to a sample of engaged key informants and a way to generate rapid insights. Twitter has been increasingly used to carry out public health research such monitoring diseases and outbreaks, gauging public opinion to emerging health threats, and implementing health education campaigns [1-6]. Prior work suggests that Twitter may aid policy-makers engage their constituents beyond what is possible through in-person meetings or passive communication (eg, listservs, newsletters) [4,5,7-9]. A growing number of patients use Twitter to connect with their peers as well as clinicians and researchers for information, psychosocial support, and research and advocacy opportunities; patients treated for breast cancer are one of the most engaged groups on social media platforms for this purpose [1,10-15].

Breast cancer is the most commonly diagnosed cancer among women in the United States, with approximately 276,000 women diagnosed each year [16]. Metastatic breast cancer, or stage 4 breast cancer, occurs when breast cancer spreads to other sites, most commonly the lungs, liver, bones, and brain. Patients with metastatic breast cancer are usually on some form of treatment for their remaining lifespan, which may include hormone blockade, targeted therapy, chemotherapy, immunotherapy, surgery, or radiation therapy. While there have been improvements in treatment options available to individuals with metastatic breast cancer, there is no cure and every year, in the United States, approximately 42,000 women will die from breast cancer [16]. Because of the intensity of the disease and its treatment, patients with metastatic breast cancer must stay engaged in the health care system and in close contact with their treatment team to receive timely care tailored to their specific needs and preferences.

The Twitter Breast Cancer Social Media (#BCSM) community was cofounded by two breast cancer patient advocates in 2011 and is the first and longest-running cancer support community on Twitter. The #BCSM community hosts weekly tweet chats (live Twitter events) using the #BCSM hashtag to provide information and virtual support to all impacted by breast cancer. Topics vary each week, and a diverse community of patients (eg, men and women, early and late stage disease), doctors and other health care providers, researchers, and representatives from advocacy organizations participate [1]. Recognizing that online patient communities can be a valuable source of timely, real-world data [17,18], which can guide policy change, the study team partnered with the #BCSM community to host a Twitter chat to gather data on barriers to care and potential policy solutions for patients living with metastatic breast cancer. Barriers to care are factors that impede or limit a patient’s access to health care services [19,20]. The study aimed to collect information on the experiences of patients living with metastatic breast cancer, health care providers, advocates, and support service providers who work with patients with metastatic breast cancer. This study was part of a larger initiative to gather stakeholder input in informing policy proposals to improve care for women with metastatic breast cancer for the next California legislative session, thus the timeliness of data gathering was critical. Our objective was to assemble a list of priority areas and policy recommendations to improve metastatic breast cancer care from chat participants on Twitter who have been highly engaged in breast cancer care advocacy, research, and care delivery. These recommendations will be included in a report intended for policy makers who are looking for ways to improve metastatic breast cancer care.

Methods

The University of California, Los Angeles institutional review board determined that the study was exempt from review due to the public nature of conversations held on Twitter. The study team designed 5 questions to solicit input about metastatic breast cancer care barriers and potential policy solutions. These questions were informed by prior research on barriers to care for women with earlier stage breast cancer [21], a literature review conducted for this study, and feedback from key stakeholders on metastatic breast cancer. We worked with #BCSM comoderator (since 2011) and study partner (DJA), to identify the ideal number of questions and wording for the chat questions so that questions could be asked within a 1-hour timeframe.

The questions were posted on November 13, 2019 on a breast cancer information blog maintained by author DJA that was shared on Twitter with the #BCSM tag to prepare and inform likely participants of the scheduled 1-hour Twitter chat on November 18, 2019. The questions were (1) What are some of the most significant health care communication barriers faced by patients with metastatic breast cancer? (2) What are the palliative care barriers faced by those with metastatic breast cancer? (3) What are the financial challenges faced by patients w with metastatic breast cancer? (4) What are barriers to obtaining disability? (5) What health system or policy changes would you suggest to improve the care experience for patients with metastatic breast cancer?

During the introduction portion of the chat, participants were informed by the moderator (DJA) that their tweets would be used for a research study, and those who did not want to participate in the study should not tweet using the #BCSM hashtag for the 1-hour duration of the scheduled chat. This study’s principal investigator (NP) was introduced at the start of the chat and was an active participant in the discussion. During the chat, the moderator posted the 5 questions sequentially and allowed for conversation and questions, per usual tweet chat routine.

Publicly available information from Symplur (Symplur LLC), a health care focused Twitter database and analytics company, was used to generate a transcript of tweets from the 1-hour chat. Only tweets that were in direct responses to each of the posed questions (ie, no retweets or off-topic entries) were evaluated. All direct and unique responses to each question were compiled for presentation in the paper by one author (RS) and checked by another author (AS). Symplur Signals (Symplur LLC) was used to categorize participants into the following groups: doctor, health care provider (not doctor), patient advocate, advocacy organization, health care organization, or research/academic. For health care stakeholders classified in more than one category, a manual review of Twitter profiles was performed to determine the category in which the participant best fit. We include the category of each participant for each Twitter chat response to help readers understand the perspective or experience of each participant.

Results

During the course of the 1-hour chat, there were 288 tweets from 42 unique participants. This generated 2.1 million impressions (a measure of tweet reach, indicating potential views). Participants included 23 patient advocates (most of whom were further identified as patients with breast cancer based on their Twitter profiles), 7 doctors, 7 researchers/academics, 3 health care providers (2 nurses, 1 clinical psychologist), and 2 representatives of advocacy organizations (Table 1). Based on their Twitter biographies, participants resided in the United States (n=40), Canadian (n=1), and unknown (n=1). More granular geolocation data were not available for all of the participants to further identify each participant's city or state of residence.

Representative tweets generated by the participants in response to each of the questions are shown in Table 2, and full results of the Twitter chat are shown in Multimedia Appendix 1. The first question inquiring about communication barriers to care generated numerous responses such as communication gaps between patient and provider, especially regarding communicating disease progression and quality of life. Timeliness of insurance company response, for example, to authorize treatments, was also a concern.

The second question focused on palliative care, which includes the management of the side effects of treatments and treatment of the symptoms of disease to optimize quality of life. Palliative care includes a number of domains of care: physical, social, cultural, emotional, spiritual, structural, psychological, and end of life [22]. In their responses, chat participants noted that palliative care is not well integrated into metastatic breast cancer care and that insurance company denials of coverage for these services were common. Regarding financial challenges, chat participants mentioned unexpected copays, changes in insurance drug formularies that made it difficult to anticipate drug costs year to year, and limits on the number of physical therapy visits covered by insurance. On the topic of disability benefits, participants expressed frustration that there is a lack of clear guidance available on how to access disability benefits. When prompted for input regarding needed health system and policy changes, participants suggested expanding the availability of nurse navigation for metastatic breast cancer, developing a guide for the range of treatment and support resources for patients at the time of metastatic breast cancer diagnosis, and improving access to clinical trials, specifically, reimbursing the cost of travel and accommodations to and from the trial site.

Table 1. Description of #BCSM community Twitter chat participants.

Participant type	Participants (N=42), n (%)	Tweets, n (%)
Patient or advocate	23 (55)	129 (45)
Doctor[a,b]	7 (17)	125 (43)
Researcher/academic	7 (17)	18 (6)
Health care provider	3 (7)	6 (2)
Advocacy organization	2 (5)	10 (4)
Total	42 (100)	288 (100)

[a]Includes #BCSM chat moderator (n=99 tweets).

[b]1 participant classified as a doctor is a practicing doctor and a patient with breast cancer in remission.

Table 2. Questions and representative responses from participants regarding barriers to care for metastatic breast cancer and potential policy solutions.

Question	Participant tweet
What are some of the most significant health care communication barriers faced by patients with metastatic breast cancer?	*MBC pts often try to look as good as possible at their onc appts and may not report QOL issues unless oncologist asks. Some don't ask.* [Patient advocate] *Big communication issue is communication style – pt usually has a preference for how detailed they want their [oncologist]. to be. Need to discuss this up front. Otherwise pt often seeks new onc.* [Patient advocate] *I have family members who have been diagnosed with early stage breast cancer and they were unaware that ~20-30% of early stage breast cancers will become MBC. I think this may be a communication gap to patients* [Patient advocate]
What are the palliative care barriers faced by those with metastatic breast cancer?	*Palliative care is not always discussed with patients and not explained well. For a long time I thought palliative care and hospice were synonymous.* [Patient advocate] *My plan's Palliative Care Team is still figuring out what they do for a living. Right now focus is mainly advanced directives and pain meds [medications]. Should get better but not yet.* [Patient advocate] *A hospital or cancer center having onsite palliative care or ability to have access to Palliative Care is a huge issue for countless patients. Many are referred to pain clinics which are NOT the same at all. There is an alarming shortage of/access to Palliative Care.* [Patient advocate]
What are the financial challenges faced by patients with metastatic breast cancer?	*Surprise co-pays for new therapies once the line of therapy is established. Then scrambling for payment assistance when already stressed by progression.* [Patient advocate] *Was recently told that I could not continue PT [physical therapy] because I had already had 35 and anything beyond would be out of pocket. So now I either pay for PT or just wait until January.* [Patient advocate] *To make matters worse/more frustrating - for treatment meds taken at home... Many private insurers change their formulary lists twice/year…MD [doctor] offices often can't keep up with those changes & pts find out after the fact.* [Health care provider]
What are barriers to obtaining disability?	*Big barrier to getting disability is the pt doesn't know the process of applying if employed, small companies don't know what to do.* [Patient advocate] *There are lots of people who are contract workers (especially in high tech). They don't have disability insurance as aren't aware of state disability.* [Patient advocate] *A lot of times pts have to be off of work consecutively for 12 weeks before their application is looked at. And you have to exhaust all of your benefit time as well.* [Advocacy organization]
What health system or policy changes would you suggest to improve the care experience for patients with metastatic breast cancer?	*Besides patient centered dosing, we need to make it easier to be in clinical trials without requiring expensive travel costs to pts. We limit who can be in a trial by pts who are unable to pay for travel.* [Patient advocate] *Would love for [insurance] companies to have a separate group just for metastatic (maybe unrealistic) but would have training and understand the unique needs of that group.* [Patient advocate] *Maybe it would be a good idea to have a national nurse navigator organization that would work like a hotline. So even remote access.* [Patient advocate]

Discussion

In partnership with an online breast cancer community, we identified several areas where legislation, policy change, or greater investment of resources can be made to improve metastatic breast cancer care. Many of the barriers to care relate to communication, care coordination, and insurance authorization. This study was part of an assessment of the barriers to care for metastatic breast cancer with the goal of offering policy solutions for the legislative session in California.

Chat participants engaged in conversations around communication barriers, echoing the need to lift communication and care coordination barriers from the patient-provider relationship [23]. Chat participants noted that patients with metastatic breast cancer, compared to patients with nonmetastatic disease have treatment decision preferences that focus far more on quality of life [24,25], and may benefit from protocols that require routine reporting of quality of life to providers. Furthermore, improved access to nurse navigation for patients with metastatic cancer could reduce the burden of care coordination on the patient and their caregiver.

Chat participants noted a number of difficulties with insurance approval of treatments. Step therapy or a fail-first protocol is an insurer's policy that requires a patient to try therapies in a specific order (ie, try a less expensive generic or biosimilar version of a therapy before moving up a step to the more expensive therapy). These processes could impose barriers to access and delays in receiving the most effective treatment [26]. For patients with metastatic breast cancer who face a 5-year survival rate of only 27% [16], fail-first protocols are especially penalizing, protracting access to a drug that may be preferred by the patient and covered by the patient's insurer, thus dually harming the health and financial well-being of women with metastatic breast cancer. There are currently 8 states (Arizona, Colorado, Illinois, Louisiana, Maryland, Minnesota, North Dakota, and Texas) that have laws restricting the use of step therapy or prior authorization protocols for patients with stage 4 or metastatic cancer; other states are considering similar legislation [27].

Even when coverage for a particular treatment is approved by insurance, cost of the medication or service remains a concern as the patient is responsible for copayment for the treatment set forth by the plan design. Financial concerns and barriers were expressed by patients in our chat. Given that many metastatic breast cancer medications are specialty or in the tie of high-cost drugs, patients are responsible for a greater share of the cost compared to that for nonspecialty drugs. This is particularly a burden for patients who are covered by high-deductible health plans [28].

Disease progression is one of several patient factors associated with financial distress [29], and patients with metastatic breast cancer may wish to address costs at the time of treatment decisions [30]. The American Society of Clinical Oncology encourages cost discussions between patients and providers [31]. The increased focus on value-based care as well as attention to financial toxicity experienced by patients undergoing treatment for cancer has made it even more important that clinicians take on a role of financial advocate for their patients, although this is an area where physicians do not always feel comfortable [32-35]. There are calls for more incentives for cost discussions and subsequent reduction in financial burden among patients using the personal spending burden (measured as personal expenditures for health care relative to income) as a quality-of-care metric [36] and calls to enact reform that makes cost transparent to patients in the prior authorization process [37]. Other efforts to lower costs to patients include the enactment of oral chemotherapy parity laws to limit patient out-of-pocket costs in line with intravenous administered drugs. For example, in 2018 California Assembly Bill 1860 [38] capped patients' out-of-pocket costs for oral chemotherapy to US $250/month per drug. A federal bill, the Cancer Drug Parity Act [39], seeks to bring parity in oral and intravenous chemotherapy to all states in the US.

Tweet chat participants noted that palliative care was often not discussed or there were restrictions regarding participation. California improved access to palliative care with the introduction of Senate Bill 1004 [40] in 2014, which required the California Department of Health Care Services to expand community-based palliative care services to its Medicaid beneficiaries. However, while the bill was a major step forward for patients with advanced diseases such as metastatic breast cancer in terms of access to palliative care, patients may be underusing these services due to lack of referral or appropriate care coordination by their oncologist and misconceptions regarding palliative care versus hospice or end-of-life care. In addition, there may be potential language and cultural barriers [41,42].

Barriers to accessing disability benefits also emerged during the chat. The federal 2019 Metastatic Breast Cancer Access to Care Act [43] would waive the current waiting periods for federal disability benefits of 5 months for Social Security disability benefits and 2 years for Medicare for those younger than 65 years. While the bill, if enacted, would lift the wait to access benefits, tweet chat participants mentioned a need for improved awareness in applying for disability programs and knowing what is available to them. Patient navigation and support communities may offer opportunities for improving literacy in accessing benefits and financial support [44].

Metastatic breast cancer advocates have continued to press for meaningful changes in policy to improve care through virtual lobbying and awareness efforts. Decisions regarding the Metastatic Breast Cancer Access to Care Act [43] and the Cancer Drug Parity Act [39] had still not been made (as of the time this paper went to press). While these two federal bills are important steps forward, state mandates and local health system policy shifts may be able to bring consequential changes to access and costs through changes in prior authorization practices and support programs.

Our analysis has several limitations. The participants in the chat were not randomly selected; those who took part in the discussions were likely those most familiar with Twitter and the weekly #BCSM chats and adept at the chat format. We do not know the geographic location, age, or disease severity of the participants. Participants in online breast cancer communities may not be representative of the average patient population [45]. We only used one particular hashtag, #BCSM, to link our tweets during the chat; there are other hashtags related to metastatic breast cancer that we did not include but might have improved engagement. Stakeholder characterization was based on self-reported information (in a Twitter user's biography) and a proprietary (Symplur LLC) algorithm. Further work is needed to determine how to best utilize the information discussed by patients on various social media platforms to thoughtfully inform public policy decisions. Despite these limitations, our findings suggest that Twitter can be an important source of timely data on the struggles and barriers being faced by patients with cancer and other health conditions.

This Twitter chat elicited a number of policy or program ideas that may improve barriers to care for patients with metastatic breast cancer. Multiple participants reported the lack of patient navigation for metastatic breast cancer and felt a policy priority could be to initiate a navigation program, hotline, or guide for all services available to a patient undergoing treatment. One participant felt metastatic cancers are so different from nonmetastatic cancers that there might be a need for physician groups who treat just metastatic cancers. Another participant mentioned clinical trial access might be improved if travel costs to and from trial sites were covered. Participants noted the need for policies related to improving palliative care and better quality-of-life reporting. Participants also noted financial and insurance barriers that might be addressable through health mandates, such as restricting formulary switching so that medications remain covered by insurance once a patient starts the treatment, limiting restrictions on number of physical therapy visits, and limiting surprise copayments. All recommendations brought up by #BCSM Twitter chat participants are included in a report that will be disseminated to policy makers working to improve timely access to care for women living with metastatic breast cancer in California.

Rapid assessments drawing from various online patient communities, not just those focused on cancer, may provide critical, timely information that can be used to ensure more responsive policy action. This has become even more important as the health care system adjusts in response to the COVID-19 pandemic.

Acknowledgments

This work was done through the support of the California Breast Cancer Research Program Policy Initiative grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1. Full list of Twitter chat responses from participants regarding barriers to care for metastatic breast cancer and potential policy solutions.

[DOCX File , 18 KB - publichealth_v7i1e23178_app1.docx]

References

1. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
2. Charalambous A. Social media and health policy. Asia Pac J Oncol Nurs 2019;6(1):24-27 [FREE Full text] [doi: 10.4103/apjon.apjon_60_18] [Medline: 30599012]
3. Kapp JM, Hensel B, Schnoring KT. Is Twitter a forum for disseminating research to health policy makers? Ann Epidemiol 2015 Dec;25(12):883-887. [doi: 10.1016/j.annepidem.2015.09.002] [Medline: 26460202]
4. Keller B, Labrique A, Jain KM, Pekosz A, Levine O. Mind the gap: social media engagement by public health researchers. J Med Internet Res 2014 Jan 14;16(1):e8 [FREE Full text] [doi: 10.2196/jmir.2982] [Medline: 24425670]
5. Park H, Reber BH, Chon M. Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement. J Health Commun 2016 Dec 30;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
6. Subbiah IM, Hamilton E, Knoll M, Shanahan K, Meisel J. A big world made small: using social media to optimize patient care. American Society of Clinical Oncology Educational Book 2019 May(39):e212-e218. [doi: 10.1200/edbk_246643]
7. Brownson RC, Dodson EA, Kerner JF, Moreland-Russell S. Framing research for state policymakers who place a priority on cancer. Cancer Causes Control 2016 Aug 14;27(8):1035-1041 [FREE Full text] [doi: 10.1007/s10552-016-0771-0] [Medline: 27299656]
8. Dodson EA, Geary NA, Brownson RC. State legislators' sources and use of information: bridging the gap between research and policy. Health Educ Res 2015 Dec 13;30(6):840-848 [FREE Full text] [doi: 10.1093/her/cyv044] [Medline: 26464418]
9. Morshed AB, Dodson EA, Tabak RG, Brownson RC. Comparison of research framing preferences and information use of state legislators and advocates involved in cancer control, United States, 2012-2013. Prev Chronic Dis 2017 Feb 02;14:E10 [FREE Full text] [doi: 10.5888/pcd14.160292] [Medline: 28152363]
10. Attai DJ, Hampton R, Staley AC, Borgert A, Landercasper J. What do patients prefer? understanding patient perspectives on receiving a new breast cancer diagnosis. Ann Surg Oncol 2016 Oct 15;23(10):3182-3189. [doi: 10.1245/s10434-016-5312-2] [Medline: 27306904]
11. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb 04;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
12. Cutshall NR, Kwan BM, Salmi L, Lum HD. "It makes people uneasy, but it's necessary. #BTSM": using Twitter to explore advance care planning among brain tumor stakeholders. J Palliat Med 2020 Jan 01;23(1):121-124 [FREE Full text] [doi: 10.1089/jpm.2019.0077] [Medline: 31170019]
13. Falisi AL, Wiseman KP, Gaysynsky A, Scheideler JK, Ramin DA, Chou WS. Social media for breast cancer survivors: a literature review. J Cancer Surviv 2017 Dec 10;11(6):808-821. [doi: 10.1007/s11764-017-0620-5] [Medline: 28601981]
14. Platt JR, Brady RR. #BCSM and #breastcancer: contemporary cancer-specific online social media communities. Breast J 2020 Apr 06;26(4):729-733. [doi: 10.1111/tbj.13576] [Medline: 31493301]
15. Sedrak MS, Attai DJ, George K, Katz MS, Markham MJ. Integrating social media in modern oncology practice and research. American Society of Clinical Oncology Educational Book 2018 May(38):894-902. [doi: 10.1200/edbk_204453]
16. American Cancer Society. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf [accessed 2021-01-04]
17. Hodgkin P, Horsley L, Metz B. The emerging world of online health communities. Stanford Social Innovation Review 2018;Available at [FREE Full text]
18. Vicari S, Cappai F. Health activism and the logic of connective action. a case study of rare disease patient organisations. Information, Communication & Society 2016 Mar 21;19(11):1653-1671. [doi: 10.1080/1369118x.2016.1154587]
19. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1. [doi: 10.2307/2137284]

20. Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski EF, editors. Changing the U.S. Health Care System: Key Issues in Health Services, Policy, and Management. San Francisco, CA: Jossey-Bass; 2001:3-30.
21. Ponce NA, Glenn BA, Shimkhada R, Scheitler AJ, Ko M. An examination of the barriers to breast cancer care in California. Am J Med Res 2017;4(2):73. [doi: 10.22381/ajmr4220174]
22. Kaufmann TL, Kamal AH. Oncology and palliative care integration: cocreating quality and value in the era of health care reform. JOP 2017 Sep;13(9):580-588. [doi: 10.1200/jop.2017.023762]
23. Lee SJC, Clark MA, Cox JV, Needles BM, Seigel C, Balasubramanian BA. Achieving coordinated care for patients with complex cases of cancer: a multiteam system approach. JOP 2016 Nov;12(11):1029-1038. [doi: 10.1200/jop.2016.013664]
24. Di Lascio S, Pagani O. Is it time to address survivorship in advanced breast cancer? a review article. The Breast 2017 Feb;31:167-172. [doi: 10.1016/j.breast.2016.10.022]
25. Spaich S, Kinder J, Hetjens S, Fuxius S, Gerhardt A, Sütterlin M. Patient preferences regarding chemotherapy in metastatic breast cancer-a conjoint analysis for common taxanes. Front Oncol 2018 Nov 21;8:535 [FREE Full text] [doi: 10.3389/fonc.2018.00535] [Medline: 30519542]
26. Gaines ME, Auleta AD, Berwick DM. Changing the game of prior authorization: the patient perspective. JAMA 2020 Feb 25;323(8):705-706. [doi: 10.1001/jama.2020.0070] [Medline: 32011646]
27. Analysis of California assembly bill 2144 step therapy. California Health Benefits Review Program. 2020. URL: http://analyses.chbrp.com/document/view.php?id=1494 [accessed 2021-01-04]
28. Leopold C, Wagner AK, Zhang F, Lu CY, Earle CC, Nekhlyudov L, et al. Total and out-of-pocket expenditures among women with metastatic breast cancer in low-deductible versus high-deductible health plans. Breast Cancer Res Treat 2018 Sep 1;171(2):449-459 [FREE Full text] [doi: 10.1007/s10549-018-4819-6] [Medline: 29855813]
29. Yabroff KR, Bradley C, Shih YT. Understanding financial hardship among cancer survivors in the United States: strategies for prevention and mitigation. JCO 2020 Feb 01;38(4):292-301. [doi: 10.1200/jco.19.01564]
30. Lei YY, Quain KM, Dizon DS, Jimenez R, Shin JA, Sepucha K, et al. Desire to address costs at time of treatment decisions among patients with metastatic breast cancer. JCO 2019 May 20;37(15_suppl):6640-6640. [doi: 10.1200/jco.2019.37.15_suppl.6640]
31. Gilligan T, Coyle N, Frankel RM, Berry DL, Bohlke K, Epstein RM, et al. Patient-clinician communication: American Society of Clinical Oncology consensus guideline. JCO 2017 Nov 01;35(31):3618-3632. [doi: 10.1200/jco.2017.75.2311]
32. Williams C, Azuero A, Kenzik K, Pisu M, Nipp R, Bhatia S, et al. Guideline discordance and patient cost responsibility in medicare beneficiaries with metastatic breast cancer. J Natl Compr Canc Netw 2019 Oct 01;17(10):1221-1228. [doi: 10.6004/jnccn.2019.7316] [Medline: 31590153]
33. Warsame R, Kennedy CC, Kumbamu A, Branda M, Fernandez C, Kimball B, et al. Conversations about financial issues in routine oncology practices: a multicenter study. JOP 2019 Aug;15(8):e690-e703. [doi: 10.1200/jop.18.00618]
34. Greenup RA, Rushing CN, Fish LJ, Lane WO, Peppercorn JM, Bellavance E, et al. Perspectives on the costs of cancer care: a survey of the American Society of Breast Surgeons. Ann Surg Oncol 2019 Oct;26(10):3141-3151 [FREE Full text] [doi: 10.1245/s10434-019-07594-3] [Medline: 31342390]
35. Rocque G, Blayney DW, Jahanzeb M, Knape A, Markham MJ, Pham T, et al. Choosing wisely in oncology: are we ready for value-based care? JOP 2017 Nov;13(11):e935-e943. [doi: 10.1200/jop.2016.019281]
36. Blumenthal D, McGinnis JM. Measuring vital signs: an IOM report on core metrics for health and health care progress. JAMA 2015 May 19;313(19):1901-1902. [doi: 10.1001/jama.2015.4862] [Medline: 25919301]
37. Insurer inaction on prior authorization reform requires federal response. American Medical Association. 2020 Jun 23. URL: https://www.ama-assn.org/press-center/press-releases/insurer-inaction-prior-authorization-reform-requires-federal-response [accessed 2020-11-14]
38. AB 1860 health care coverage: cancer treatment. California State Legislature. URL: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB1860 [accessed 2021-01-04]
39. HR 1730 - cancer drug parity act. 116th US Congress. URL: https://www.congress.gov/bill/116th-congress/house-bill/1730/text [accessed 2021-01-04]
40. SB 1004 health care: palliative care. California State Legislature. URL: https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB1004 [accessed 2021-01-04]
41. Enguidanos S, Rahman A, Hoe D, Meyers K. Provider-identified barriers to palliative care for Medicaid patients. Innovation in Aging 2019;3:S689. [doi: 10.1093/geroni/igz038.2540]
42. Hoe D, Wang YH, Meyers K, Enguidanos S. Palliative care…what's that?: Medicaid patient-identified barriers to palliative care. Innovation in Aging 2019;3:S917. [doi: 10.1093/geroni/igz038.3342]
43. HR 2178 - metastatic breast cancer access to care act. 116th US Congress. URL: https://www.congress.gov/bill/116th-congress/house-bill/2178/text [accessed 2021-01-04]
44. Baik SH, Gallo LC, Wells KJ. Patient navigation in breast cancer treatment and survivorship: a systematic review. JCO 2016 Oct 20;34(30):3686-3696. [doi: 10.1200/jco.2016.67.5454]
45. Kashian N, Jacobson S. Factors of engagement and patient-reported outcomes in a stage iv breast cancer Facebook group. Health Commun 2020 Jan;35(1):75-82. [doi: 10.1080/10410236.2018.1536962] [Medline: 30351185]

Abbreviations

#BCSM: Breast Cancer Social Media community
COVID-19: coronavirus disease 2019

Edited by T Sanchez; submitted 04.08.20; peer-reviewed by C Ure, J Feliciano, D Juzwishin; comments to author 11.09.20; revised version received 08.10.20; accepted 27.11.20; published 15.01.21.

Please cite as:
Shimkhada R, Attai D, Scheitler AJ, Babey S, Glenn B, Ponce N
Using a Twitter Chat to Rapidly Identify Barriers and Policy Solutions for Metastatic Breast Cancer Care: Qualitative Study
JMIR Public Health Surveill 2021;7(1):e23178
URL: http://publichealth.jmir.org/2021/1/e23178/
doi:10.2196/23178
PMID:33315017

Original Paper

Use of Geosocial Networking Apps and HIV Risk Behavior Among Men Who Have Sex With Men: Case-Crossover Study

Justin Knox[1], PhD; Yi-No Chen[2], MPH; Qinying He[3], BMed; Guowu Liu[4], MSc; Jeb Jones[2], PhD; Xiaodong Wang[5], BSc; Patrick Sullivan[2], DVM, PhD; Aaron Siegler[2], PhD

[1]Columbia University, New York, NY, United States
[2]Emory University, Atlanta, GA, United States
[3]Center for Disease Control and Prevention, Chengdu, China
[4]Center for Disease Control and Prevention, Beijing, China
[5]Chengdu Tongle Social Work Service Center, Chengdu, China

Corresponding Author:
Aaron Siegler, PhD
Emory University
1518 Clifton Road NE, Suite 406
Atlanta, GA, 30322
United States
Phone: 1 404 712 9733
Email: asiegle@emory.edu

Abstract

Background: HIV disproportionately affects men who have sex with men (MSM) in China. The HIV epidemic is largely driven by unprotected anal sex (ie, sex not protected by condoms or HIV pre-exposure prophylaxis [PrEP]). The possible association between unprotected anal sex and the use of geospatial networking apps has been the subject of scientific debate.

Objective: This study assessed whether users of a gay geospatial networking app in China were more likely to use condoms when they met their partners online versus offline. A case-crossover analysis, with each person serving as his own control, was employed to address the potential bias that men looking for sex partners through an online dating medium might have inherently different (and riskier) patterns of sexual behavior than men who do not use online dating media.

Methods: A cross-sectional survey was administered in 2018 to adult male users of Blued—a gay geospatial networking app—in Beijing, Tianjin, Sichuan, and Yunnan, China. A case-crossover analysis was conducted among 1311 MSM not taking PrEP who reported engaging in both unprotected and protected anal sex in the previous 6 months. Multivariable conditional logistic regression was used to quantify the association between where the partnership was initiated (offline or online) and the act of unprotected anal sex, controlling for other interval-level covariates. Four sensitivity analyses were conducted to assess other potential sources of bias.

Results: We identified 1311 matched instances where a person reported having both an unprotected anal sex act and a protected anal sex act in the previous 6 months. Of the most recent unprotected anal sex acts, 22.3% (292/1311), were initiated offline. Of the most recent protected anal sex acts, 16.3% (214/1311), were initiated offline. In multivariable analyses, initiating a partnership offline was positively associated with unprotected anal sex (odds ratio 2.66, 95% CI 1.84 to 3.85, $P<.001$) compared with initiating a partnership online. These results were robust to each of the different sensitivity analyses we conducted.

Conclusions: Among Blued users in 4 Chinese cities, men were less likely to have unprotected anal sex in partnerships that they initiated online compared with those that they initiated offline. The relationship was strong, with over 2.5 times the likelihood of engaging in unprotected anal sex in partnerships initiated offline compared with those initiated online. These findings suggest that geospatial networking apps are a proxy for, and not a cause of, high-risk behaviors for HIV infection; these platforms should be viewed as a useful venue to identify individuals at risk for HIV transmission to allow for targeted service provision.

(JMIR Public Health Surveill 2021;7(1):e17173) doi:10.2196/17173

KEYWORDS

HIV; case-crossover study; dating apps; geosocial networking apps; men who have sex with men; sexual risk behavior

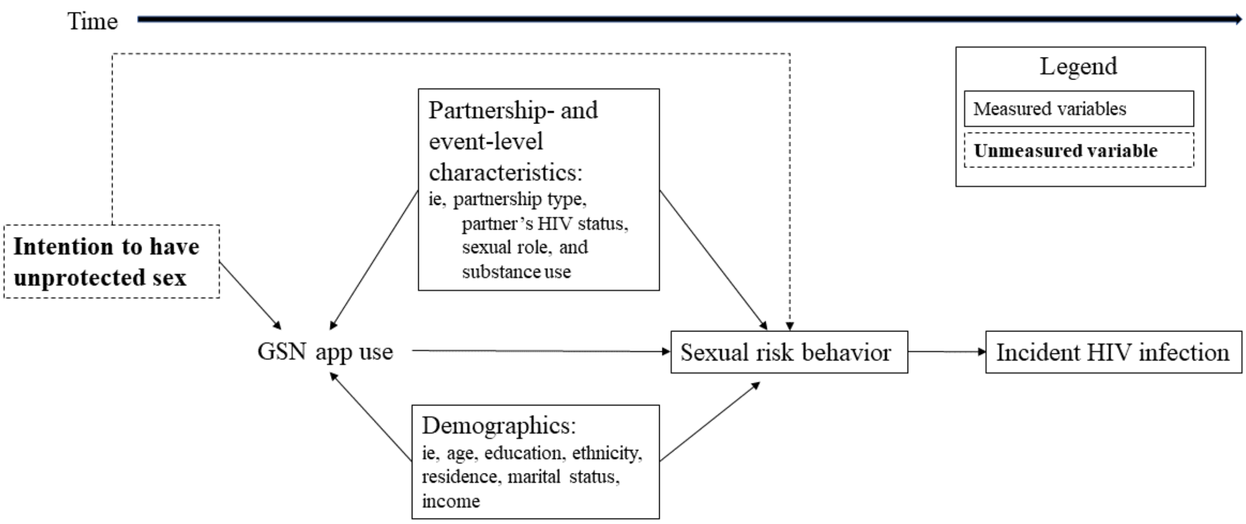
Introduction

HIV disproportionately affects men who have sex with men (MSM) in China [1]. In 2014, a national meta-analysis reported a pooled 7% prevalence of HIV among MSM in China [2]. In Beijing, a cohort study [3] of MSM found an annualized HIV incidence of 5.9 per 100 person-years—an alarming level, similar to intense HIV epidemics among MSM in Thailand [4], South Africa [5], and the southern United States [6]. According to data from China's national sentinel surveillance system, the HIV prevalence among MSM increased from 6% to 8% from 2010 to 2014 [7]. MSM in China represented 12% of new case diagnoses in the country in 2010 and 26% of new case diagnoses in 2014 [8]. MSM is the only risk group with increasing HIV diagnoses, and the estimated HIV incidence was higher for MSM in China than for any other key population [7]. The heightened risk of HIV infection among MSM is largely driven by unprotected anal sex (ie, intercourse not protected by condoms or HIV pre-exposure prophylaxis [PrEP]). Unprotected anal sex is one of the most efficient modes of HIV transmission, yielding a 17-times higher per-act transmission probability than vaginal sex [9,10].

An association between meeting partners through geospatial networking apps (eg, dating apps or "hook-up" apps) and having unprotected anal sex has been hypothesized [11-19]. As smartphone ownership grows increasingly ubiquitous worldwide [20], including in China where there are currently over 800 million smartphone users (more than one-half of the population) [21], there is concern that, if true and causal, such an association might contribute to increased levels of HIV transmission. Studies that have sought to explore this have produced mixed results: some studies conducted among MSM linked the use of geospatial networking apps to high-risk sexual behavior and adverse sexual health outcomes [11-17], but others found no association [18,19]. Mixed results have also been reported in China specifically, where one study reported a higher prevalence of HIV in MSM who met sex partners through the internet [17], while another found no difference in the number of condomless anal sex partners who were male between MSM who used partner-seeking mobile apps and those who did not [19].

More recently, a study in China investigated the impact of the use of a geospatial networking app on incident HIV infection using longitudinal data from a cohort of MSM, finding that incident infection in the follow-up period was associated with ever using a dating app (but, interestingly, not with recent app use) [22]. A shortcoming of the study, and the others mentioned previously that examined the association between app use and HIV infection, is that a substantial source of confounding—the intention to have unprotected sex—was not adequately accounted for. As we elaborate using a directed acyclic graph in Figure 1, if one hypothesizes that exposure to a geospatial networking app causes an increase in HIV risk behaviors, which in turn causes increased HIV incidence, then intention to have unprotected sex (in dashed boxes with bold text) is a confounder because it is associated with both the exposure of interest (one reason that a person might download and use a geospatial networking app is their intention to have unprotected sex) and the outcome of interest (intention to have unprotected sex increases HIV risk, which in turn causes increased HIV incidence). Therefore, not accounting for this variable adequately would result in the observation of a spurious association, even in the event that the underlying variables of interest (ie, use of geospatial networking apps and HIV incidence) were not associated. Without taking intention to have unprotected sex into account, there is no way to determine whether using a dating app increases risk or (alternative hypothesis) people are using the dating app because they intended to have unprotected sex before they even downloaded it. To evaluate a potential causal association, a study design was needed to address the inherent selection bias that imbalances a comparison between men who choose to use dating apps and those who do not. The case-crossover design is one mechanism to address this potential bias in that each person serves as his own control, and the comparison is between the types of sexual activity during periods of exposure (with a partner met online, through a dating app or website) and the types of sexual activity in nonexposure periods (with a partner met offline).

Figure 1. Directed acyclic graph showing the hypothesized pathway from geosocial networking (GSN) app use to sexual risk behavior and incident HIV infection, highlighting the unmeasured confounding effect of intention to have unprotected sex.



Therefore, we used a case-crossover approach in order to assess whether users of a geospatial networking app in China were more likely to have sex not protected by condoms when they initiated partnerships online versus when they initiated partnerships offline. We used data from a cross-sectional study [23] conducted among users of Blued, the largest gay geospatial networking app in the world, with over 40 million registered users in China [24].

Methods

Recruitment

A cross-sectional survey was administered from May 6 to 17, 2018, to adult (≥18 years of age) Blued app users who were male at birth and located in Beijing, Tianjin, Sichuan, or Yunnan, China. Recruitment was conducted through the built-in advertising functions within the Blued app, which included pop-up messages, clickable links, banner advertising, and text message notifications. Text message notifications were sent to 631,963 randomly selected Blued users. Advertisements were designed by the study team and implemented by collaborators at Blued. A total of 34,701 Blued users clicked the link to the survey webpage. Advertisements asked potential participants if they would like to participate in a short research survey. Participants who clicked on the advertisements were taken to a survey platform designed by Sojump within the Blued app, which provided information about the study and a way to navigate to the eligibility screener. All eligible participants were consented electronically before completing the online survey, which was administered using the Sojump survey platform. Although 6040 Blued users initiated the survey process, 1368 of them did not complete the informed consent or did not meet the screening criteria, leaving 4672 participants eligible for these analyses. For analysis, data were exported from Sojump and stored on an encrypted computer at Blued in Beijing.

Ethical approval was provided by China's National Center for AIDS/STD Control and Prevention (NCAIDS) (KX180117492), which is registered with the US Office for Human Research Protections (IRB0000227) and has a Federal Wide Assurance (FWA00002958).

Measures

In the survey, participants were asked to describe their last sex acts involving anal sex with condom use and without condom use, respectively. This allowed us to create a "case interval" (ie, last act of unprotected anal sex) and a "control interval" (ie, last act of protected anal sex) for each individual participant. The survey also collected other characteristics of the sex act (ie, interval-level characteristics), including where participants initiated the partnership (online [eg, through Blued or another app or website] or offline [eg, through friends, at a bar, in a park]), the partnership type (ie, main or committed partner, casual partner with multiple sex acts, or casual partner with a single sex act [one-time]), the participant's knowledge of the partner's HIV status (ie, negative, positive, or unknown status), the participant's role in anal sex (ie, receptive, insertive, or both), and the participant's engagement in substance use (eg, alcohol or illicit drugs) before sexual intercourse.

The survey also collected data regarding the participant's demographic characteristics, sexual identity, substance use in the past 6 months (rush poppers, methamphetamine, methylenedioxymethamphetamine, gamma-hydroxybutyrate, ketamine, and ecstasy), sexually transmitted infections (STIs) in the past 6 months, and HIV status.

Statistical Analysis

Primary Analyses

The objective of the study was to test the association between the use of an online dating medium for finding sex partners and engagement in unprotected anal sex (without the use of condoms or PrEP). We used a case-crossover analysis to isolate the effect of initiating partnerships online (versus offline) on engagement in unprotected anal sex by comparing instances where men initiated partnerships through an online dating medium with instances where they initiated partnerships through an offline medium. Survey respondents who reported always or never using condoms during anal sex in the past 6 months were excluded because they were not informative with respect to the effect of meeting partners online or offline. In addition, respondents who were currently taking PrEP were excluded from the case-crossover analysis.

We used chi-square tests (or Fisher exact test if categorical data were sparse) and two-sample equal variance *t* tests (or Mann-Whitney *U* test if continuous data were not normally distributed) to compare the distribution of individual-level characteristics by participants' eligibility status for the case-crossover analysis. We also compared the distribution of interval-level characteristics between participants who initiated partnerships online and those who initiated partnerships offline in both case and control intervals. We developed bivariable and multivariable models fitting conditional logistic regression to quantify the association between where the partnership was initiated (offline or online) and unprotected anal sex. In the multivariable model, partnership type, partner's HIV status, participant's role in anal sex, and participant's substance use before sex were included as covariates. For all hypothesis tests conducted in this analysis, 95% CIs and a two-tailed *P* value of <.05 were used to assess statistical significance. All data were analyzed using SAS software (version 9.4; SAS Institute Inc).

Sensitivity Analyses

Sensitivity analyses were used to assess possible sources of bias in the analysis. First, the use of online dating media for finding sexual partners could in turn be a determinant of partnership type (ie, sex with one-time, casual, or main partner), which could be associated with unprotected anal sex. Thus, partnership type might be along the causal pathway between initiating a partnership offline and unprotected anal sex, making it a mediator of this relationship and therefore inappropriate to control for it as a potential confounder. To assess this possibility, we conducted multivariable analyses that did not include partnership type as a covariate (as it was in the main analyses). Second, to assess whether the association between initiating a partnership online and having unprotected anal sex varied by partnership type, we stratified the sample by partnership type and conducted analyses within strata. Third, to account for potential differences in sexual behavior among those who met partners exclusively offline or exclusively online, we restricted the analyses to participants who initiated partnerships both online and offline. Fourth, to account for potential differences in sexual behavior among HIV-positive men, we restricted the analyses to participants whose self-reported HIV status was either negative or unknown.

Results

Sample Characteristics

A total of 4672 participants completed the survey. Of those, 3361 participants (71.9%) were excluded from this analysis because they reported not having anal sex in the last 6 months (n=1215), never having condomless anal sex (n=267), always having condomless anal sex (n=1873), or currently using PrEP (n=6). This left a remaining sample of 1311 participants (28.1%) who reported data required for the case-crossover analysis: engaging in both protected and unprotected anal sex. Table 1 presents data on demographic characteristics, sexual orientation, drug use, HIV status, and STI status by inclusion in the analysis. The median age of the participants was 27 years (IQR 23-33 years). More than one-half (2751/4672, 58.9%) of the participants had completed college or more, 20.4% (952/4672) were students, and 71.9% (3360/4672) were employed. Most (3330/4672, 71.3%) of the participants identified as being homosexual. Compared with participants excluded from the analysis, participants included in the analysis were less likely to have attended college, were less likely to be students, were more likely to identify as homosexual, were more likely to report engaging in substance use in the past 6 months, were more likely to have had syphilis in the past 6 months, were more likely to have had gonorrhea in the past 6 months, and were more likely to be HIV positive.

Table 1. Characteristics of 4672 surveyed adult Blued app users in 4 provinces in China.

Characteristic	Total (N=4672), n (%)	Included in analysis (n=1311), n (%)	Excluded from analysis[a] (n=3361), n (%)	*P* value
Age (years), median (IQR)[b]	27 (10)	27 (10)	27 (10)	.39
Ethnicity: Han Chinese	4332 (92.7)	1226 (93.5)	3106 (92.4)	.19
Education: college or above	2751 (58.9)	666 (50.8)	2085 (62.0)	<.001
Current employment				
In the workforce	3360 (71.9)	935 (71.3)	2425 (72.2)	.57
Student	952 (20.3)	234 (17.8)	718 (21.4)	.007
Sexual orientation				
Homosexual	3330 (71.2)	993 (75.7)	2337 (69.5)	<.001
Bisexual	1277 (27.3)	307 (23.4)	970 (28.9)	<.001
Substance use in past 6 months[c]	1611 (34.5)	661 (50.4)	950 (28.3)	<.001
Self-reported STIs[d] in past 6 months				
Not tested	3192 (68.3)	819 (62.5)	2375 (70.7)	N/A[e]
Any STI	232 (5.0)	110 (8.4)	122 (3.6)	<.001
Syphilis	126 (2.7)	65 (5.0)	61 (1.8)	<.001
Gonorrhea	16 (0.3)	11 (0.8)	5 (0.1)	.003
HPV[f]/genital warts	99 (2.1)	36 (2.7)	63 (1.9)	.50
Self-reported HIV status				
Never tested	1416 (30.3)	295 (22.5)	1121 (33.4)	N/A
HIV positive	406 (8.7)	160 (12.2)	246 (7.3)	<.001

[a]Participants who reported to never or always have had condom-protected anal sex 6 months prior to the survey collection were excluded from the study analysis because of the case-crossover study design.

[b]Age distribution was skewed. Wilcoxon Mann-Whitney *U* test was used to assess the distribution difference by eligibility. Median and IQR were used to characterize the distribution.

[c]Drugs included rush poppers, methamphetamine, methylenedioxymethamphetamine, gamma-hydroxybutyrate, ketamine, and ecstasy.

[d]STIs: sexually transmitted infections.

[e]N/A: Not applicable.

[f]HPV: human papillomavirus.

Interval-Level Partnership Characteristics by Offline or Online Initiation, Stratified By Unprotected Anal Sex

Table 2 compares the distribution of interval-level characteristics in partnerships initiated offline and those initiated online within unprotected anal sex acts and within protected anal sex acts. Of the 1311 most recent unprotected anal sex acts, 1019 (77.7%) were initiated online. Of the 1311 most recent protected anal sex acts, 1097 (83.7%) were initiated online.

We found similar bivariate associations across case (unprotected sex) and control (protected sex) intervals. In both intervals, partnerships initiated offline were less likely to be with one-time partners (unprotected anal sex acts: 24.7% vs 52.6%, *P*<0.001; protected anal sex acts: 30.8% vs 54.9%, *P*<.001) and more likely to be with a partner who the participant believed to be HIV negative (unprotected anal sex acts: 42.1% vs 26.7%, *P*<.001; protected anal sex acts: 41.6% vs 23.2%, *P*<.001) than partnerships initiated online. In partnerships initiated offline, the participant was less likely to only engage in receptive anal sex (40.7% vs 49.3%, *P*=.02) and more likely to engage in substance use prior to having sex (10.7% vs 6.7%, *P*=.04) than in partnerships initiated online in the protected anal sex interval.

Table 2. Interval-level characteristics of sexual acts among 1311 adult Blued app users in 4 provinces in China who reported engaging in both unprotected and protected anal sex in the past 6 months.

Characteristic	All participants (N=1311)		UAI[a]			No UAI[b]		
	UAI[a], n (%)	No UAI[b], n (%)	Offline[c] (n=292), n (%)	Online[d] (n=1019), n (%)	*P* value	Offline[c] (n=214), n (%)	Online[d] (n=1097), n (%)	*P* value
Partnership type								
One-time partner	608 (46.4)	668 (51.0)	72 (24.7)	536 (52.6)	<.001	66 (30.8)	602 (54.9)	<.001
Casual partner	433 (33.0)	410 (31.3)	136 (46.6)	297 (29.1)	<.001	90 (42.1)	320 (29.2)	<.001
Main partner	270 (20.6)	233 (17.8)	84 (28.8)	186 (18.3)	<.001	58 (27.1)	175 (16.0)	<.001
Partner's HIV status								
Negative	395 (30.1)	344 (26.2)	123 (42.1)	272 (26.7)	<.001	89 (41.6)	255 (23.2)	<.001
Positive	45 (3.4)	55 (4.2)	12 (4.1)	33 (3.2)	.47	14 (6.5)	41 (3.7)	.06
Not sure	871 (66.4)	912 (69.6)	157 (53.8)	714 (70.1)	<.001	111 (51.9)	801 (73.0)	<.001
Participant's sexual role								
Receptive	622 (47.4)	628 (47.9)	124 (42.5)	498 (48.9)	.05	87 (40.7)	541 (49.3)	.02
Insertive	524 (40.0)	501 (38.2)	130 (44.5)	394 (38.7)	.07	92 (43.0)	409 (37.3)	.12
Both	165 (12.6)	182 (13.9)	38 (13.0)	127 (12.5)	.80	35 (16.4)	147 (13.4)	.25
Participant's substance use before sex (vs no use)	89 (6.8)	96 (7.3)	25 (8.6)	64 (6.3)	.17	23 (10.7)	73 (6.7)	.04

[a]UAI: unprotected anal sex interval.

[b]No UAI: protected anal sex interval.

[c]Offline: initiated partnership offline.

[d]Online: initiated partnership online.

Associations Between Interval-Level Partnership Characteristics and Unprotected Anal Sex

Table 3 presents data on associations between interval-level covariates and unprotected anal sex. In multivariable analyses, the belief that partners were HIV negative (adjusted odds ratio [aOR] 1.57, 95% CI 1.09 to 2.27, *P*=.02) was associated with unprotected anal sex, compared with not being sure about the partner's HIV status. Our primary outcome, initiating a partnership offline, was positively associated with unprotected anal sex (aOR 2.66, 95% CI 1.84 to 3.85, *P*<.001), compared with initiating a partnership online.

Table 3. Interval-level characteristics and unprotected anal sex among 1311 adult Blued app users in 4 provinces in China.

Characteristic	Bivariable model				Multivariable model[a]			
	OR[b]	95% CI	Wald χ^2	*P* value	aOR[c]	95% CI	Wald χ^2	*P* value
Partnership initiated offline (vs online)[d]	2.95	2.06-4.22	34.96	<.001	2.66	1.84-3.85	26.84	<.001
Partnership type[e]								
One-time partner	Reference				Reference			
Casual partner	1.41	1.08-1.83	6.51	.01	1.16	0.87-1.54	1.06	.30
Main partner	1.68	1.24-2.27	11.13	.001	1.39	1.00-1.93	3.75	.05
Partner's HIV status[f]								
Negative	1.92	1.38-2.68	14.80	<.001	1.57	1.09-2.27	5.81	.02
Positive	0.67	0.35-1.31	1.35	.25	0.58	0.29-1.15	2.45	.12
Not sure	Reference				Reference			
Participant's sexual role[g]								
Receptive	Reference				Reference			
Insertive	1.42	0.93-2.16	2.68	.10	1.47	0.95-2.28	3.01	.08
Both	0.77	0.48-1.24	1.15	.28	0.74	0.45-1.19	1.55	.21
Participant's substance use before sex (vs no use)[h]	0.72	0.39-1.32	1.13	.29	0.58	0.31-1.10	2.81	.09

[a]Model log-likelihood: –876.1.
[b]OR: odds ratio.
[c]aOR: adjusted odds ratio.
[d]Model log-likelihood: –888.6.
[e]Model log-likelihood: –902.0.
[f]Model log-likelihood: –899.6.
[g]Model log-likelihood: –905.6.
[h]Model log-likelihood: –908.1.

Sensitivity Analyses

We conducted four sensitivity analyses in order to assess other potential biases. First, not including partnership type as a covariate in multivariable analyses had virtually no effect on the results (Multimedia Appendix 1). Second, the associations between initiating a partnership offline and unprotected anal sex stratified by partnership type were all positive and were not significantly different from each other (Multimedia Appendix 2). Third, restricting the sample to participants who met sexual partners both online and offline had virtually no effect on the results (Multimedia Appendix 3). Fourth, restricting the sample to participants whose self-reported HIV status was either negative or unknown had virtually no effect on the results (Multimedia Appendix 4).

Discussion

Principal Findings

In numerous previous assessments observing app users and comparing them with nonapp users, including several studies conducted in China, the causal model proposed was that exposure to a dating app caused an increase in HIV risk behaviors, which in turn caused increased HIV incidence [11-15,17,22]. We hypothesize that an underlying factor might make individuals who download, install, create an account, and use a dating app inherently different from those who do not: intention to have unprotected sex. In this concept, the app is a venue (akin to a bar or other meeting place), with intention to have unprotected sex causing both app use and subsequent unprotected sex with a higher HIV transmission risk. This analysis provides the first data to take into account the intention to have unprotected sex by those using dating apps by using a case-crossover design. Our primary finding—consistent across bivariable, multivariable, and sensitivity analyses—is that individuals were more likely to have unprotected anal sex in partnerships that they initiated offline compared with partnerships that they initiated online. The relationship was substantial, with over 2.5 times increased likelihood of engaging in unprotected anal sex in partnerships initiated offline compared with those initiated online.

Although our findings are in contrast to previous assessments that were mentioned, there are also studies that have reported similar results. For example, a study among young MSM using Grindr (a geosocial online dating app) found significantly higher rates of condom use with partners met on Grindr relative to partners that they met elsewhere [18]. However, these same authors reported in a subsequent study that familiarity with Grindr (using it for at least 1 year and meeting more partners through Grindr in the past month) was associated with increased sexual risk behavior (ie, condomless anal intercourse with their most recent Grindr-met partner) [16]. Specific to China, a study found that MSM who used partner-seeking mobile apps (eg, Jack'd, Grindr, Blued) did not report more condomless sex than men who did not use apps [19]. The current study improves upon these previous studies by looking within app users and employing a case-crossover design to avoid making potentially biased comparisons.

We also found that participants in a large majority of partnerships (68.0%) were not sure about their partner's HIV status. There are two potential explanations for this. One explanation is that there is a high proportion of Chinese MSM who do not know their HIV status, which was reported in two previous studies, with 39.1% [25] and 33.2% [26] of MSM having never been tested for HIV. In this study, nearly one-third (30.3%) of the 4672 MSM from the entire sample had never been tested for HIV, and more than one-half (57.1%) had not been tested for HIV in the previous 6 months. Another potential explanation for the high frequency of sexual events among partners with unknown HIV status is that there is limited communication about and disclosure of HIV status among MSM prior to having sex, which has been reported previously [27]. Furthermore, partnerships were more likely to feature unprotected anal sex when a participant believed that his partner was HIV negative (compared with participants reporting that they were unsure of their partner's HIV status). This type of safer sex negotiation in light of perceived risk has been reported previously in numerous studies [28,29].

Limitations

We acknowledge several limitations in our study. First, data were cross-sectional and required recall from the previous 6 months (although we targeted a specific event in that time span, potentially reducing recall bias). Recall bias, if any, would be expected to bias toward the null. Second, data were based on self-report, and unprotected anal sex may have been under-reported. Again, this would be expected to bias toward the null. Third, although the comparisons made were within Blued app users and interval-level covariates were controlled for in multivariable analyses, there is the possibility that even though app use appears to be associated with condom use, this may not generalize to all sexual risk behaviors. For example, app users could have an elevated risk of HIV transmission through other means, such as by having more sexual partners.

Conclusions

These limitations notwithstanding, ours is the first study to take into account the inherent intentions of individuals using geosocial networking apps, compared with those not using such apps, when looking at their effect on sexual risk behavior. Our assertion is that the factor motivating individuals to download, install, create an account, and use a dating app (ie, intent to have unprotected sex) contributes to the associations between app use and sexual risk behavior reported in previous studies. These findings have important policy implications. If dating apps are believed to contribute to increased risk of HIV transmission, then these apps might be perceived as a problem to be addressed by health agencies. However, if instead apps are a proxy for, and not a cause of, increased HIV risk, which our results suggest, then dating app platforms should be viewed as a useful venue to identify individuals at increased risk for HIV transmission to allow for targeted service provision such as HIV testing, condoms, and PrEP.

Acknowledgments

This work was supported by the National Institute of Allergy and Infectious Diseases (R01AI143875) and the National Institute of Mental Health (R01MH114692). The work was facilitated by the Emory Center for AIDS Research (P30AI050409). Funding for Dr Knox's contribution to the present study was supported by NIDA (T32DA031099; PI: Hasin). We would also like to express gratitude to Dr Mitchel Klein, who assisted us with the statistical analysis used in the case-crossover analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interval-level characteristics (excluding partnership type) and unprotected anal sex among 1311 adult Blued app users in 4 provinces in China.
[DOCX File , 14 KB - publichealth_v7i1e17173_app1.docx]

Multimedia Appendix 2
Associations between initiating a partnership offline and unprotected anal sex in three partnership-type strata among 1311 adult Blued app users in 4 provinces in China.
[DOCX File , 14 KB - publichealth_v7i1e17173_app2.docx]

Multimedia Appendix 3

Interval-level characteristics and unprotected anal sex among 158 adult Blued app users in 4 provinces in China who initiated partnerships both offline and online.
[DOCX File , 16 KB - publichealth_v7i1e17173_app3.docx]

Multimedia Appendix 4
Interval-level characteristics and unprotected anal sex among 1151 adult Blued app users in 4 provinces in China who did not self-report as HIV positive (HIV negative or unknown status).
[DOCX File , 15 KB - publichealth_v7i1e17173_app4.docx]

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
2. Zhou Y, Li D, Lu D, Ruan Y, Qi X, Gao G. Prevalence of HIV and syphilis infection among men who have sex with men in China: a meta-analysis. Biomed Res Int 2014;2014:620431 [FREE Full text] [doi: 10.1155/2014/620431] [Medline: 24868533]
3. Liu G, Lu H, Wang J, Xia D, Sun Y, Mi G, et al. Incidence of HIV and Syphilis among Men Who Have Sex with Men (MSM) in Beijing: An Open Cohort Study. PLoS One 2015;10(10):e0138232 [FREE Full text] [doi: 10.1371/journal.pone.0138232] [Medline: 26426271]
4. van Griensven F, Holtz TH, Thienkrua W, Chonwattana W, Wimonsate W, Chaikummao S, et al. Temporal trends in HIV-1 incidence and risk behaviours in men who have sex with men in Bangkok, Thailand, 2006-13: an observational study. Lancet HIV 2015 Feb;2(2):e64-e70. [doi: 10.1016/S2352-3018(14)00031-9] [Medline: 26424462]
5. Lane T, Osmand T, Marr A, Struthers H, McIntyre JA, Shade SB. Brief Report: High HIV Incidence in a South African Community of Men Who Have Sex With Men: Results From the Mpumalanga Men's Study, 2012-2015. J Acquir Immune Defic Syndr 2016 Dec 15;73(5):609-611. [doi: 10.1097/QAI.0000000000001162] [Medline: 27851715]
6. Sullivan PS, Rosenberg ES, Sanchez TH, Kelley CF, Luisi N, Cooper HL, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. Ann Epidemiol 2015 Jun;25(6):445-454 [FREE Full text] [doi: 10.1016/j.annepidem.2015.03.006] [Medline: 25911980]
7. Cui Y, Guo W, Li D, Wang L, Shi CX, Brookmeyer R, et al. Estimating HIV incidence among key affected populations in China from serial cross-sectional surveys in 2010-2014. J Int AIDS Soc 2016;19(1):20609 [FREE Full text] [doi: 10.7448/IAS.19.1.20609] [Medline: 26989062]
8. National Health and Family Planning Commission of the People's Republic of China. 2015 China AIDS Response Progress Report. UNAIDS. 2015 May. URL: https://www.unaids.org/sites/default/files/country/documents/CHN_narrative_report_2015.pdf [accessed 2020-12-28]
9. Patel, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS 2014 Jul 19;28(10):1509-1519 [FREE Full text] [doi: 10.1097/QAD.0000000000000298] [Medline: 24809629]
10. Baggaley RF, White RG, Boily MC. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. Int J Epidemiol 2010 Aug;39(4):1048-1063 [FREE Full text] [doi: 10.1093/ije/dyq057] [Medline: 20406794]
11. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
12. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
13. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, et al. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional sample of men who have sex with men in Los Angeles County. Sex Transm Infect 2014 Nov;90(7):567-572 [FREE Full text] [doi: 10.1136/sextrans-2013-051494] [Medline: 24926041]
14. Holloway IW, Pulsipher CA, Gibbs J, Barman-Adhikari A, Rice E. Network Influences on the Sexual Risk Behaviors of Gay, Bisexual and Other Men Who Have Sex with Men Using Geosocial Networking Applications. AIDS Behav 2015 Jan 9;19(S2):112-122. [doi: 10.1007/s10461-014-0989-3]
15. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup App Use, Sexual Behavior, and Sexual Health Among Adolescent Men Who Have Sex With Men in the United States. J Adolesc Health 2018 Jun;62(6):708-715 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.01.001] [Medline: 29784114]
16. Winetrobe H, Rice E, Bauermeister J, Petering R, Holloway IW. Associations of unprotected anal intercourse with Grindr-met partners among Grindr-using young men who have sex with men in Los Angeles. AIDS Care 2014;26(10):1303-1308. [doi: 10.1080/09540121.2014.911811] [Medline: 24754563]

17. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, National MSM Survey Group. HIV and syphilis prevalence among men who have sex with men: a cross-sectional survey of 61 cities in China. Clin Infect Dis 2013 Jul;57(2):298-309 [FREE Full text] [doi: 10.1093/cid/cit210] [Medline: 23580732]
18. Rice E. Sex Risk among Young Men who have Sex with Men who use Grindr, a Smartphone Geosocial Networking Application. J AIDS Clinic Res 2012;01(S4). [doi: 10.4172/2155-6113.s4-005]
19. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay Apps for Seeking Sex Partners in China: Implications for MSM Sexual Health. AIDS Behav 2015 Jul;19(6):941-946 [FREE Full text] [doi: 10.1007/s10461-014-0994-6] [Medline: 25572834]
20. Silver L. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-01-04]
21. McCarthy N. China Now Boasts More Than 800 Million Internet Users And 98% Of Them Are Mobile [Infographic]. Forbes. 2018 Aug 23. URL: https://www.forbes.com/sites/niallmccarthy/2018/08/23/china-now-boasts-more-than-800-million-internet-users-and-98-of-them-are-mobile-infographic/?sh=4136d4097092 [accessed 2021-01-04]
22. Xu J, Yu H, Tang W, Leuba SI, Zhang J, Mao X, et al. The Effect of Using Geosocial Networking Apps on the HIV Incidence Rate Among Men Who Have Sex With Men: Eighteen-Month Prospective Cohort Study in Shenyang, China. J Med Internet Res 2018 Dec 21;20(12):e11303 [FREE Full text] [doi: 10.2196/11303] [Medline: 30578225]
23. Hernandez J. Building a Community, and an Empire, With a Gay Dating App in China. The New York Times. 2016 Dec 16. URL: https://www.nytimes.com/2016/12/16/world/asia/building-a-community-and-an-empire-with-a-gay-dating-app-in-china.html [accessed 2020-12-28]
24. Blued. URL: https://www.blued.com/ [accessed 2020-12-28]
25. Han L, Wei C, Muessig KE, Bien CH, Meng G, Emch ME, et al. HIV test uptake among MSM in China: Implications for enhanced HIV test promotion campaigns among key populations. Glob Public Health 2017 Jan;12(1):31-44 [FREE Full text] [doi: 10.1080/17441692.2015.1134612] [Medline: 26785328]
26. Li X, Lu H, Raymond HF, Sun Y, Jia Y, He X, et al. Untested and undiagnosed: barriers to HIV testing among men who have sex with men, Beijing, China. Sex Transm Infect 2012 May;88(3):187-193. [doi: 10.1136/sextrans-2011-050248] [Medline: 22158932]
27. Knox J, Reddy V, Kaighobadi F, Nel D, Sandfort T. Communicating HIV status in sexual interactions: assessing social cognitive constructs, situational factors, and individual characteristics among South African MSM. AIDS Behav 2013 Jan;17(1):350-359 [FREE Full text] [doi: 10.1007/s10461-012-0337-4] [Medline: 23065127]
28. Yang C, Latkin C, Tobin K, Seal D, Koblin B, Chander G, et al. An Event-Level Analysis of Condomless Anal Intercourse with a HIV-Discordant or HIV Status-Unknown Partner Among Black Men Who Have Sex with Men from a Multi-site Study. AIDS Behav 2018 Jul;22(7):2224-2234 [FREE Full text] [doi: 10.1007/s10461-018-2161-y] [Medline: 29779160]
29. Siegler AJ, Sullivan PS, Khosropour CM, Rosenberg ES. The role of intent in serosorting behaviors among men who have sex with men sexual partnerships. J Acquir Immune Defic Syndr 2013 Nov 01;64(3):307-314 [FREE Full text] [doi: 10.1097/QAI.0b013e3182a0e880] [Medline: 23846562]

Abbreviations

aOR: adjusted odds ratio
MSM: men who have sex with men
NCAIDS: National Center for AIDS/STD Control and Prevention
OR: odds ratio
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by Z Ma, A Bwanika Naggirinya; comments to author 10.02.20; revised version received 06.03.20; accepted 09.12.20; published 15.01.21.

Please cite as:
Knox J, Chen YN, He Q, Liu G, Jones J, Wang X, Sullivan P, Siegler A
Use of Geosocial Networking Apps and HIV Risk Behavior Among Men Who Have Sex With Men: Case-Crossover Study
JMIR Public Health Surveill 2021;7(1):e17173
URL: http://publichealth.jmir.org/2021/1/e17173/
doi:10.2196/17173
PMID:33448934

Original Paper

A Moderated Mediation Analysis of Condom Negotiation and Sexual Orientation on the Relationship Between Sexual Coercion and Condom Use in Chinese Young Women: Cross-Sectional Study

Wen Zhang[1], MNurs; Edmond Pui Hang Choi[1], PhD; Daniel Yee-Tak Fong[1], PhD; Janet Yuen-Ha Wong[1], PhD

School of Nursing, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Hong Kong

Corresponding Author:
Janet Yuen-Ha Wong, PhD
School of Nursing
Li Ka Shing Faculty of Medicine
University of Hong Kong
21 Sassoon Road
Pokfulam
Hong Kong
Hong Kong
Phone: 852 3917 6641
Email: janetyh@hku.hk

Abstract

Background: The high prevalence of sexual coercion against young women has become a significant public health issue in China and other regions around the world. Young women are also especially vulnerable to engage in inconsistent condom use because of low sexual control. Although the relationship between sexual coercion and condom use has been widely demonstrated, the mechanism of this relationship is still unclear.

Objective: The objective of this study was to test condom negotiation as a mediator of the relationship between sexual coercion and condom use in young Chinese women and to investigate whether sexual orientation is a moderator.

Methods: Data were collected using web-based questionnaires and a total of 402 young Chinese women were included in the analysis. Sexual coercion was measured using a subscale of the Revised Conflict Tactics Scales and condom negotiation was measured using a subscale of the UCLA Multidimensional Condom Attitudes Scale. Sexual orientation was assessed using an item adopted from a previous study and condom use was calculated by the total number of times condoms were used divided by the total number of times sexual intercourse was engaged in during the past 3 months. Moderated mediation analyses were conducted with sexual coercion as the independent variable, condom use consistency as the dependent variable, condom negotiation as the mediator variable, and sexual orientation as a moderator.

Results: The moderated mediation analysis indicated that the relationship between sexual coercion and condom use was significantly mediated by condom negotiation and moderated by sexual orientation. The indirect effect of condom negotiation was significant in heterosexual women (indirect effect: –0.80, 95% boot CI –1.67 to –0.36) but not in sexual minority women (indirect effect: –0.33, 95% boot CI –0.86 to 0.31).

Conclusions: The results showed that sexual orientation meaningfully affects the relationship between sexual coercion and condom negotiation. The difference in the mechanism of the relation between sexual coercion and sexual behaviors in heterosexual and sexual minority women should be considered for future research and interventions aimed at mitigating the adverse effects of sexual coercion.

(JMIR Public Health Surveill 2021;7(1):e24269) doi:10.2196/24269

KEYWORDS

sex offenses; sex orientation; unprotected sex; online research; women’s health

Introduction

Sexual coercion against women remains a significant global health problem [1]. Previous studies have defined sexual coercion as behaviors, ranging from verbal manipulation to physical force, employed to complete or attempt sexual activities without the partner's free consent [2,3]. A national survey in the United States found that approximately one-fifth of women reported experiencing sexual violence in their lifetime; one-half of these women reported that intimate partners were the offenders [4]. This national survey further indicated that more than 1 in every 3 female survivors of rape was first raped in her college-aged years (18-24 years) [4]. According to research by Planty et al [5], the risk of sexual coercion was higher in the age group of 18 to 34 years than in other age groups.

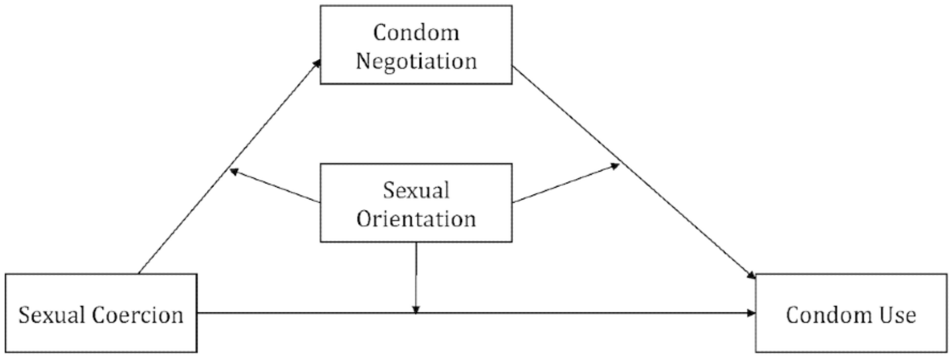
Sexual coercion against young women in China has also become an emerging public health issue that deserves attention [6]. The prevalence of sexual coercion against Chinese college women was approximately 13% in 2008 [7], and a similar prevalence was found in 2015, despite the improved status of women in Hong Kong, China [8]. Young women are also especially vulnerable to inconsistent condom use, since low sexual control has consistently been reported in young Chinese women [9]. A national survey in China reported that 1.6% of Chinese female college students had multiple sexual partners [10] and another study reported that only 17.2% of sexually active college women in China consistently used condoms [11]. An association between sexual coercion and inconsistent condom use has been observed [12,13]. Most of the data indicated that individuals with a history of sexual coercion (versus those without) reported a higher level of inconsistent condom use, which resulted in a higher risk of contracting sexually transmitted infections (STIs) [14]. Although previous studies have explored the mechanism of the relation between a history of sexual violence and condom use in female sex workers [15] and HIV-positive women [16], the specific mechanism of the relation between a history of sexual coercion and condom use in college women remains unclear. To improve interventions aimed at reducing sexual risk among college women, it is necessary to understand the mechanism underlying the relation between sexual coercion and condom use in this group.

Condom negotiation is one of the strongest predictors of condom use [17]. A previous study indicated that condom negotiation might play a crucial role in the relation between sexual violence and condom use [15]. Condom negotiation is closely related to condom use by college women [18], and women with a history of sexual coercion are less likely to negotiate or use condoms than women without experiences of sexual coercion. According to the traumagenic dynamics model [19,20], sexual coercion is viewed as a traumatic event with psychological sequelae, such as a negative attitude arising from the powerlessness experienced during sexual coercion. This negative attitude then contributes to maladaptive behavioral patterns. Women with an abusive experience could be at a disadvantage in their condom negotiations with their sexual partners because they seek to avoid nonphysical coercion [21]. Thus, they may be more likely to have limited or no control over condom decision making, which contributes to inconsistent condom use. Taken together, these patterns suggest that condom negotiation is a potential mediator between sexual coercion and condom use.

Previous studies found that sexual minority women (women who identify as having a sexual orientation other than heterosexual or who engage in same-sex sexual behavior, experience same-sex attraction, or self-identify as lesbian or bisexual [22]) experience a significantly higher incidence of sexual coercion than heterosexual women [23]. This suggests that although the experience of sexual coercion has a disincentivizing effect on the consistency of condom use, it does not affect all women equally. Sexual orientation also plays a role in individuals' condom use and the negotiation process. Young bisexual women exhibited a greater likelihood of inconsistent condom use in vaginal intercourse than heterosexual women [24]. Skakoon-Sparling and Cramer [25] found that the process of condom negotiation can be impacted by sexual orientation. This finding might suggest that personal characteristics such as sexual orientation moderate the association between the experience of sexual coercion and consistency of condom use.

In this study, we explored the relationship between sexual coercion and condom use in a sample of Chinese college women. Based on previous literature, we investigated the mediating effect of condom negotiation on the relation between sexual coercion and condom use in this population. Unique to this study, we tested whether sexual orientation moderated the hypothesized mediation of the relation between the experience of sexual coercion and condom use by condom negotiation. The proposed moderated mediation model is shown in Figure 1.

Figure 1. Proposed moderated mediation model.



Methods

Data Collection

The baseline data of an interactive computer-based intervention (ICBI) project [26] were used to perform the mediated moderation analysis. This project was a randomized controlled trial that estimated the relative effectiveness of an ICBI and the provision of basic information in terms of promoting consistent condom use. The baseline data were collected from September 2018 to December 2018. The protocol for the parent study was approved by Institutional Review Boards (IRB NO.UW-17029) and is registered at ClinicalTrials.gov (NCT03695679). Signed informed consent was obtained from each participant via the study's website. Data were collected via an anonymous web-based survey conducted at 5 universities in Hong Kong SAR, China. Vouchers of HK $300 (US$38.69) were delivered to participants who completed the project; participants who only completed the baseline questionnaire did not receive vouchers.

Participants and Procedures

We recruited female students from 5 universities in Hong Kong by bulk email using the corresponding institution's bulk email delivery service. In addition, we displayed posters on the campuses and set up campus booths to distribute leaflets. Students who were interested in the study were asked to complete a registration form with their contact information via a Google Form. The invitation email with the website registration information was sent to all interested students, and they were then screened after they logged into the website. Participants who met the inclusion criteria were asked to provide informed consent; they then completed a baseline assessment. Participants could not submit the questionnaire if any information was missing. The inclusion criteria in this study were female college students who were aged 18 years or older, were unmarried, reported having intimate partners in the past 12 months, and had engaged in sexual activity in the past 3 months. Women were excluded if they were unwilling to complete the questionnaire, were pregnant or had recently given birth, or had psychiatric illness. We screened 1503 students, of whom 805 did not meet the eligibility criteria and 292 refused to participate. Of the 406 eligible participants, 4 participants were excluded after data checking (3 participants reported having no sexual experience but also reported engaging in sexual activity in the past 3 months and 1 participant had a missing sexual coercion scale because of a technical problem), giving a validity rate of 99.0%. Ultimately, 402 female university students were included in the study. The average age of the participants was 21.90 (SD 2.74) years and the average age at first sexual intercourse was 19.48 (SD=2.40) years. Among the 402 participants, 87.6% (n=352) had never smoked, 10.7% (n=43) were quitting smoking, and 1.7% (n=7) were smokers; 33.3% (n=134) never drank, 16.4% (n=66) were quitting drinking, and 50.2% (n=202) drank. Approximately 70.6% (284/402) of the participants were born in Hong Kong and 29.4% (118/402) were born elsewhere.

Measures

Sexual coercion was measured using a 7-item subscale of the Revised Conflict Tactics Scales [27]. Participants rated items to indicate how often the behavior occurred during the past year on a 7-point Likert scale, where higher scores indicated higher frequency. This scale has been widely used in the Chinese population and has shown a satisfactory reliability [28]. In this study, Cronbach was 0.63 for this subscale.

Sexual orientation was assessed using an item that was adopted from the longitudinal Growing Up Today Study [29], which had been ongoing since 1996. There were 6 options: completely heterosexual (attracted only to the opposite sex), mostly heterosexual, bisexual (attracted to both the opposite and the same sex), mostly homosexual, completely homosexual (attracted only to the same sex), and unsure. Referring to the definition of sexual minority women [22] and a previous study in the Chinese population [30], completely heterosexual was coded as "heterosexual," and mostly heterosexual, bisexual, mostly homosexual, completely homosexual, and unsure were combined into a "sexual minority" group.

Condom use was measured by the consistency of condom use, which was defined as the total number of times condoms were used during vaginal intercourse divided by the total number of times vaginal intercourse occurred in the past 3 months. This assessment was recommended by a systematic review of condom use measurement that examined 56 studies of sexual risk behavior [31].

Condom negotiation was measured using a subscale of the UCLA Multidimensional Condom Attitudes Scale [32]. This subscale is used to evaluate attitudes toward condom negotiation and use (eg, “When I suggest using a condom, I am almost always embarrassed,” “I am comfortable talking about condoms with my partner,” “I never know what to say when my partner and I need to talk about condoms or other protection,” and “It is easy to suggest to my partner that we use a condom”). These items were answered using a 7-point Likert scale from “strongly agree” to “strongly disagree.” Higher scores indicate a more positive attitude regarding communication and negotiation of condom use. A previous study has shown acceptable validity and reliability in the Chinese population [33]. In this study, Cronbach was 0.87.

Demographic variables examined in the study included participant characteristics such as age, age at first sexual intercourse, smoking status, drinking status, and place of birth.

Statistical Analyses

Descriptive statistics and bivariate correlation analyses were conducted of the studied variables as preliminary analyses. The Shapiro-Wilk test was used as a test of normality and $P<.05$ was considered evidence for nonnormality. For skewed data, the median and IQR were used to describe the nonnormal variables, the Mann-Whitney U test was used to test the difference between the heterosexual group and the sexual minority group, and Spearman rank correlation analyses were conducted to identify the correlations between the nonnormal variables. The mediation effect of condom negotiation was tested using model 4 of Hayes’ PROCESS macro for SPSS (version 25.0; IBM Corp) [34]. Moderated mediation analysis was conducted using model 59 of PROCESS to examine whether the indirect path was moderated by sexual orientation [34]. Since the data on the consistency of condom use may be nonnormally distributed, a bootstrapping procedure with 5000 samples was used to test the proposed conditional direct and indirect effects using the PROCESS macro for SPSS. Age and age at first sexual intercourse were added as covariates.

Results

Normality and Description of the Study Variables

The results of the Shapiro-Wilk test showed that condom use ($P<.001$), sexual coercion ($P<.001$), and condom negotiation ($P<.001$) were not normally distributed. The descriptive statistics and differences in the study variables between the heterosexual group and the sexual minority group are presented in Table 1. The results of the Mann-Whitney U test indicated that there was a significant difference in condom use (U=7841, Z=–6.23, $P<.001$) and condom negotiation (U=10740.5, Z=–2.43, $P=.02$), between the heterosexual group and the sexual minority group. Further, Spearman rank correlation analyses showed that in the heterosexual group, those who had more frequent sexual coercion experiences reported significantly less condom use ($r_s=-0.36$, $P<.001$) and were less positive about condom negotiation ($r_s=-0.28$, $P<.001$); there was a significant positive correlation between condom use and condom negotiation ($r_s=0.30$, $P<.001$). In the sexual minority group, only condom use was positively related to condom negotiation ($r_s=0.24$, $P=.03$).

Table 1. Description of the study variables and results of the Mann-Whitney U test.

Variables	Sexual orientation		Mann-Whitney U test		
	Heterosexual (n=321), median (IQR)	Sexual minority (n=81), median (IQR)	U	Z	*P* value
Sexual coercion	0 (2)	0 (0)	12428.5	–0.76	.45
Condom use	100 (33.33)	0 (100)	7841	–6.23	<.001
Condom negotiation	30 (9)	28 (9.5)	10740.5	–2.43	.02

Tests of Mediation

The results of the mediation analysis regarding sexual coercion and condom use, after adjustments for age and age at first sexual intercourse, showed that an experience of sexual coercion was a negative predictor of condom negotiation (coefficient a=–0.16, 95% boot CI –0.31 to –0.10) (Table 2), indicating that participants who experienced sexual coercion were less likely to engage in condom negotiation. Condom negotiation was a positive predictor of condom use (coefficient b=2.02, 95% boot CI 1.32 to 2.70), which indicated that participants who were more positive about condom negotiation were more likely to be consistent in terms of condom use. A significant indirect and negative effect of sexual coercion on the consistency of condom use through condom negotiation was found (indirect effect: coefficient a=–0.32, 95% boot CI –0.67 to –0.18). The direct effect of sexual coercion on condom use became nonsignificant (coefficient c=–0.33, 95% boot CI –0.86 to 0.31). The indirect effect accounted for 49.2% of the total effect of sexual coercion on condom use.

Table 2. Mediation results for condom negotiation.

Outcome, Predictor	Coefficient	Boot SE	Boot LLCI[a]	Boot ULCI[b]
Condom negotiation				
Sexual coercion	–0.16	0.06	–0.31	–0.10
Condom use				
Sexual coercion	–0.33	0.29	–0.86	0.31
Condom negotiation	2.02	0.35	1.32	2.70
Direct effect	–0.33	0.29	–0.86	0.31
Indirect effect	–0.32	0.13	–0.67	–0.18
Total effect	–0.65	0.28	–1.25	–0.12

[a]LLCI: lower limit confidence interval.

[b]ULCI: upper limit confidence interval.

Tests of Moderated Mediation

After adjusting for age and age at first sexual intercourse, the results of the moderated mediation analyses for sexual coercion and condom use showed that the interaction term between sexual coercion and sexual orientation was significant (coefficient c= 0.36, 95% boot CI 0.16 to 0.74) (Table 3 and Figure 2), which suggested that sexual orientation moderated the association between sexual coercion and condom negotiation. To further explore the moderation effect, the conditional indirect effect of sexual coercion on condom use via condom negotiation was estimated by using the pick-a-point approach in both sexual orientation groups. A significant indirect effect was seen in the heterosexual group (effect = –0.80, 95% boot CI –1.67 to –0.36), while the indirect effect became insignificant in the sexual minority group (effect = –0.14, 95% boot CI –0.31 to 0.004) (Figure 3 and Table 4).

Table 3. The moderating effects of sexual orientation.

Outcome, Predictor	Coefficient	Boot SE	Boot LLCI[a]	Boot ULCI[b]
Condom negotiation				
Sexual coercion	–0.43	0.14	–0.80	–0.24
Sexual orientation	–1.96	0.77	–3.53	–0.50
Inter 1[c]	0.36	0.15	0.16	0.74
Condom use				
Sexual coercion	–0.29	0.68	–1.80	0.91
Condom negotiation	1.87	0.39	–1.10	2.62
Sexual orientation	–30.65	25.56	–78.20	23.27
Inter 1	0.13	0.78	–1.26	1.77
Inter 2[d]	0.04	0.90	–1.76	1.74

[a]LLCI: lower limit confidence interval.

[b]ULCI: upper limit confidence interval.

[c]Inter 1 = (sexual coercion) × (sexual orientation).

[d]Inter 2 = (condom negotiation) × (sexual orientation).

Figure 2. Tested moderated mediation model.

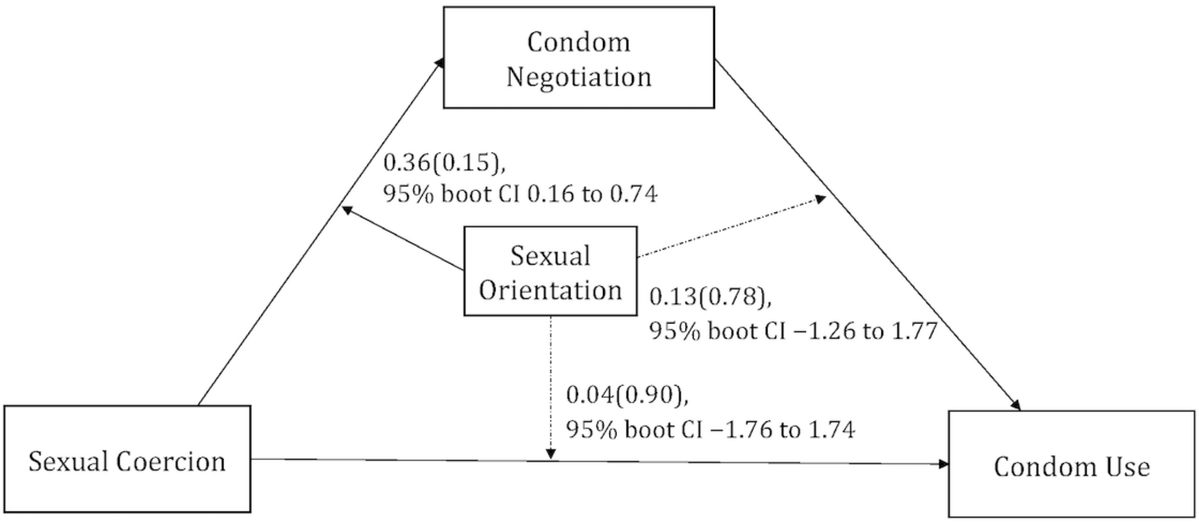


Figure 3. The moderation of the relationship between sexual coercion and condom negotiation by sexual orientation.

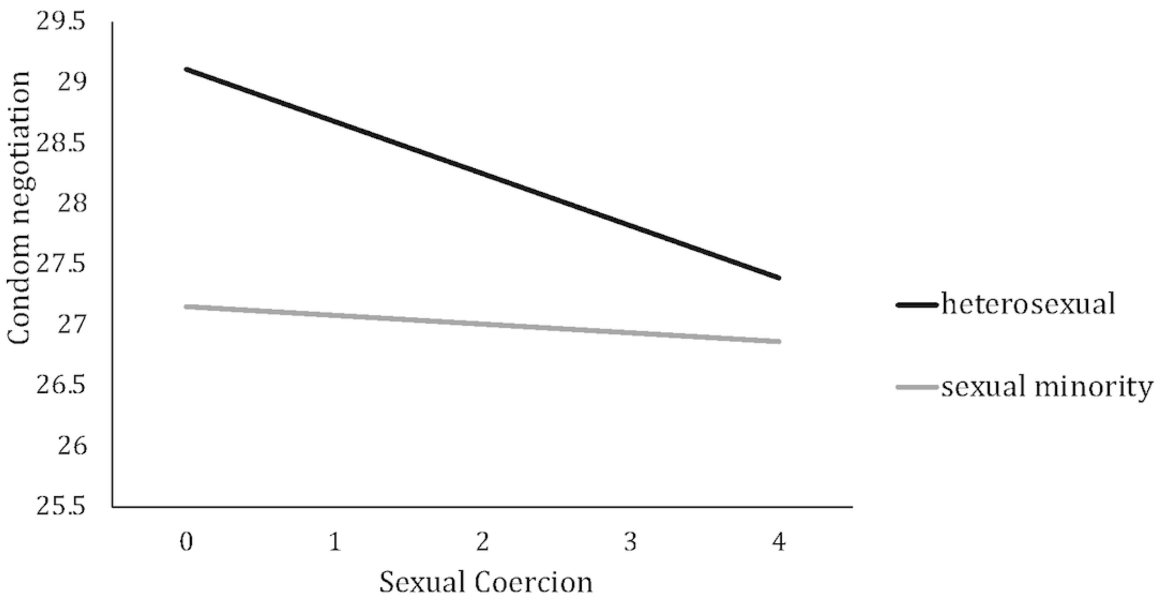


Table 4. Conditional indirect effects of the experience of sexual coercion on condom use.

Sexual orientation	Effect	Boot SE	Boot LLCI[a]	Boot ULCI[b]
Heterosexual	–0.80	0.33	–1.67	–0.36
Sexual minority	–0.14	0.10	–0.31	0.004

[a]LLCI: lower limit confidence interval.

[b]ULCI: upper limit confidence interval.

Discussion

Principal Findings

This study found a moderated mediation model of the pathway from sexual coercion to condom use via condom negotiation in a sample of Chinese female college students. Based on the Traumagenic Dynamics Model [20], the mediation effect of condom negotiation was tested and the results indicate that the relationship between sexual coercion and condom use is mediated by the level of condom negotiation. We found that a higher level of sexual coercion decreased condom negotiation, which in turn decreased condom use. The results were consistent with a previous study that was conducted on female sex workers from two West African countries [15].

In this study, sexual orientation moderated the indirect effect of sexual coercion on condom use. To our knowledge, this is the first known study to present such an intersectional analysis of the role of sexual orientation in the indirect effect of sexual coercion on condom use in young women. Altogether, these findings support the difference between heterosexual women and sexual minority women regarding the pattern of sexual behaviors in those who have experienced sexual coercion, emphasizing that sexual orientation meaningfully affects the relationship between sexual coercion and attitude toward condom negotiation. A significant indirect effect was found in the heterosexual women. This result is in line with a previous study in which the frequency of condom negotiation mediated the association between psychological intimate partner violence and condom use [35]. However, a nonsignificant indirect effect was found in sexual minority women, mainly because of the absence of condom negotiation. We also found no significant association between sexual coercion and condom negotiation or between condom negotiation and condom use in sexual minority women in this study. The belief that same-sex activities present a low risk for STIs is common in women who have sex with women (WSW) [36], and Formby [37] found that approximately 2 in 5 sexual minority women believe that they cannot get STIs from having sex with women. However, more recent research reported an infectious rate in WSW that was higher or similar to that in women who have sex exclusively with men [38,39]. The above mistaken belief mainly results in the absence of condom negotiation in WSW and their partners [40], which might contribute to the nonsignificant indirect effect in sexual minority women. Research by Walls [41] indicated that most sexual minority women seldom negotiate safe sex practices with their partners because they do not think they will contract STIs. This contributed to the nonsignificant indirect effect of condom negotiation on the relation between sexual coercion and condom use because condom negotiation might not be a critical factor affecting condom use in sexual minority women who have experienced sexual coercion. Instead of increasing condom negotiation skills, more information about the risks of STIs in sexual minority women and about appropriate protection methods should be provided.

Study Limitations

Our findings should be interpreted with caution. One limitation is that our measurements relied on self-reports of sensitive information and often stigmatized experiences and behaviors, even though an anonymous process was used to minimize social desirability bias. Self-reports of sensitive information, such as sexual coercion experiences, are vulnerable to cognitive and motivational processes that can bias recall-based responses [42]. The other limitation is that a small sample of sexual minority women was included in the data analysis. Small sample sizes are a common problem in the same-sex sexual violence research field [43]. Larger sample sizes of sexual minority women with regional and religious diversity are needed to increase the statistical power.

Future Work

Despite the above limitations, our study provided some new insights and implications for future studies examining condom use in women. One potential avenue for future research is to improve condom negotiation among sexual coercion survivors, given that the indirect effect of sexual negotiation accounted for nearly one-half (49.2%) of the total effect of sexual coercion on the consistency of condom use. The other implication is related to the different needs of women with different sexual orientations. In previous research, the difference in the mechanism of the relation between sexual coercion and sexual behaviors in heterosexual women and sexual minority women was ignored, and these two groups were included in the same intervention when addressing sexual coercion [44]. Our findings suggest that future interventions should not simply combine heterosexual women and sexual minority women. More qualitative and quantitative research to determine how sexual coercion experiences affect behavior changes in sexual minority women should be conducted.

Conclusions

Condom negotiation was found to mediate the association between sexual coercion and condom use in young women. A further moderated mediation emerged, with the indirect effects of sexual coercion on the consistency of condom use via condom negotiation differing between sexual minority women and heterosexual women. The results emphasized that sexual orientation meaningfully affects the relationship between sexual coercion and condom negotiation, and the different patterns for individuals of different sexual orientations should be considered for future research and for interventions designed to mitigate the adverse effects of sexual coercion.

Acknowledgments

We would like to thank the Food and Health Bureau, Hong Kong Special Administrative Region Government, for its support of the primary project through the Health and Medical Research Fund (Ref no: 14150971).

Conflicts of Interest

None declared.

References

1. Young B, Desmarais SL, Baldwin JA, Chandler R. Sexual Coercion Practices Among Undergraduate Male Recreational Athletes, Intercollegiate Athletes, and Non-Athletes. Violence Against Women 2017 Jun;23(7):795-812. [doi: 10.1177/1077801216651339] [Medline: 27247201]
2. Benbouriche M, Parent G. Sexual coercion: Thinking and understanding sexual violence beyond sexual offenders. Sexologies 2018 Apr;27(2):e15-e19. [doi: 10.1016/j.sexol.2018.02.001]
3. Abbey A, Wegner R, Woerner J, Pegram SE, Pierce J. Review of Survey and Experimental Research That Examines the Relationship Between Alcohol Consumption and Men's Sexual Aggression Perpetration. Trauma, Violence, & Abuse 2014 Apr 27;15(4):265-282. [doi: 10.1177/1524838014521031]
4. Smith SG, Zhang X, Basile KC, Merrick MT, Wang J, Kresnow MJ. The national intimate partner and sexual violence survey: 2015 data brief–updated release. National Center for Injury Prevention and Control. URL: https://www.cdc.gov/violenceprevention/pdf/2015data-brief508.pdf [accessed 2020-10-21]
5. Planty M, Langton L, Krebs C, Berzofsky MH, Smiley-McDonald H. Female victims of sexual violence, 1994-2010. Bureau of Justice Statistics. 2013. URL: https://www.bjs.gov/index.cfm?iid=4594&ty=pbdetail [accessed 2020-10-21]
6. Wang X, Ho SY. "Female Virginity Complex" Untied: Young Chinese Women's Experience of Virginity Loss and Sexual Coercion. Smith College Studies in Social Work 2011 Apr;81(2-3):184-200. [doi: 10.1080/00377317.2011.589336]
7. Chan KL, Straus MA. Prevalence and Correlates of Physical Assault on Dating Partners. TOSSJ 2008 Oct 30;1(1):5-14. [doi: 10.2174/1874945300801010005]
8. Wong JYH, Choi EPH, Lo HHM, Wong W, Chio JHM, Choi AWM, et al. Intimate Partner Sexual Violence and Mental Health Indicators Among Chinese Emerging Adults. J Interpers Violence 2019 Sep 03. [doi: 10.1177/0886260519872985] [Medline: 31478438]
9. Wang B, Davidson P. Sex, lies, and videos in rural China: a qualitative study of women's sexual debut and risky sexual behavior. J Sex Res 2006 Aug;43(3):227-235. [doi: 10.1080/00224490609552321] [Medline: 17599245]
10. Sun X, Liu X, Shi Y, Wang Y, Wang P, Chang C. Determinants of risky sexual behavior and condom use among college students in China. AIDS Care 2013;25(6):775-783. [doi: 10.1080/09540121.2012.748875] [Medline: 23252705]
11. Ma Q, Ono-Kihara M, Cong L, Pan X, Xu G, Zamani S, et al. Behavioral and psychosocial predictors of condom use among university students in Eastern China. AIDS Care 2009 Feb;21(2):249-259. [doi: 10.1080/09540120801982921] [Medline: 19229696]
12. Purdie MP, Abbey A, Jacques-Tiura AJ. Perpetrators of intimate partner sexual violence: are there unique characteristics associated with making partners have sex without a condom? Violence Against Women 2010 Oct;16(10):1086-1097 [FREE Full text] [doi: 10.1177/1077801210382859] [Medline: 20980229]
13. Coker AL, Hopenhayn C, DeSimone CP, Bush HM, Crofford L. Violence against Women Raises Risk of Cervical Cancer. J Womens Health (Larchmt) 2009 Aug;18(8):1179-1185. [doi: 10.1089/jwh.2008.1048] [Medline: 19630537]
14. Stockman JK, Campbell JC, Celentano DD. Sexual violence and HIV risk behaviors among a nationally representative sample of heterosexual American women: the importance of sexual coercion. J Acquir Immune Defic Syndr 2010;53(1):136-143 [FREE Full text] [doi: 10.1097/QAI.0b013e3181b3a8cc] [Medline: 19734802]
15. Wirtz AL, Schwartz S, Ketende S, Anato S, Nadedjo FD, Ouedraogo HG, et al. Sexual violence, condom negotiation, and condom use in the context of sex work: results from two West African countries. J Acquir Immune Defic Syndr 2015 Mar 01;68 Suppl 2:S171-S179. [doi: 10.1097/QAI.0000000000000451] [Medline: 25723982]
16. Clum GA, Chung S, Ellen JM, Perez LV, Murphy DA, Harper GW, et al. Victimization and sexual risk behavior in young, HIV positive women: exploration of mediators. AIDS Behav 2012 May;16(4):999-1010 [FREE Full text] [doi: 10.1007/s10461-011-9931-0] [Medline: 21452050]
17. Noar SM, Carlyle K, Cole C. Why communication is crucial: meta-analysis of the relationship between safer sexual communication and condom use. J Health Commun 2006 Jun;11(4):365-390. [doi: 10.1080/10810730600671862] [Medline: 16720536]
18. Holland KJ, French SE. Condom negotiation strategy use and effectiveness among college students. J Sex Res 2012;49(5):443-453. [doi: 10.1080/00224499.2011.568128] [Medline: 21732866]
19. Stockman JK, Lucea MB, Campbell JC. Forced sexual initiation, sexual intimate partner violence and HIV risk in women: a global review of the literature. AIDS Behav 2013 Mar;17(3):832-847 [FREE Full text] [doi: 10.1007/s10461-012-0361-4] [Medline: 23143750]
20. Finkelhor D. The Trauma of Child Sexual Abuse. J Interpers Violence 1987;2(4):348-366. [doi: 10.1177/088626058700200402]
21. Neilson EC, Gilmore AK, Stappenbeck CA, Gulati NK, Neilon E, George WH, et al. Psychological Effects of Abuse, Partner Pressure, and Alcohol: The Roles of in-the-Moment Condom Negotiation Efficacy and Condom-Decision Abdication on Women's Intentions to Engage in Condomless Sex. J Interpers Violence 2019 Jun 27. [doi: 10.1177/0886260519857160] [Medline: 31246143]
22. Farmer GW, Jabson JM, Bucholz KK, Bowen DJ. A population-based study of cardiovascular disease risk in sexual-minority women. Am J Public Health 2013;103(10):1845-1850 [FREE Full text] [doi: 10.2105/AJPH.2013.301258] [Medline: 23948018]

23. Edwards KM, Sylaska KM, Barry JE, Moynihan MM, Banyard VL, Cohn ES, et al. Physical dating violence, sexual violence, and unwanted pursuit victimization: a comparison of incidence rates among sexual-minority and heterosexual college students. J Interpers Violence 2015 Feb;30(4):580-600. [doi: 10.1177/0886260514535260] [Medline: 24923891]
24. Kerr DL, Ding K, Thompson AJ. A comparison of lesbian, bisexual, and heterosexual female college undergraduate students on selected reproductive health screenings and sexual behaviors. Womens Health Issues 2013;23(6):e347-e355. [doi: 10.1016/j.whi.2013.09.003] [Medline: 24183409]
25. Skakoon-Sparling S, Cramer KM. Are We Blinded by Desire? Relationship Motivation and Sexual Risk-Taking Intentions during Condom Negotiation. J Sex Res 2020;57(5):545-558. [doi: 10.1080/00224499.2019.1579888] [Medline: 30884967]
26. Wong J, Zhang W, Wu Y, Choi E, Lo H, Wong W. An Interactive Sexual Health Literacy Programme for Safe Sex Practice in Female Chinese University Students: A Multicentre, Randomised Controlled Trial. Journal of medical Internet Research Preprints. URL: https://preprints.jmir.org/preprint/22564/accepted [accessed 2021-01-04]
27. STRAUS MA, HAMBY SL, BONEY-McCOY S, SUGARMAN DB. The Revised Conflict Tactics Scales (CTS2). Journal of Family Issues 2016 Jun 30;17(3):283-316. [doi: 10.1177/019251396017003001]
28. Chan KL. Correlates of wife assault in Hong Kong Chinese families. Violence Vict 2004 Apr;19(2):189-201. [doi: 10.1891/vivi.19.2.189.64104] [Medline: 15384454]
29. Fredriksen-Goldsen KI, Kim H, Barkan SE, Balsam KF, Mincer SL. Disparities in health-related quality of life: a comparison of lesbians and bisexual women. Am J Public Health 2010 Nov;100(11):2255-2261. [doi: 10.2105/AJPH.2009.177329] [Medline: 20864722]
30. Wong JY, Choi EP, Lo HH, Wong W, Chio JH, Choi AW, et al. Dating violence, quality of life and mental health in sexual minority populations: a path analysis. Qual Life Res 2017 Apr;26(4):959-968. [doi: 10.1007/s11136-016-1415-2] [Medline: 27679496]
31. Noar SM, Cole C, Carlyle K. Condom use measurement in 56 studies of sexual risk behavior: review and recommendations. Arch Sex Behav 2006 Jun;35(3):327-345. [doi: 10.1007/s10508-006-9028-4] [Medline: 16799837]
32. Helweg-Larsen M, Collins BE. The UCLA Multidimensional Condom Attitudes Scale: documenting the complex determinants of condom use in college students. Health Psychol 1994 May;13(3):224-237. [Medline: 8055858]
33. Choi EPH, Fong DYT, Wong JYH. The use of the Multidimensional Condom Attitude Scale in Chinese young adults. Health Qual Life Outcomes 2020 Oct 08;18(1):331 [FREE Full text] [doi: 10.1186/s12955-020-01577-9] [Medline: 33032622]
34. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York City: Guilford publications; 2017. ISBN-10: 1462534651.
35. Peasant C, Sullivan TP, Ritchwood TD, Parra GR, Weiss NH, Meyer JP, et al. Words can hurt: The effects of physical and psychological partner violence on condom negotiation and condom use among young women. Women Health 2018;58(5):483-497 [FREE Full text] [doi: 10.1080/03630242.2017.1316345] [Medline: 28402194]
36. McNair R. Risks and prevention of sexually transmissible infections among women who have sex with women. Sex Health 2005;2(4):209-217. [doi: 10.1071/sh04046] [Medline: 16402667]
37. Formby E. Lesbian and bisexual women's human rights, sexual rights and sexual citizenship: negotiating sexual health in England. Cult Health Sex 2011 Nov;13(10):1165-1179. [doi: 10.1080/13691058.2011.610902] [Medline: 21972785]
38. Singh D, Fine DN, Marrazzo JM. Chlamydia trachomatis infection among women reporting sexual activity with women screened in Family Planning Clinics in the Pacific Northwest, 1997 to 2005. Am J Public Health 2011 Jul;101(7):1284-1290. [doi: 10.2105/AJPH.2009.169631] [Medline: 20724697]
39. Austin SB, Roberts AL, Corliss HL, Molnar BE. Sexual violence victimization history and sexual risk indicators in a community-based urban cohort of "mostly heterosexual" and heterosexual young women. Am J Public Health 2008 Jun;98(6):1015-1020. [doi: 10.2105/AJPH.2006.099473] [Medline: 17901440]
40. Richters J, Clayton S. The practical and symbolic purpose of dental dams in lesbian safer sex promotion. Sex Health 2010 Jun;7(2):103-106. [doi: 10.1071/sh09073] [Medline: 20648734]
41. Kodee L. Barriers To Safer Sex Practices For Lesbian And Bisexual Women. In: Barriers To Safer Sex Practices For Lesbian And Bisexual Women. Muncie, Indiana, USA: Ball State University; 2016.
42. Reis H, Gable S. Event-sampling and other methods for studying everyday experience. In: Handbook of research methods in social and personality psychology. . Cambridge University Press. Cambridge, UK: Cambridge University Press; 2000:190-222.
43. Stephenson R, Freeland R, Finneran C. Intimate partner violence and condom negotiation efficacy among gay and bisexual men in Atlanta. Sex Health 2016 Apr 28;13(4):366-372. [doi: 10.1071/SH15212] [Medline: 27120351]
44. DeGue S, Valle LA, Holt MK, Massetti GM, Matjasko JL, Tharp AT. A systematic review of primary prevention strategies for sexual violence perpetration. Aggress Violent Behav 2014;19(4):346-362 [FREE Full text] [doi: 10.1016/j.avb.2014.05.004] [Medline: 29606897]

Abbreviations

ICBI: interactive computer-based intervention

STI: sexually transmitted infection
WSW: women who have sex with women

Edited by G Eysenbach; submitted 11.09.20; peer-reviewed by X Yang, B Gupta; comments to author 01.10.20; revised version received 03.11.20; accepted 16.11.20; published 19.01.21.

Please cite as:
Zhang W, Choi EPH, Fong DYT, Wong JYH
A Moderated Mediation Analysis of Condom Negotiation and Sexual Orientation on the Relationship Between Sexual Coercion and Condom Use in Chinese Young Women: Cross-Sectional Study
JMIR Public Health Surveill 2021;7(1):e24269
URL: http://publichealth.jmir.org/2021/1/e24269/
doi:10.2196/24269
PMID:33464220

Original Paper

Toward a Working Definition of eCohort Studies in Health Research: Narrative Literature Review

Vasileios Nittas[1], MSc; Milo Alan Puhan[1], MD, Prof Dr; Viktor von Wyl[1], Prof Dr

Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Corresponding Author:
Vasileios Nittas, MSc
Epidemiology, Biostatistics and Prevention Institute
University of Zurich
Hirschengraben 84
Zurich, 8001
Switzerland
Phone: 41 44 63 44946
Email: vasileios.nittas@uzh.ch

Abstract

Background: The wide availability of internet-connected devices and new sensor technologies increasingly infuse longitudinal observational study designs and cohort studies. Simultaneously, the costly and time-consuming nature of traditional cohorts has given rise to alternative, technology-driven designs such as eCohorts, which remain inadequately described in the scientific literature.

Objective: The aim of this study was to outline and discuss what may constitute an eCohort, as well as to formulate a first working definition for health researchers based on a review of the relevant literature.

Methods: A two-staged review and synthesis process was performed comparing 10 traditional cohorts and 10 eCohorts across the six core steps in the life cycle of cohort designs.

Results: eCohorts are a novel type of technology-driven cohort study that are not physically linked to a clinical setting, follow more relaxed and not necessarily random sampling procedures, are primarily based on self-reported and digitally collected data, and systematically aim to leverage the internet and digitalization to achieve flexibility, interactivity, patient-centeredness, and scalability. This approach comes with some hurdles such as data quality, generalizability, and privacy concerns.

Conclusions: eCohorts have similarities to their traditional counterparts; however, they are sufficiently distinct to be treated as a separate type of cohort design. The novelty of eCohorts is associated with a range of strengths and weaknesses that require further exploration.

(JMIR Public Health Surveill 2021;7(1):e24588) doi:10.2196/24588

KEYWORDS

cohorts; digital epidemiology; eCohorts; eHealth

Introduction

Background

The term “cohort” is derived from Latin and was initially used to describe Roman military units; its epidemiological meaning describes a defined group of people, observed over a period of time to determine certain health outcomes [1,2]. Cohort studies provide invaluable information on the determinants of health, disease, and death [1]. Much of modern medicine’s knowledge, including the consequences of smoking and alcohol, the impact of socioeconomic factors on health outcomes, and the role of physical activity on chronic disease, is the result of large cohort studies [3-5]. Nonetheless, performing these studies remains a largely complex, expensive, and time-consuming endeavor, often embedded within resource-limited environments [6]. These limitations have led to the development of novel technology-driven approaches that aim to mitigate some of these challenges [7,8].

eCohorts, also often referred to as online or web-based cohorts, are the inevitable result of recent technological advances as well as societal developments. eCohorts harness the reach and flexibility of the internet to deal with some of the inherent complexities of traditional approaches, including the need for time-consuming and costly recruitment of large sample sizes, slow communication methods, and participant retention [8,9]. The growing acceptance of mobile health and wearables has enabled the continuous and relatively simple self-monitoring of health and risk, where patients can generate and access their personal health data supported by health apps that provide personalized content, ranging from primary prevention to therapy support and rehabilitative coaching [10-12]. In parallel, as landline telephone and postal communication use declines, social media and online communities are increasingly being used as platforms for health information sharing, in which users actively engage and contribute [8,9]. eCohorts are shaped by these developments, which promise reach, flexibility, retention, and efficiency [8,9].

Defining "eCohort"

A clear definition of what constitutes an eCohort study has not yet been established. Existing research uses a variety of terms, including "web-based," "online," "digital," and "internet" cohorts, while emphasizing different methodological aspects from web-based recruitment to online data collection and digital follow up [7,8,13,14]. We believe that the first step toward establishing a comprehensive eCohort definition is to perform a methodological comparison of traditional cohorts to eCohorts, considering all steps in the design and life cycle of these studies. The main research questions addressed were as follows: (1) Can we define eCohorts based on what we know about traditional cohorts? (2) How similar or how different are eCohorts from traditional cohorts? Thus, the primary aim of this study was to provide the first directions toward answering these questions.

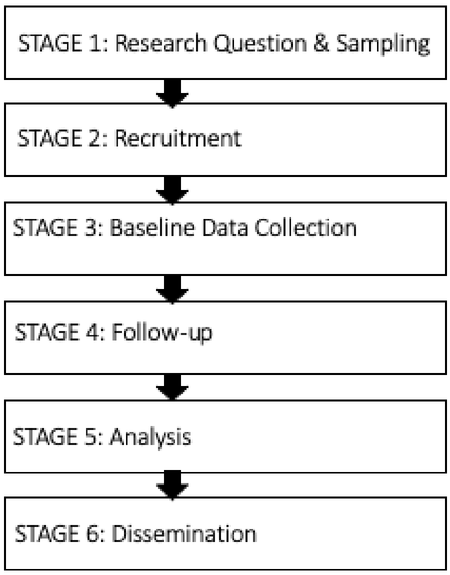
Aims

Based on a literature search and our own experiences, we aimed to outline and discuss what may constitute an eCohort and how these elements can be brought together to formulate a first working definition for health researchers. This definition should go beyond basic technical characteristics to provide a holistic description of all steps along the life cycle of an eCohort study, as well as its distinct strengths, weaknesses, risks, and challenges. As a first step to achieve this goal, we conducted a literature search to contrast the characteristics of eCohorts with those of well-defined traditional cohorts, which can facilitate a better understanding of their differences and potential similarities. We also aim to use the findings of this narrative review to inform the design of an upcoming comprehensive scoping review on eCohorts.

Methods

Our approach was based on a two-staged iterative review and synthesis process. The paper is organized as follows. Initially, we compare traditional and eCohort studies across the 6 core steps in their life cycle as outlined in Figure 1. We then continue discussing eCohorts in the context of additional characteristics such as flexibility, interactivity, usability, security, scalability, and costs. Our synthesis relied on a (1) narrative literature synthesis and (2) our own experiences with traditional cohorts and eCohorts, such as with the Women's Interagency HIV Study [15], Swiss HIV Cohort Study [16], and Swiss Multiple Sclerosis Registry [17].

Figure 1. Core stages in the life cycle of a cohort study.



We searched PubMed and Google Scholar (first 5 pages) using the terms "cohort profile," "cohort description," "cohort methods," as well as "e-cohort," "web-based cohort," "online cohort," and "digital cohort." For traditional cohorts, to avoid an unmanageable number of hits, we set the filter to observational studies in PubMed, which automatically captures only publications from 2012 onward. To include older cohorts, we used Google Scholar. This was followed by the selection of 10 traditional cohort and 10 eCohort studies. As we did not aim to provide a detailed synthesis of all existing literature, we arbitrarily set the cutoff at 10, based on narrowing down iteratively and pragmatically. Our selection was guided by the criteria outlined in Textbox 1. First, we selected studies with titles and abstracts that clearly indicated a detailed methodological account. We then proceeded iteratively to select 10 traditional and 10 eCohort studies that together provided the most rich and broad methodological information, with as few as possible overlaps, while also fulfilling the third criterion of Textbox 1.

Textbox 1. Literature review inclusion criteria.

- Criterion 1: each publication should provide sufficient descriptive information on at least three of the stages in the life cycle of a cohort study (see Figure 1).
- Criterion 2: in total, the set of included publications should provide sufficient content on all stages in the life cycle of a cohort study.
- Criterion 3: in total, the set of included publications should provide a good balance between older and newer cohorts as well as between general population and disease-specific cohorts.

The seven life-cycle stages of a cohort study guided our data extraction procedure (Figure 1). Each stage received a code and for each code we iteratively developed several subcodes. The subcodes emerged during the full-text appraisal. Coded sections were then transferred to an Excel file, and were synthesized and analyzed thematically. Multimedia Appendix 1 provides a list of our codes.

Results

Included Studies

PubMed yielded 275 hits for traditional cohorts and 46 hits for eCohorts. Google Scholar yielded an additional 150 publications for traditional cohorts and 200 publications for eCohorts. Following our inclusion criteria, we selected 10 illustrative traditional and 10 eCohort publications, which are listed in Table 1.

Table 1. Included cohort studies.

Cohort name	Year of start	Reference
Traditional cohorts		
Framingham Heart Study	1948	Tsao and Vasan [18]
The National Child Development Study	1958	Power and Elliott [19]
Nurses' Health Study	1976	Bao et al [20]
Swiss HIV Cohort Study	1988	Schoeni-Affolter et al [16]
Cohort of Norway (CONOR)	1994	Næss et al [21]
The Danish National Birth Cohort	1996	Olsen et al [22]
PIAMA Birth Cohort	1996	Wijga et al [23]
UK Millennium Cohort Study	2000	Connelly and Platt [24]
The Chronic Kidney Disease in Children Cohort Study	2005	Furth et al [25]
The lidA Cohort Study	2011	Hasselhorn et al [26]
eCohorts		
NINFEA Birth Cohort	2005	Firestone et al [8]
The Nurses and Midwives e-cohort Study	2006	Turner et al [9], Huntington et al [27]
Snart-Gravid Cohort	2007	Christensen et al [14]
ELF Cohort	2007	Firestone et al [8]
Etude NutriNet-Sante´ e-cohort	2009	Andreeva et al [13], Andreeva et al [28], Hercberg et al [29]
UK Cosmos	2009	Toledano et al [30]
SnartForaeldre	2011	Christensen et al [14]
French G-GrippeNet cohort	2014	Loubet et al [31]
Swiss MS Registry	2016	Puhan et al [7]

Stage 1: Research Question and Sampling

Every well-grounded epidemiological study is based on a well-defined research question. Traditionally, cohort studies are based on broad and multipurpose questions, dynamically changing over time based on new insights, theory, and expected future challenges [16,19,21-23,26]. New questions are commonly answered by previously collected data, as these are usually rich and highly practical. eCohort research questions do not deviate substantially from this traditional approach. Nonetheless, eCohorts are more easily aligned with the principles of citizen science (eg, involving patients in the process) and may entail technology validation components (eg, exploring the use of technology in the implementation of cohort studies) [7,9,27]. Although eCohort questions may also dynamically change over time, they are usually answered prospectively rather than using existing data, enabled by lower logistical hurdles and higher flexibility.

Identifying and recruiting a sample of adequate size and representativeness is a crucial and yet challenging step of cohort studies. Traditional cohort designs rely on well-established sampling processes, aiming for samples that are representative of the target population in terms of characteristics that potentially impact results (eg, measures of disease occurrence or specific associations). Commonly, participants are randomly selected from a predefined population group (eg, random sample of all inhabitants in a city), defined by specific events (eg, births within a certain period), and framed around specific exposures or diseases, as well as combinations of these, which may include multiple stages and stratification schemes [6,18,20,21,23,24,26]. For eCohorts, if any actual sampling process takes place, it tends to be more inclusive and less systematic than that used for traditional cohort studies. Participants are usually self-selected volunteers who are reached through various online, as well as offline, community outreach and advertisement efforts [7,9,31].

Stage 2: Recruitment

Study recruitment is the direct “engine” of any prospective cohort study. With an emphasis on sampling, potential participants are commonly preidentified and invited to participate, rendering comprehensive advertisement campaigns of lower importance. Traditional population-based cohorts primarily rely on traditional recruitment processes such as mailed invitation letters, and paper-based (or face-to-face) informed consent forms and reminders [18,20,21,23,25,26]. Participants are usually recruited and enrolled in a clinical context (eg, by physicians or nurses) [16,22,23,25]. Overall, the recruitment and study settings are very much interlinked with the clinical or community context.

By contrast, eCohorts are less attached to a clinical setting and instead rely on mixed, but mostly online and passive recruitment [9,13,27,31]. As samples are often self-selected, advertisement plays a key role. Beyond conventional methods (eg, flyers, posters), online advertising (eg, forums, social media) is becoming increasingly common, with invitations and reminders primarily sent digitally [8,9,14,27,29-31]. These approaches aim to direct potential study participants to dedicated web pages that provide all relevant study information and the option to register [8,14,29]. This is followed by the assignment of unique study identification codes and the completion of electronic consent if the legal context allows [7,9,13,27,30,31]. Self-selection, the unequal access to resources (eg, technology ownership), and the unequal distribution of skills (eg, digital literacy) may lead to selective samples of younger, better educated, high-income, health-conscious, and digitally affine participants, thereby impacting the external validity of eCohorts (ie, generalizability of study findings) [8,13,27,28,31].

Stage 3: Baseline Data Collection

The collection of baseline information (eg, exposures, current health) sets the foundation of all future comparisons. At baseline, traditional cohorts usually rely on combinations of paper-based questionnaires, environmental surveys, existing records, medical examinations, biosampling, and interviews [6,16,18-20,22,23]. These approaches are now often complemented by web-based approaches (eg, online questionnaires), aiming to reduce printing and administrative costs [18]. In contrast, web-based data collection, mostly in the form of online surveys, is the norm in eCohorts [7,9]. Paper-based survey options and medical record data are used in a complementary manner to overcome limited digital literacy and validate self-reported information [7,9,30].

Traditional cohorts rely on multiple streams of data, which, if complete, are widely considered as valid and robust, allowing for multiple control mechanisms and information triangulation [21,25,26]. In contrast, the quality, reliability, and internal validity of digitally generated data, which constitute the core of eCohorts, remain under scrutiny. Information is primarily self-reported, and may be unstructured, incomplete, or generated by devices of unclear accuracy (eg, wearables). To mitigate these limitations, eCohorts often utilize customized, automated, interactive, and responsive online surveys that minimize missing or inaccurate data [8,14,27]. Skip logics remove irrelevant questions and improve user-friendliness, consistency checks and data entry formatting reduce missing data, intermitted saving options allow for questionnaire completion over multiple sittings, and altered and feedback messages ensure that inaccurate or incomplete information is kept to a minimum [7,14,27].

Stage 4: Follow Up

Prospective cohort studies usually include longer follow-up periods, throughout which data are collected over multiple time points; this holds for both traditional and eCohort designs. Traditional designs use multiple approaches for follow up, which may be similar to the approaches used at baseline. These include regular mailed or telephone surveys, medical examinations, in-clinic biosampling, medical record linkages, data retrieval from disease and death registries, as well as personal interviews [18,19,21,22,26]. Attrition can be mitigated through record linkages (eg, school registries) that keep participant contact details updated, as well as by regularly requesting participants to update their records. Response rates are enhanced through repeated contact, combinations of multiple follow-up methods, as well as with the support of motivating health care professionals [19,20,24]. Although the time intervals between data collection points vary, they tend to be lengthy (eg, multiple years) [23,24,26].

The follow up of eCohorts is predominantly based on self-reported digital data collection (eg, online surveys, web-based diaries), which may or may not be complemented (or validated) by clinical data [7,29]. Although offline alternatives are not uncommon, they remain secondary [7]. Attrition is mitigated through (personalized) digital reminders (eg, email, SMS text messages, social media), as well as online requests to update contact details [9,27,30]. The flexibility of the internet equips eCohorts with a variety of tools to maintain response rates throughout the follow-up period. These include (1) participatory and citizen science approaches, (2) personalized and understandable feedback (eg, online data summaries), (3) tailored electronic reminders, and (4) interactive and responsive data collection methods (eg, real-time completion status, visual cues, error messages, instant feedback) [7,9,27,29-31]. Social networks (eg, Facebook, Twitter) may be utilized as contact and outreach tools, providing study updates, and allowing for direct and continuous contact with participants [7,30]. Biospecimen collection is rarer in eCohorts than in traditional cohorts, but can be accommodated remotely when needed, such as through mail-in self-kits [8]. Considering that most eCohort data are self-reported digitally, the time intervals between data collection points are flexible and more frequent than those used in traditional cohorts [31].

Stage 5: Analysis

At the data analysis stage, concerns about data quality and approaches to mitigate these seem to be a key difference between traditional and eCohorts. The analyses of traditional cohorts are largely built upon combinations of pre-existing and prospectively collected clinical data (eg, medical records, biosampling results), which are widely considered as valid and robust [21,25,26]. Subjective and self-reported data (eg, surveys) are validated through and complemented by parallel streams of clinical information (eg, health insurance claims) [21,25,26]. Although challenges and biases (eg, low response rates, loss to follow up, limited data usefulness, low sample representativeness, social desirability bias) are not uncommon, concerns inherent to data quality are minimized through the use of well-established data collection instruments, a combination of data streams, and increasingly modernized data transfer and storage practice [16,26]. Analyses usually follow lengthy data collection processes.

As outlined above (Stage 3: Baseline Data Collection), the quality and reliability of eCohort data are often scrutinized, requiring considerate data management efforts and careful adjustments to data collection instruments to mitigate a negative impact on analyses. Part of these efforts is the complementary use of clinical data (eg, medical records) to increase validity, reliability, and overall quality [7,29]. Recent analytic advances in multiple imputations of missing data have the potential to mitigate these problems in both traditional and eCohorts. Despite these challenges, digital data collection has its advantages. Data access is improved, while data collection time frames can be shorter, thereby facilitating the completion of preliminary analyses without the need for lengthy gap periods [27].

Stage 6: Dissemination

Details on the dissemination of cohort findings were scarce in the included traditional cohort publications. Dissemination seems to be focused on scientific publications, which, if added to the lengthy data collection and analysis completion periods, seems to have a rather delayed character. Nonetheless, traditional cohorts may have dedicated websites through which publications and key findings can be retrieved. A further element that could be described as integral to the dissemination strategy of traditional cohorts is the use of findings for the development and dissemination of clinical tools such as risk prediction scores [18,23,25].

By contrast, the dissemination of findings received greater emphasis in the included eCohort publications. The internet (eg, websites, newsletters, and social media) seems to be the primary tool for communicating updates and findings [7,9,27,30]. As mentioned in the previous section, the flexibility of digitalization may allow for faster data access and therefore more possibilities for preliminary analyses. In turn, this enables more frequent communication of findings and less lengthy gaps between updates [7,27]. Communication of updates, relevant news, and findings, including community outreach (eg, by webinars) and presentations to health care staff, participants, and patients, may be a part of overall strategies for maintaining participant motivation and mitigating attrition [7,30]. An important opportunity arising from eCohort (and digital health) research is that of reproducibility. Although science is undoubtedly facing a reproducibility crisis, the internet and its inherent possibilities for data availability and accessibility may eliminate replication barriers [32]. As data and technology availability increases, the practical challenges and costs of replicating research (eg, rerunning analyses) diminish. This is further facilitated by initiatives such as open science registries that aim for transparency and wide access to public research data [32]. If done correctly, the findings of eCohorts can facilitate reproducibility and open science, turning a crisis into a strength.

An overall summary comparing traditional cohorts to eCohorts is shown in Table 2.

Table 2. Comparison of eCohorts to traditional cohorts.

Characteristic	eCohort	Traditional cohort
Research question	Broad, multipurpose, interdisciplinary questions; questions may be rooted in citizen science and attached to methodological elements (eg, use of technology in epidemiological studies), change dynamically, and may be answered prospectively	Broad, multipurpose questions; questions change dynamically and are mostly answered with existing data
Sampling	Usually nonrandom sampling with self-selected volunteers	Random samples or clinic populations defined by event, exposure, or disease
Recruitment	Primarily online advertisement (eg, webpages, newsletters, forums, social media), but can be complemented by offline approaches (eg, flyers, posters)	Primarily offline advertisement (eg, flyers, posters, newspaper advertisements), but increasingly complemented by online approaches
	Recruitment usually online, through dedicated study webpages, possible at any place, any time	Recruitment usually within clinical (eg, by health care providers) or community setting and appointment-based
	Electronic consent procedures	Consent procedures usually face to face and paper-based
Baseline data collection	Primarily online and usually directly reported by participants (eg, web-based surveys). Sometimes complemented by offline data collection (eg, mailed surveys) and nonself-reported data (eg, medical record data)	Primarily offline (eg, paper-based questionnaires, data retrieval from existing records, personal interviews), and may be combined with medical examinations and biosampling
Follow up	Primarily online and usually directly reported by participants (eg, web-based surveys, personalized email, or SMS text message reminders)	Primarily offline (eg, paper-based questionnaires, data retrieval from existing records, personal interviews, medical examinations, and biosampling, mailed reminders)
	Rarely linked to medical care. Use of internet (eg, study website, social media, newsletters) for outreach and participant contact/engagement	Usually linked to medical care; personal relationship (or at least personal interactions) between participant and study coordinators
	Data quality, reliability, and internal validity may be a concern	Strong focus on data quality, reliability, and internal validity
	Data quality tradeoffs due to self-reporting; need for simpler questions, better data management, and user-friendliness	
Analysis	Usually built on self-reported data	Built upon multiple data streams, and a combination of clinical and self-reported data
	Easier data access, preliminary analyses possible in shorter time frames	Longer process, preliminary analyses more difficult in short time frames
	Analyses tend to have a stronger participant (patient) focus	Analyses tend to have a stronger clinical/biomedical focus
Dissemination	In addition to publications, through a variety of online channels (eg, websites, social media)	Primarily focused on scientific publications
	More frequent dissemination of findings	Subject to larger time gaps
	Dissemination may be a part of an overall strategy to keep participants engaged	Dissemination of findings in form of clinical tools (eg, risk scores)
	Opportunities for reproducibility and open science	

Additional Considerations for eCohorts

Flexibility and Interactivity

The digitalized nature of eCohorts allows for a certain degree of flexibility and interactivity along all stages. Internet-based recruitment and participation are not bound to a certain physical location and allow for a larger geographic reach, even if (prospective) participants are on the move [27]. Electronic data collection can be designed to be personalized and interactive, such as online questionnaires that provide real-time feedback (eg, error messages, completion status), which can be completed over multiple sittings and quickly accessed from anywhere for long periods [8,9,29]. Similarly, automated and tailored electronic reminders and follow ups such as through email, SMS, or social media allow for cheaper, faster, more frequent, and interactive communication, thereby rapidly connecting and diverting participants to study websites (eg, through click-through links) [30]. Study websites can be interactively designed, aiming to engage participants and enhance compliance [27].

Usability

Inherently, the design and functioning of eCohorts require a certain level of participant engagement. Participants have to proactively access and engage with study websites, independently self-register, and repeatedly self-report their data, often without any physical interaction with project staff or health care providers. Inevitably, to motivate and sustain this engagement, the usability of involved technology is central. Some examples include barrier-free and tailored digital interfaces, simple online recruitment and registration processes, flexible and personalized data collection approaches, as well as functioning control and guidance systems [7,8,29].

Ethics and Security

Security plays an equally important role in motivating and sustaining participation. Ethical and privacy issues are inherent to the internet, which comes with certain vulnerabilities and risks related to various stages of a cohort design, including recruitment, advertising, and data collection [7,9]. Targeted advertising (eg, through social media platforms) requires the use of data that might be considered as private (eg, demographics, education) before a person is even aware of a study's existence and long before they consent to participate [33,34]. Along similar lines, showing interest in an online advertised study (eg, by clicking on an advertisement) leaves an online trail that can be easily used by advertising companies for further profiling and targeted commercial advertising [33,34]. The tracking of our online behavior is inherent to the internet; nonetheless, this is challenging from an ethical and privacy perspective, especially in the context of sensitive health research. Further issues may arise from a certain loss of control over advertising, especially if that involves the sharing of advertisements by third parties and through various social media networks. Such uncontrolled spread might lead to losing sight of where a cohort is promoted, as well as of potential comments or questions that might have been posted across the internet [35]. Obtaining informed consent is an essential aspect of recruiting participants in a cohort. When conducted in a face-to-face manner, questions and concerns can be addressed interactively, which is lost if informed consent is obtained online and without individual contact. Filling this gap requires carefully designed online consent procedures that are transparent, understandable, and contain all elements of regular informed consent [36]. Finally, the internet makes it easier for sensitive data to be accessed without authorization, as well as hacked or replicated [33]. Although individual risk can be kept low if data are anonymized, some argue that the ease in which digital information is linked, shared, and merged renders all data potentially identifiable or traceable [37]. Therefore, adequate security features that keep risk at a minimum are inevitable [7]. Some of these features include robust password protections, high-standard information technology security, encrypted communication and data transfer, strict access controls, data deidentification, as well as the separation of personal information and unidentifiable data [7]. The emphasis on security also increases the responsibility that participants themselves have to carry, including adequate password protection, correct communication with study sites, and ensuring that devices and software are up to date.

Scalability and Costs

The internet adds a significant resource for fostering scalability and breadth [8]. Being predominantly online, eCohorts have the advantage of not being limited by physical location, having a larger sampling frame, and reaching populations who might have been otherwise difficult to reach [8,14]. Data can be collected over large geographic areas, even across borders, fostering collaborations while being managed from a single site [27]. Low-cost online recruitment and data collection techniques, facilitated by social media and their wide reach, may allow for longer recruitment and follow-up periods, thereby adding scale without prohibitively burdensome financial requirements [8,30]. Scalability is commonly associated with high costs and immense complexity, which is a major barrier of traditional cohort designs [27,30]. Nonetheless, the inherent flexibilities of eCohorts have the potential to keep costs substantially lower than those of their traditional counterparts [8,27]. Targeted online advertising can increase efficiencies, while online recruitment, invitations, and data collection can reduce labor, printing, and mailing costs [9,14,27,30]. These cost-efficiencies can nonetheless be rapidly offset. Large eCohorts require adequate resources (eg, call center, information technology personnel, digital experts, technical backups) and extensive error testing for solving arising problems as well as dealing with participant queries, all of which are costly [30]. Additional costs can also occur for the design of web platforms and data collection instruments, as well as for subsequent data security infrastructures, both of which are essential for data quality and misuse prevention, requiring maintenance throughout the full study duration [27,30].

Discussion

Principal Findings

Our comparison of traditional cohorts to eCohorts suggests a certain level of conceptual overlap. Assuming that a large proportion of eCohorts is run by people experienced with traditional cohorts, this is not a surprise. Although the stages, overall aims, and methodological basis are fairly similar, their realization differs between traditional cohorts and eCohorts across several aspects. Knowledge of traditional cohorts can be used to understand the methodological aims of eCohorts; however, the same knowledge cannot be used to derive implementation of the latter. Therefore, we consider the design of an eCohort to be a variant of its traditional counterpart.

The novelty and flexibility of eCohorts inherently bring some advantages over traditional cohort designs. The reach of the internet allows for wider, more flexible advertisement and recruitment that is not limited to a single physical setting, and may cover larger geographic regions as well as cross borders [8,14,27]. In combination with electronic data collection methods, this flexibility ultimately allows for easier scale up at potentially lower costs [8]. Digital data collection may also enable easier data availability and access (by researchers and participants), faster analyses, and more frequent dissemination of findings, all of which may foster the interest and engagement of participants [7,27]. The internet does not simply enable a wider reach, but if utilized correctly, also provides a targeted, personalized, engaging, and participant-centered process [7,8,29]. Online processes require some degree of interactivity and participant proactiveness, both of which are enhanced by digital communication methods such as personalized emails, SMS, and social media [8,9,29,30].

Inevitably, with novelty comes new risks and challenges. One of these is the generalizability of findings resulting from digitally collected data. eCohort samples often consist of volunteers that may not be representative of reference populations, posing potential external validity limitations [13], which may apply to both questions on measures of disease risks and occurrence (ie, descriptive epidemiology) and on associations (ie, inferential epidemiology). In contrast, population subgroups with lower digital literacy skills might be systematically left out, often being those that face additional sociodemographic disadvantages [38]. Concerns around privacy, security, and transparency require constant attention, especially in relation to data access, ownership, and sharing. Added to these issues, vaguely formulated and nontransparent privacy regulations create novel ethical challenges that cannot be ignored [39,40]. Finally, eCohorts require technological and analytical expertise that is carefully combined with traditional epidemiological skills and an overall motivation to keep up with the fast pace of technological innovation [40]. The promise of improved data collection and management, as well as cost-efficiency, can only be realized with carefully designed digital interfaces, effective participation incentives, and data quality assurances, which, if missing, can lead to observed moderate to low response rates and offset costs [8,9,14,27]. At the same time, given the inherent challenges in the management and analysis of eCohorts described above, some classical concepts of observational epidemiology may require adaptation to electronic contexts, including self-selection, limited potential for data management, mitigation of information biases, missing data, as well as the systematic integration and analysis of external electronic data (eg, secondary data from medical records).

Hybrid Designs

To address our aim of gaining a better understanding of eCohorts, we contrasted their design to more traditional approaches. Nonetheless, cohort studies that combine both digital and traditional elements are an increasingly common phenomenon. As indicated in our results, traditional cohorts might be enhanced by digital components such as online recruitment and data collection, while largely eCohorts may also include complementary offline elements such as physical recruitment, conventional advertising methods (eg, flyers, posters), paper-based data collection, as well as the inclusion or collection of clinical data and biospecimens. In the future, and as technology advances, hybrid cohort designs will likely be inevitable. Digitalization may support traditional cohorts to stay up to date, reach younger populations, and deal with increased mobility, while increasing efficiency and reducing costs. In turn, eCohorts may benefit from traditional approaches for reaching nondigitally native populations and increasing the validity of their data.

Working Definition

Based on our findings, a working definition of epidemiological eCohort studies could be formulated as follows. eCohorts are a novel type of cohort study, which (1) use the internet and technology as the primary delivery mode across most stages, from advertisement to recruitment, follow up, and dissemination; (2) are not entirely physically linked to a clinical setting; (3) follow more relaxed, not necessarily random, sampling procedures; (4) are primarily based on self-reported, digitally collected data, and usually have a strong patient focus; and (5) systematically aim to leverage the internet and digitalization to achieve scalability and efficiencies. We consider studies that have technology and the internet as their basis, but include hybrid elements (eg, on-site recruitment, paper-based data collection) within the scope of that definition.

Limitations

As this is relatively novel territory, we aimed for a mix of methodological control and iterative exploration, for which our findings need to be viewed in light of the following limitations. Our sample did not aim to provide a comprehensive picture of the existing literature, but rather mainly a snapshot of existing work to provide a good basis for a comparison of traditional and eCohort designs. For this purpose, we kept our searches simple and pragmatic, and our final selection of included studies was iterative. For traditional cohorts, we decided to use a more specific search, adding methodological terms to reduce the sensitivity and number of hits, whereas for eCohorts, we chose a more sensitive sample as we expected fewer hits. Of note, an extended search would also include prominent examples of digital studies, which may not strictly follow principles of cohort studies but bear close resemblance (eg, the Apple Heart Study, Project Baseline). Furthermore, very recent (unpublished) or currently ongoing eCohorts that have not been captured by our search might well emphasize additional key aspects (eg, sensor measurements in combination with telehealth consultations and patient-reported data), for which we deem a follow up of our work necessary. We aimed to counteract the potential impact of our iterative selection approach by complementing our findings with the research team's experience in traditional and eCohort studies. Our findings are primarily framed from an epidemiological perspective, which strongly impacted our focus and ultimately the definition we propose. Capturing and fully understanding all aspects of eCohorts would require further research, ideally exploring eCohorts through various angles, including an eHealth and ethical perspective. Such work would ultimately help us further refine the definition and conceptualization of eCohorts.

Conclusion

This study provides a working definition of eCohorts, facilitating a better understanding of their implementation from an epidemiological and traditional cohort perspective. Our synthesis indicates that eCohorts may have many similarities to their traditional counterparts; however, eCohorts are sufficiently distinct to be treated as a separate type of cohort design. Sampling and recruitment are more flexible, the use of the internet and technology is prominent across all cohort stages, and analyses are primarily based on self-reported and digitally collected data. The novelty of eCohorts comes with a range of strengths, weaknesses, as well as uncertainties that require further exploration. Finally, eCohorts inherently offer new insights on how the internet and emerging technology can contribute to and blend in with epidemiological and broader health research.

Acknowledgments

The salary of VN was paid by the National Research Programme “Digital Transformation” (NRP 77) of the Swiss National Science Foundation (grant number: 407740_187356).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coding.
[DOCX File , 13 KB - publichealth_v7i1e24588_app1.docx]

References

1. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg 2010 Dec;126(6):2234-2242 [FREE Full text] [doi: 10.1097/PRS.0b013e3181f44abc] [Medline: 20697313]
2. Morabia A. A History of Epidemiologic Methods and Concepts. Basel: Birkhäuser; 2004.
3. Boston University, National Institutes of Health, National Heart, Lung and Blood Institute. Framingham Heart Study. URL: https://framinghamheartstudy.org [accessed 2020-08-12]
4. Brigham and Women's Hospital, Harvard Medical School, Harvard T.H. Chan School of Public Health. Nurses' Health Study. URL: https://www.nurseshealthstudy.org/ [accessed 2020-07-07]
5. UCL Centre for Longitudinal Studies. 1958 National Child Development Study. URL: https://cls.ucl.ac.uk/cls-studies/1958-national-child-development-study/ [accessed 2020-08-12]
6. Ward H, Toledano M, Shaddick G, Davies B, Elliott P. Oxford Handbook of Epidemiology for Clinicians. Oxford: Oxford University Press; 2012.
7. Puhan MA, Steinemann N, Kamm CP, Müller S, Kuhle J, Kurmann R, Swiss Multiple Sclerosis Registry Smsr. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. Swiss Med Wkly 2018;148:w14623 [FREE Full text] [doi: 10.4414/smw.2018.14623] [Medline: 29767828]
8. Firestone R, Cheng S, Pearce N, Douwes J, Merletti F, Pizzi C, et al. Internet-Based Birth-Cohort Studies: Is This the Future for Epidemiology? JMIR Res Protoc 2015 Jun 12;4(2):e71 [FREE Full text] [doi: 10.2196/resprot.3873] [Medline: 26071071]
9. Turner C, Bain C, Schluter PJ, Yorkston E, Bogossian F, McClure R, NursesMidwives e-cohort Group. Cohort Profile: The Nurses and Midwives e-Cohort Study--a novel electronic longitudinal study. Int J Epidemiol 2009 Feb 17;38(1):53-60. [doi: 10.1093/ije/dym294] [Medline: 18202083]
10. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
11. Kugler C, Gottlieb J, Dierich M, Haverich A, Strueber M, Welte T, et al. Significance of patient self-monitoring for long-term outcomes after lung transplantation. Clin Transplant 2010;24(5):709-716. [doi: 10.1111/j.1399-0012.2009.01197.x] [Medline: 20047613]
12. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc 2013;53(2):172-181 [FREE Full text] [doi: 10.1331/JAPhA.2013.12202] [Medline: 23571625]
13. Andreeva VA, Deschamps V, Salanave B, Castetbon K, Verdot C, Kesse-Guyot E, et al. Comparison of Dietary Intakes Between a Large Online Cohort Study (Etude NutriNet-Santé) and a Nationally Representative Cross-Sectional Study (Etude Nationale Nutrition Santé) in France: Addressing the Issue of Generalizability in E-Epidemiology. Am J Epidemiol 2016 Nov 01;184(9):660-669 [FREE Full text] [doi: 10.1093/aje/kww016] [Medline: 27744386]
14. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
15. Bacon MC, von Wyl V, Alden C, Sharp G, Robison E, Hessol N, et al. The Women's Interagency HIV Study: an observational cohort brings clinical sciences to the bench. Clin Diagn Lab Immunol 2005 Sep;12(9):1013-1019 [FREE Full text] [doi: 10.1128/CDLI.12.9.1013-1019.2005] [Medline: 16148165]
16. Swiss HIV Cohort Study, Schoeni-Affolter F, Ledergerber B, Rickenbach M, Rudin C, Günthard HF, et al. Cohort profile: the Swiss HIV Cohort study. Int J Epidemiol 2010 Oct 30;39(5):1179-1189. [doi: 10.1093/ije/dyp321] [Medline: 19948780]
17. Steinemann N, Kuhle J, Calabrese P, Kesselring J, Disanto G, Merkler D, Swiss Multiple Sclerosis Registry. The Swiss Multiple Sclerosis Registry (SMSR): study protocol of a participatory, nationwide registry to promote epidemiological and patient-centered MS research. BMC Neurol 2018 Aug 13;18(1):111 [FREE Full text] [doi: 10.1186/s12883-018-1118-0] [Medline: 30103695]

18. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015 Dec 23;44(6):1800-1813 [FREE Full text] [doi: 10.1093/ije/dyv337] [Medline: 26705418]
19. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol 2006 Feb;35(1):34-41. [doi: 10.1093/ije/dyi183] [Medline: 16155052]
20. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, Methods, and Evolution of the Three Nurses' Health Studies. Am J Public Health 2016 Sep;106(9):1573-1581. [doi: 10.2105/AJPH.2016.303338] [Medline: 27459450]
21. Naess O, Søgaard AJ, Arnesen E, Beckstrøm AC, Bjertness E, Engeland A, et al. Cohort profile: cohort of Norway (CONOR). Int J Epidemiol 2008 Jun 04;37(3):481-485 [FREE Full text] [doi: 10.1093/ije/dym217] [Medline: 17984119]
22. Olsen J, Melbye M, Olsen SF, Sørensen TI, Aaby P, Andersen AM, et al. The Danish National Birth Cohort--its background, structure and aim. Scand J Public Health 2001 Dec;29(4):300-307. [doi: 10.1177/14034948010290040201] [Medline: 11775787]
23. Wijga AH, Kerkhof M, Gehring U, de Jongste JC, Postma DS, Aalberse RC, et al. Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. Int J Epidemiol 2014 Apr 11;43(2):527-535. [doi: 10.1093/ije/dys231] [Medline: 23315435]
24. Connelly R, Platt L. Cohort profile: UK Millennium Cohort Study (MCS). Int J Epidemiol 2014 Dec 17;43(6):1719-1725. [doi: 10.1093/ije/dyu001] [Medline: 24550246]
25. Furth SL, Cole SR, Moxey-Mims M, Kaskel F, Mak R, Schwartz G, et al. Design and methods of the Chronic Kidney Disease in Children (CKiD) prospective cohort study. Clin J Am Soc Nephrol 2006 Sep;1(5):1006-1015 [FREE Full text] [doi: 10.2215/CJN.01941205] [Medline: 17699320]
26. Hasselhorn HM, Peter R, Rauch A, Schröder H, Swart E, Bender S, et al. Cohort profile: the lidA Cohort Study-a German Cohort Study on Work, Age, Health and Work Participation. Int J Epidemiol 2014 Dec 11;43(6):1736-1749 [FREE Full text] [doi: 10.1093/ije/dyu021] [Medline: 24618186]
27. Huntington A, Gilmour J, Schluter P, Tuckett A, Bogossian F, Turner C. The Internet as a research site: establishment of a web-based longitudinal study of the nursing and midwifery workforce in three countries. J Adv Nurs 2009 Jun;65(6):1309-1317. [doi: 10.1111/j.1365-2648.2009.04995.x] [Medline: 19445011]
28. Andreeva VA, Salanave B, Castetbon K, Deschamps V, Vernay M, Kesse-Guyot E, et al. Comparison of the sociodemographic characteristics of the large NutriNet-Santé e-cohort with French Census data: the issue of volunteer bias revisited. J Epidemiol Community Health 2015 Sep 01;69(9):893-898. [doi: 10.1136/jech-2014-205263] [Medline: 25832451]
29. Hercberg S, Castetbon K, Czernichow S, Malon A, Mejean C, Kesse E, et al. The Nutrinet-Santé Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. BMC Public Health 2010 May 11;10(1):242 [FREE Full text] [doi: 10.1186/1471-2458-10-242] [Medline: 20459807]
30. Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to Establish and Follow up a Large Prospective Cohort Study in the 21st Century--Lessons from UK COSMOS. PLoS One 2015 Jul 6;10(7):e0131521 [FREE Full text] [doi: 10.1371/journal.pone.0131521] [Medline: 26147611]
31. Loubet P, Guerrisi C, Turbelin C, Blondel B, Launay O, Bardou M, et al. First nationwide web-based surveillance system for influenza-like illness in pregnant women: participation and representativeness of the French G-GrippeNet cohort. BMC Public Health 2016 Mar 11;16(1):253 [FREE Full text] [doi: 10.1186/s12889-016-2899-y] [Medline: 26969654]
32. Stupple A, Singerman D, Celi LA. The reproducibility crisis in the age of digital medicine. NPJ Digit Med 2019;2:2. [doi: 10.1038/s41746-019-0079-z] [Medline: 31304352]
33. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
34. Curtis BL. Social networking and online recruiting for HIV research: ethical challenges. J Empir Res Hum Res Ethics 2014 Feb;9(1):58-70 [FREE Full text] [doi: 10.1525/jer.2014.9.1.58] [Medline: 24572084]
35. Fileborn B. Participant recruitment in an online era: A reflection on ethics and identity. Res Ethics 2015 Sep 23;12(2):97-115. [doi: 10.1177/1747016115604150]
36. Flicker S, Haans D, Skinner H. Ethical dilemmas in research on internet communities. Qual Health Res 2004 Jan;14(1):124-134. [doi: 10.1177/1049732303259842] [Medline: 14725180]
37. Mittelstadt B, Benzler J, Engelmann L, Prainsack B, Vayena E. Is there a duty to participate in digital epidemiology? Life Sci Soc Policy 2018 May 09;14(1):9 [FREE Full text] [doi: 10.1186/s40504-018-0074-1] [Medline: 29744694]
38. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
39. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb 9;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
40. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012 Jul 26;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by G Fagherazzi, V Ameli, E Nelson; comments to author 21.10.20; revised version received 06.11.20; accepted 09.12.20; published 21.01.21.

Please cite as:
Nittas V, Puhan MA, von Wyl V
Toward a Working Definition of eCohort Studies in Health Research: Narrative Literature Review
JMIR Public Health Surveill 2021;7(1):e24588
URL: http://publichealth.jmir.org/2021/1/e24588/
doi:10.2196/24588
PMID:33475521

Original Paper

Assessment of Strategies and Epidemiological Characteristics of Tuberculosis in Henan Province, China: Observational Study

Hui Jiang[1,2], PhD; Guolong Zhang[3], MD; Jinfeng Yin[1,2], MPhil; Dongyang Zhao[3], MD; Fangchao Liu[1,2], PhD; Yuxia Yao[3], MPH; Chao Cai[4], PhD; Jiying Xu[3], MPH; Xinwei Li[5], BSc; Wangli Xu[5], PhD; Weimin Li[1,6,7], PhD

[1]Beijing Chest Hospital, Capital Medical University, Beijing, China

[2]Beijing Tuberculosis and Thoracic Tumor Research Institute, Beijing, China

[3]Institute of Tuberculosis Control and Prevention, Henan Center for Disease Control and Prevention, Henan, China

[4]Beijing Youan Hospital, Capital Medical University, Beijing, China

[5]School of Statistics, Renmin University of China, Beijing, China

[6]Beijing Municipal Key Laboratory of Clinical Epidemiology, School of Public Health, Capital Medical University, Beijing, China

[7]National Tuberculosis Clinical Lab of China, Beijing Tuberculosis and Thoracic Tumour Research Institute, Beijing Key Laboratory in Drug Resistance Tuberculosis Research, Beijing, China

Corresponding Author:
Weimin Li, PhD
Beijing Chest Hospital
Capital Medical University
District 1, No. 9 Beiguan Street, Tongzhou District
Beijing, 101149
China
Phone: 86 1089509359
Email: lwm_18@aliyun.com

Abstract

Background: In 2005, China established an internet-based Tuberculosis Information Management System (TBIMS) to monitor changes in tuberculosis (TB). Many scholars have conducted epidemiological research using TBIMS; however, few studies assessing control strategies have been performed based on this platform data. Henan province is a high TB incidence area in China where, in addition to following the nationwide TB strategies, a series of local intervention combinations have been implemented.

Objective: Our study aims to evaluate the impact of nationwide TB intervention combinations on epidemiological changes and determine whether Henan province can achieve the World Health Organization's (WHO) goal of reducing TB incidence by 50% and TB mortality by 75% by the year 2025.

Methods: We used descriptive statistical methods to show the spatial and temporal distribution of pulmonary tuberculosis (PTB) reported to the TBIMS database from 2005 to 2018, and logistic regression analysis was performed to identify the risk factors of bacteriological-positive TB. The dynamic compartmental model and Bayesian melding approach was adopted to estimate the burden of TB under the impact of different TB control policies.

Results: In total, 976,526 PTB cases were notified to the TBIMS in Henan in a period of 14 years. Although the overall incidence of PTB declined from $91.4/10^5$ to $58.5/10^5$, and the overall incidence of bacteriological-positive PTB declined from $44.5/10^5$ to $14.7/10^5$, the WHO's 2025 goal could not be met. The distribution of high incidence and poverty-stricken counties were basically overlapped. Men, farmers and herdsmen (in rural areas), and subjects aged ≥60 years were more likely to develop bacteriological-positive PTB. The increasing treatment success for drug-susceptible tuberculosis and multidrug-resistant tuberculosis has not provided the desired reduction in incidence and mortality.

Conclusions: To achieve the targeted goal, while improving the cure rate of TB, new active (rather than passive) detection and intervention strategies should be formulated based on epidemiological characteristics in Henan province.

(JMIR Public Health Surveill 2021;7(1):e24830) doi:10.2196/24830

KEYWORDS

notified pulmonary tuberculosis; Tuberculosis Information Management System; epidemiological characteristics; dynamic compartmental model; policy evaluation;

Introduction

Implementation of the directly observed treatment short-course (DOTS) chemotherapy strategy led to a 65% reduction in the prevalence of smear-positive tuberculosis (TB) in China between 1990 and 2010 [1]. Nevertheless, China still had the second-highest number of TB infections globally [2]. In 2005, China established an internet-based Tuberculosis Information Management System (TBIMS) as the national TB surveillance system, and all Chinese TB health facilities have been required to report diagnosed pulmonary tuberculosis (PTB) cases directly into the TBIMS [3]. The TBIMS platform allows for real-time monitoring of TB diagnosis, treatment, and outcomes in China, especially for PTB. Recently, many scholars have conducted epidemiological investigations using TBIMS; however, few studies have assessed control strategies based on this platform data.

The World Health Organization's (WHO) 2025 goal was to reduce TB incidence by 50% and TB mortality by 75% [4] from 2015 to 2025. Recently, research studies have shown that China is unlikely to meet the global TB-related targets by intensifying its current strategies, which were passive measures taken only for TB symptomatic persons [5,6]. Henan province is a high incidence area, accounting for 10% of TB infections in China. In addition to following the nationwide TB strategies, Henan province has also implemented a series of nonfragmented local intervention combinations, including an annual investment from the local authorities of 1.42 million US dollars to purchase antituberculosis drugs and diagnostic reagents since 2010, the provision of free screening for latent TB infection to enrolled first-year students since 2017, and a payment plan targeting single-disease TB treatment since 2018. However, whether these measures would change the epidemiological characteristics and help Henan province achieve the WHO's goal for 2025 has not been determined.

In this study, based on the TBIMS database, we collected details of PTB as well as demographic, epidemiological, geographical, and laboratory information and related policies in Henan province for 2005-2018. We aim to explore whether the epidemiological characteristics of TB have changed in the past 14 years under the guidance of the national and provincial policies, and to use a dynamic compartmental model to evaluate various TB control and prevention policies in order to determine whether Henan province can achieve the WHO's 2025 goal.

Methods

Data Collection

Since January 1, 2005, PTB cases are reported to the TBIMS, the national TB surveillance system, within 24 hours of diagnosis [3]. From the TBIMS database, we collected information from PTB cases on basic demographic information (sex, birth date, home address, occupation, and treatment of classification), time of illness onset and diagnosis, and laboratory outcomes (sputum smear results and sputum culture results). The percentage values of multidrug-resistant tuberculosis (MDR-TB) infections in all cases, new patients, and re-treated patients were obtained by analysis of multidrug resistance of tuberculosis. The mortality was obtained from the national disease surveillance system's tuberculosis death analysis report and the Ministry of Health's prevalence survey in 2010.

Data Analysis

Descriptive statistical methods were adopted to analyze continuous variables and categorical variables. The annual incidence rates of PTB (per 100,000 people) and bacteriological-positive PTB were calculated. We used the Arc Map's (version 10.2; ESRI Inc) ring map to show the spatiotemporal patterns of PTB incidence. In order to illustrate the seasonal patterns of PTB in different regions, we created heatmaps for the proportion of PTB cases identified during each month of the year. A hierarchical clustering method was used to identify similar regions based on the overall and bacteriological-positive PTB incidence rates, which were compared with the distribution of poverty-stricken regions in Henan province. Univariable and multivariable logistic regression models were applied in order to investigate the factors associated with bacteriological-positive PTB, and unadjusted odds ratios (OR) and adjusted odds ratios (AOR) were also estimated.

We used a dynamic compartmental model [5,7] to predict the incidence and mortality of TB epidemics, and the main parameters included natural history parameters, mortality rate or birth rate, and program parameters such as patient visit rate and long-term cure rate. We found the present background parameters values through a review of the literature [5,6], including the long-term cure rate for new cases and re-treatment cases in different medical settings. Based on experts' opinions and a local epidemiological survey from Henan Center for Disease Control and Prevention, we selected and set the local background parameters and further simulated 4 main scenarios for the future. We assumed that these scenarios would be implemented in 2018 and estimated their effect by 2025 using the model. In scenario 1 (the current status in Henan), we estimated the ranges of the decline in incidence and mortality if the current strategies are maintained until 2025. In scenario 2, we estimated the change of incidence and mortality by year if the treatment success rate for drug-susceptible tuberculosis (DS-TB) increased to 92% for new treatment and 90% for re-treatment. In scenario 3, we considered using better diagnosis technologies and increasing treatment success for MDR-TB to show the impact of the measures, and the long-term cure rate with a second-line drug for MDR-TB was 82%. In scenario 4, a combination of all conducted measures was represented; that is, scenarios 1, 2, and 3 were delivered simultaneously. A detailed description of the statistical analyses is provided in Multimedia Appendix 1.

In addition, the data of TB burden in Henan province were used to calibrate the model by adopting the Bayesian melding approach [8]. Then, the impact of intervention strategies on the epidemiology of TB was predicted using the fitted model. We predicted the incidence, mortality, and multidrug-resistant percentage for all cases and new cases for each scenario. The posterior simulations produced 95% credible intervals. Furthermore, the post-2015 global tuberculosis targets were compared with the variation in incidence and mortality that we calculated.

Ethics Statement

The ethics review committee of the Henan Tuberculosis Control Institute provided approval for this study. Additionally, patients' information was anonymized to ensure privacy.

Results

Demographic Characteristics

A total of 976,526 PTB cases were reported to TBIMS in the 2005–2018 period. Of these, 381,598 (39.08%) were cases of bacteriological-positive tuberculosis. In all PTB cases, the overall male-to-female ratio was 2.38:1; however, this pattern was not uniform across the years (Figure 1A), and the gender difference in the 0-14 years age group was the lowest (Figure 1B). The median age was 48 (IQR 28–63) years, and the median age for bacteriological-positive cases (51 years, IQR 31-65) was higher than for negative cases (47 years, IQR 28-63).

Moreover, 95.0% of patients were new cases, and 98.2% had received antituberculosis therapy (Table 1).

Figure 1. Number of pulmonary tuberculosis (PTB) cases by age and sex in Henan province, China, from 2005 to 2018. A: Number of PTB cases by sex and year. B: Number of PTB cases by age group, sex, and year.

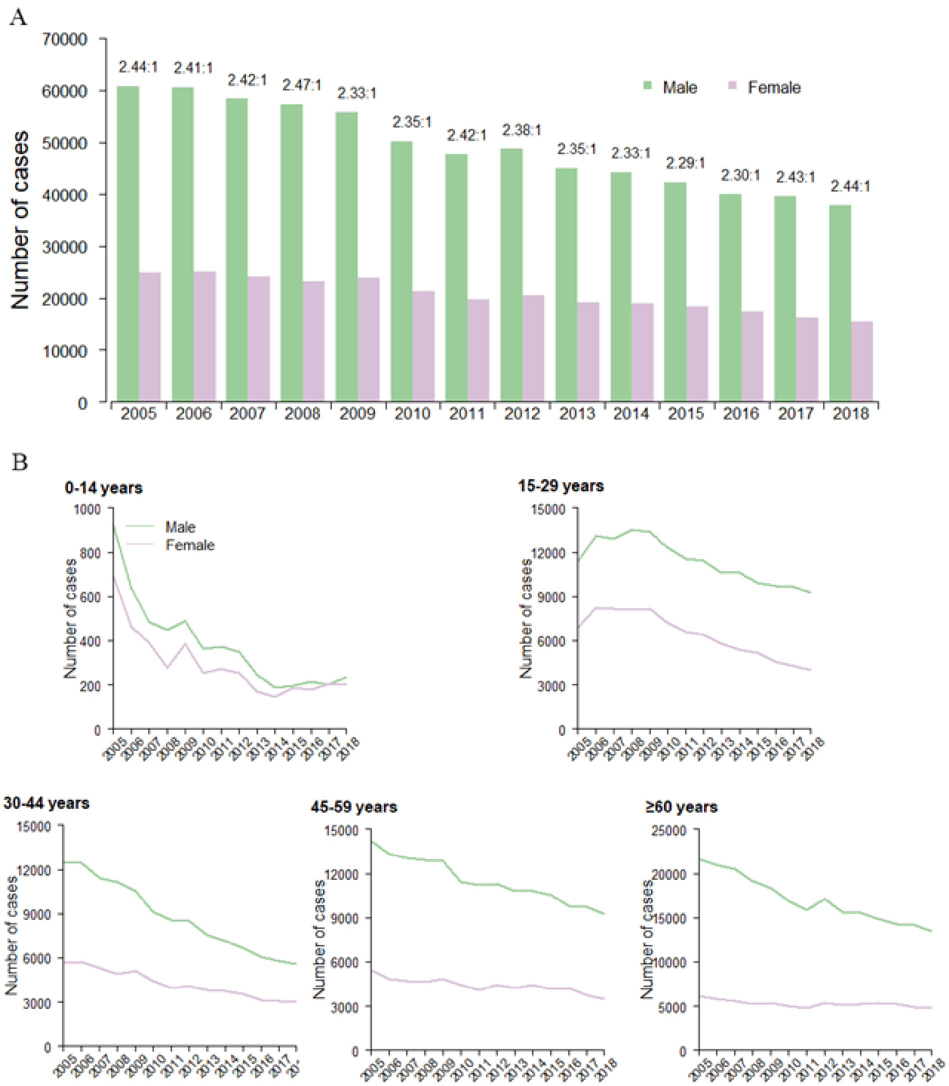


Table 1. Characteristics of pulmonary tuberculosis (PTB) cases in Henan province, China, from 2005 to 2018.

Characteristic	Total PTB cases (n=976,526)	Laboratory results	
		Bacteriological-positive (n=381,598)	Bacteriological-negative (n=526,888)
Gender			
Male, n (%)	687,996 (70.45)	287,143 (74.54)	358,916 (68.12)
Female, n (%)	288,530 (29.55)	98,055 (25.46)	167,972 (31.88)
Age groups in years, n (%)			
0-14	9,434 (0.97)	1,086 (0.28)	4,668 (0.89)
15-29	248,376 (25.43)	87,272 (22.66)	137,771 (26.15)
30-44	182,613 (18.70)	69,732 (18.10)	99,220 (18.83)
45-59	223,073 (22.84)	90,222 (23.42)	120,923 (22.95)
≥60	312,620 (32.01)	136,734 (35.50)	164,116 (31.15)
Unknown	410 (0.04)	152 (0.04)	190 (0.04)
Occupation, n (%)			
Nursery children	307 (0.03)	8 (0.00)	79 (0.01)
Scattered children	729 (0.07)	36 (0.01)	246 (0.05)
Students	47,760 (4.89)	11,171 (2.90)	29,690 (5.63)
Workers	26,229 (2.69)	9,924 (2.58)	14,135 (2.68)
Farmers and herdsmen	795,893 (81.50)	325,259 (84.44)	425,466 (80.75)
Commercial service stratum	4704 (0.48)	1607 (0.42)	2650 (0.50)
Others	98,414 (10.08)	36,181 (9.39)	53,395 (10.13)
Unknown	2490 (0.25)	1012 (0.26)	1227 (0.23)
Residence, n (%)			
Rural	786,210 (80.51)	317,547 (82.44)	419,554 (79.63)
Urban	180,375 (18.47)	63,582 (16.51)	102,130 (19.38)
Unknown	9941 (1.02)	4069 (1.06)	5204 (0.99)
Treatment of classification, n (%)			
New case	927,250 (94.95)	358,170 (92.98)	505,423 (95.93)
Re-treated case	49,276 (5.05)	27,028 (7.02)	21,465 (4.07)
Anti-tuberculosis therapy			
Yes	958,965 (98.20)	379,016 (98.44)	517,660 (98.31)
No	16,978 (1.74)	6017 (1.56)	8917 (1.69)
Unknown	583 (0.06)	165 (0.04)	311 (0.06)
Time from illness onset to first hospital visit, in days, n (%)			
0-30	646,116 (66.16)	235,774 (61.21)	365,316 (69.33)
31-60	119,264 (12.21)	50,124 (13.01)	62,298 (11.82)
61-90	41,171 (4.22)	19,310 (5.01)	19,938 (3.78)
>90	53,414 (5.47)	26,919 (6.99)	24,511 (4.65)
Unknown	116,561 (11.94)	53,071 (13.78)	54,825 (10.41)
Time from onset to confirmation, in days, n (%)			
0-30	564,081 (57.76)	205,505 (53.35)	318,226 (60.40)
31-60	192,248 (19.69)	76,306 (19.81)	103,770 (19.69)
61-90	68,368 (7.00)	30,570 (7.94)	34,263 (6.50)

Characteristic	Total PTB cases (n=976,526)	Laboratory results	
		Bacteriological-positive (n=381,598)	Bacteriological-negative (n=526,888)
>90	96,450 (9.88)	48,874 (12.69)	44,032 (8.36)
Unknown	55,379 (5.67)	23,943 (6.22)	26,597 (5.05)
Time from onset to treatment end, in months, n (%)			
0-6	47,621 (4.88)	19,284 (5.01)	24,999 (4.74)
6-12	715,251 (73.24)	277,827 (72.13)	404,604 (76.79)
>12	60,994 (6.25)	22,312 (5.79)	22,390 (4.25)
Unknown	152,660 (15.63)	65,775 (17.08)	74,895 (14.21)

Incidence and Seasonality

The annual average PTB incidence was $75.3/10^5$, and the overall incidence trend declined from $91.4/10^5$ in 2005 to $58.5/10^5$ in 2017. The annual average incidence for patients with bacteriological-positive PTB was $30.1/10^5$, and it decreased from $44.5/10^5$ in 2005 to $14.7/10^5$ in 2017 (Figure 2A).

PTB incidence showed broad age-specific variations, and patients aged ≥60 years and 15-29 years ranked first and second, respectively. In addition, the incidence in rural regions was higher than in urban areas (P<.05; Table 2).

As can be observed from the PTB's geographical distribution across Henan in the 2005-2018 period, the high incidence areas of PTB in Henan province remained unchanged (Figure 2B). In addition, from 2005 to 2010, the overall incidence of bacteriological-positive cases remained high, except in Zhengzhou city (Figure 2C). The incidence rates of PTB, cases of bacteriological-positive tuberculosis, and poverty-stricken counties were basically the same, especially in Nanyang city, Xinyang city, Zhoukou city, and Zhumadian city (Figure 3A-C).

Figure 2. Spatiotemporal distribution of pulmonary tuberculosis (PTB) cases by city and epidemic curve in Henan province, China, from 2005 to 2018. A: PTB epidemic curve on the number of cases reported weekly. B: Spatiotemporal distribution of all PTB cases. C: Spatiotemporal distribution of bacteriological-positive PTB cases. D: Time series for weekly reported PTB cases (standardized by the number of annual cases). E: Cluster analysis for seasonal distribution of PTB cases.

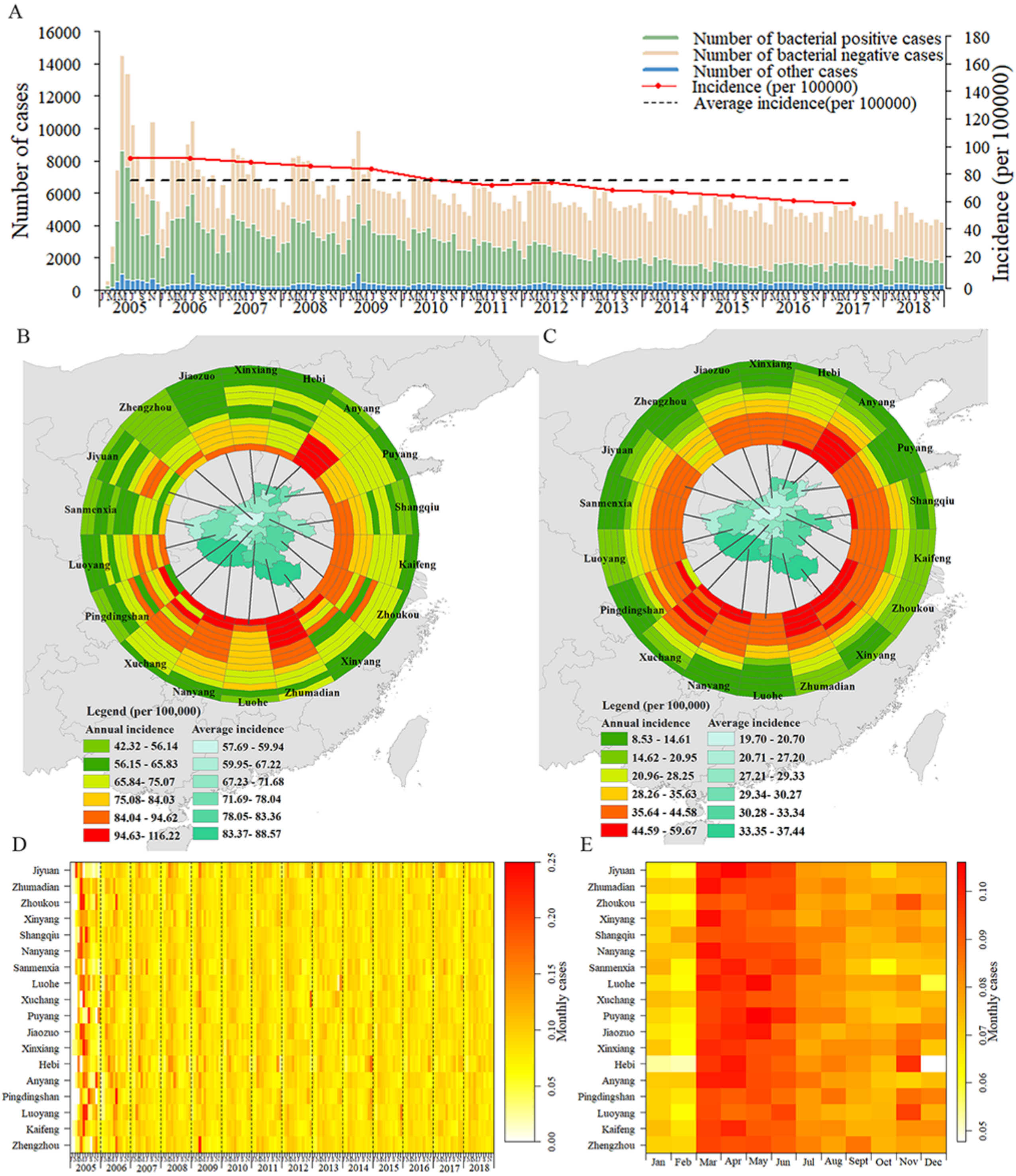
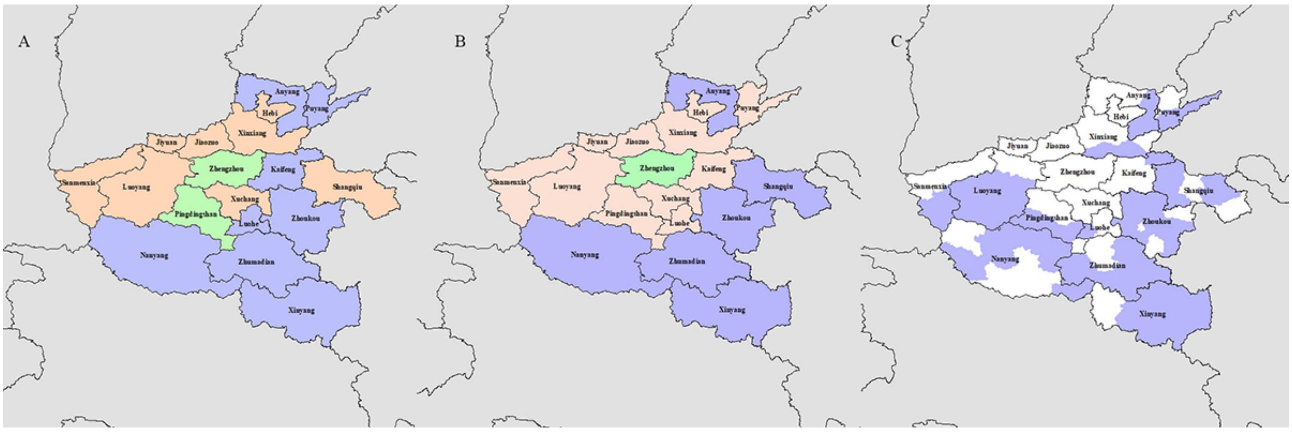


Table 2. Incidence of pulmonary tuberculosis (PTB) cases by age group and residence in Henan province, China, from 2005 to 2017.

Characteristic	Year (2005-2017)												
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Residence													
Urban	46.4	48.1	42.9	39.6	40.8	35.1	31.0	31.7	29.8	29.8	27.1	25.6	25.8
Rural	108.5	110.6	110.4	109.6	108.2	100.5	98.1	103.1	96.9	96.9	95.7	92.3	90.7
Age group, in years													
0-14	8.1	5.63	4.6	3.8	4.8	3.1	3.3	3.2	2.1	1.6	1.9	1.9	2.0
15-29	—[a]	104.2	100.9	103.9	103.5	84.5	82.9	83.2	81.7	77.4	74.9	71.9	74.6
30-44	—	72.2	68.5	68.8	69.1	62.5	58.4	61.9	59.1	54.8	53.7	48.6	48.7
45-59	—	103.7	96.0	86.6	87.1	90.8	86.8	82.7	73.3	76.0	71.5	67.2	60.1
≥60	—	237.5	243.0	218.2	186.0	182.0	156.2	158.2	143.6	147.7	135.1	127.3	121.8

[a] — not available.

Figure 3. Classification of pulmonary tuberculosis (PTB) epidemiological regions through cluster analysis in Henan province, China, from 2005 to 2018. A: Classification of all PTB epidemiological regions. B: Classification of bacteriological-positive PTB epidemiological regions. C: Distribution of poverty-stricken counties.

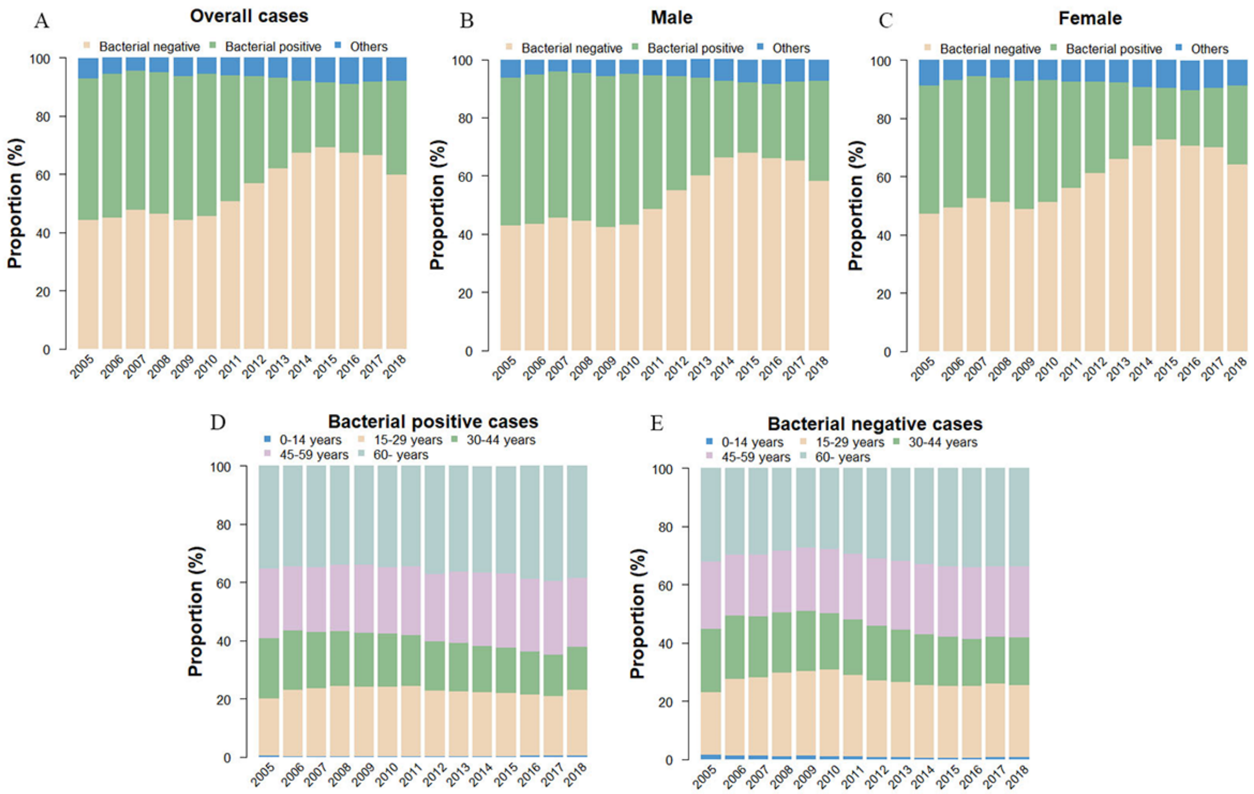


On the provincial scale, it can be observed that people developed PTB throughout the year, while at the city-level, the incidence did not reach a peak at the same time across Henan; nevertheless, the annual seasonal peaks were noticeable. A seasonal pattern was observed in March, April, May, and June, and the peak season occurred in March and April. Cases decreased during early June and after mid-October, and the PTB cases increased again, with transmission reaching a small peak, in November. The 2 months with the lowest PTB incidence were January and February (Figure 2D-E).

Changing Trends of Bacteriological-Positive Results

Bacteriological results showed that the proportion of bacteriological-positive cases decreased from 48.79% in 2005 to 23.64% in 2016. However, the proportion of bacteriological-positive infection rose in 2017 and 2018 to 25.20% and 32.27%, respectively (Figure 4A). Further, the proportion of males with bacteriological-positive tuberculosis was higher than that of females (Figure 4B-C). In addition, the proportion of patients with bacteriological-positive infection in the ≥60 years age group was the highest, and the proportion increased continuously from 35.42% in 2005 to 38.42% in 2018 (Figure 4D). The proportion of patients with bacteriological-negative infection across different age groups was consistent with that of patients with bacteriological-positive infection (Figure 4E).

Figure 4. Proportion of laboratory-tested pulmonary tuberculosis (PTB) cases by sex and bacteriological type in Henan province, China, from 2005 to 2018. A: based on overall cases; B: based on male cases; C: based on female cases; D: based on bacteriological-positive cases; E: based on bacteriological-negative cases.



Risk Factors for Bacteriological-Positive Infection

Male sex, residence in rural areas, and re-treated cases were factors associated with the development of bacteriological-positive PTB, and the AORs for these groups were 1.32, 1.17, and 1.70, respectively. In addition, a trend of substantial increase in PTB risk with bacteriological-positive infection was observed with increasing age; the AORs were 1.91, 1.92, 1.97, and 2.17 for the 15-29, 30-44, 45-59, and ≥60 years age groups, respectively. Moreover, workers, farmers and herders, and the commercial service stratum were more susceptible to developing bacteriological-positive PTB than nursery children and students; the AORs for these groups were 3.78 (95% CI 1.93-8.54), 3.69 (95% CI 1.88-8.34), and 3.44 (95% CI 1.75-7.79), respectively (Table 3).

Table 3. Risk factors for bacteriological-positive infection cases in Henan province, 2005-2018.

Characteristic	Laboratory results, n (%)		Odds ratio (95% CI)	Adjusted odds ratio (95% CI)
	Bacteriological-positive	Bacteriological-negative		
Sex				
Male	287,143 (74.54)	358,916 (68.12)	1.37 (1.35-1.38)	1.32 (1.31-1.34)
Female	98,055 (25.46)	167,972 (31.88)	1	1
Age groups, in years				
0-14	1086 (0.28)	4668 (0.89)	1	1
15-29	87,272 (22.67)	137,771 (26.16)	2.72 (2.54-2.91)	1.91 (1.79-2.06)
30-44	69,732 (18.11)	99,220 (18.84)	3.02 (2.82-3.23)	1.92 (1.79-2.06)
45-59	90,222 (23.43)	120,923 (22.96)	3.20 (3.00-3.42)	1.97 (1.84-2.12)
≥60	136,734 (35.51)	164,116 (31.16)	3.58 (3.35-3.82)	2.17 (2.02-2.33)
Occupation				
Nursery children	8 (0.00)	79 (0.02)	1	1
Scattered children	36 (0.01)	246 (0.06)	1.44 (0.67-3.46)	1.33 (0.61-3.22)
Students	11,171 (2.91)	29,690 (5.65)	3.71 (1.91-8.35)	2.22 (1.14-5.02)
Workers	9924 (2.58)	14,135 (2.69)	6.93 (3.56-15.58)	3.78 (1.93-8.54)
Farmers and herdsmen	325,259 (84.66)	425,466 (80.94)	7.54 (3.88-16.96)	3.69 (1.88-8.34)
Commercial service stratum	1607 (0.42)	2650 (0.50)	5.98 (3.07-13.48)	3.44 (1.75-7.79)
Others	36,181 (9.42)	53,395 (10.16)	6.69 (3.44-15.04)	3.61 (1.84-8.15)
Residence				
Urban	63,582 (16.68)	102,130 (19.58)	1	1
Rural	317,547 (83.32)	419,554 (80.42)	1.21 (1.20-1.22)	1.17 (1.15-1.18)
Treatment history				
New case	358,170 (92.98)	505,423 (95.93)	1	1
Retreated case	27,028 (7.02)	21,465 (4.07)	1.77 (1.74-1.80)	1.70 (1.66-1.73)

Impact of Interventions on TB Control Policies

In scenario 1 (the current status in Henan), if the strategies are kept unchanged, the incidence and mortality are expected to gradually decline by 22.6% (95% CI 21.7%-23.6%) and 27.9% (95% CI 27.0%-28.3%), respectively, from 2015 to 2025 (Figure 5A-B, Table 4).

Figure 5. Expected impact of different scenarios of tuberculosis (TB) control on the prevalence of TB in Henan province, China, 2015-2025. A: Expected impact of different scenarios of TB control on the annual TB incidence. B: Expected impact of different scenarios of TB control on the annual TB mortality. C: Expected impact of different scenarios of TB control on the percentage of multidrug-resistant tuberculosis (MDR-TB) in the general population. D: Expected impact of different scenarios of TB control on the percentage of MDR-TB in all TB cases. E: Expected impact of different scenarios of TB control on the percentage of MDR-TB in new TB cases.

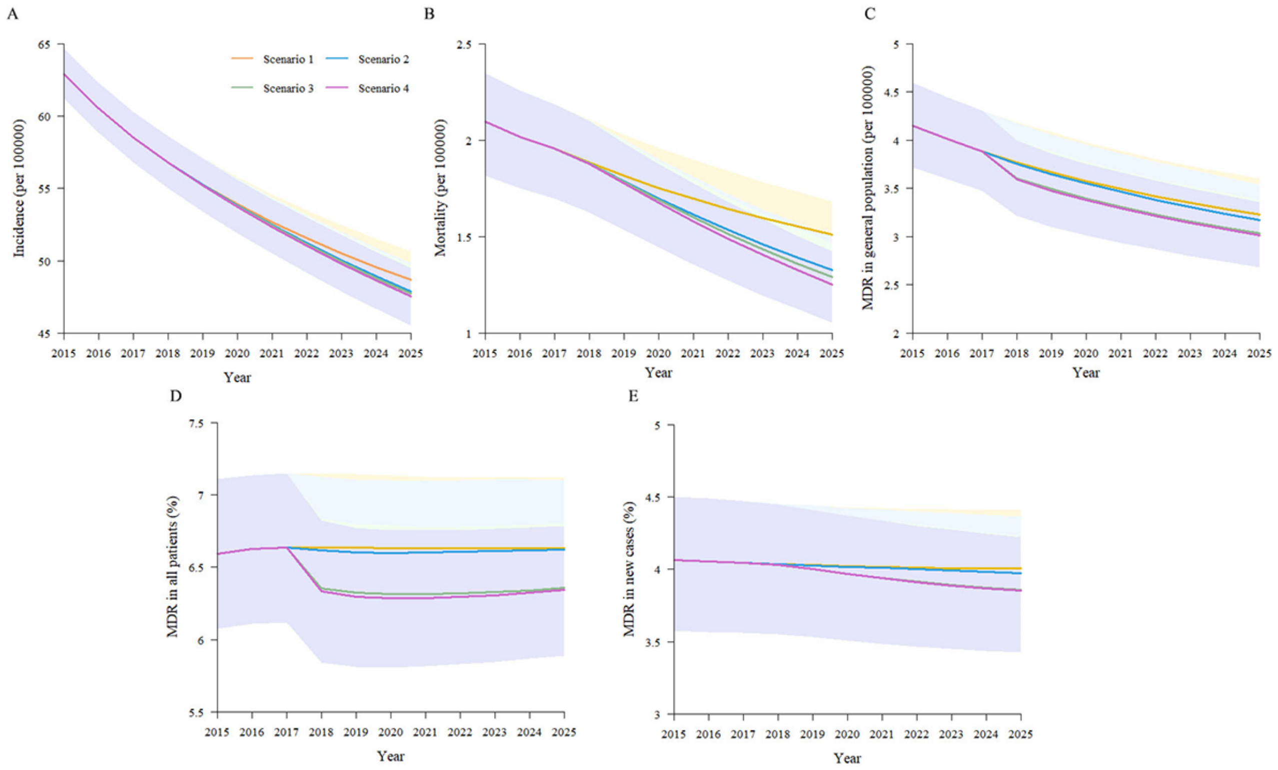


Table 4. Impacts of different intervention scenarios on pulmonary tuberculosis (PTB) epidemiology in Henan province, China, from 2015 to 2025.

Scenario	Tuberculosis incidence		Tuberculosis-related death	
	Cumulative number, n (95% CI)	Cumulative reduction, % (95% CI)	Cumulative number, n (95% CI)	Cumulative reduction, % (95% CI)
1	665,394 (643,266-688,305)	22.61 (21.72-23.62)	21,599 (18,716-24,151)	27.92 (27.02-28.28)
2	662,400 (640,116-685,372)	23.90 (22.96-25.03)	20,741 (17,881-23,273)	36.80 (36.06-38.04)
3	661,772 (639,449-684,757)	24.16 (23.20-25.31)	20,582 (17,725-23,112)	38.47 (37.71-39.85)
4	660,979 (638,633-683,974)	24.46 (23.48-25.63)	20,384 (17,528-22,907)	40.35 (39.39-42.08)

In scenario 2, if the treatment success rate of new and re-treatment cases is increased to 92% and 90%, respectively, the incidence and mortality will be decreased by 23.9% (95% CI 23.0%-25.0%) and 36.8% (95% CI 36.1%-38.0%), respectively, by 2025 (Figure 5A-B, Table 4).

In scenario 3, if 90% of TB smear-positive patients are tested for resistance testing, and the long-term cure rate for MDR-TB can reach 82%, the incidence and mortality was found to decline by 24.2% (95% CI 23.2%-25.3%) and 38.5% (95% CI 37.7%-39.9%), respectively (Figure 5A-B, Table 4). In addition, we observed that the proportions of MDR-TB in all patients with TB and new cases were declining slowly in Henan with the current measures. Moreover, the use of better diagnostic technologies and increasing treatment success for MDR-TB would yield the greatest percentage reduction in MDR-TB in the general population, all patients with TB, and new cases, by 26.9%, 3.6%, and 5.1%, respectively, from 2015 to 2025 (Figure 5C-E).

The greatest reduction was observed in scenario 4, where the combined strategy would yield reductions in incidence and mortality of 24.5% (95% CI 23.5%-25.6%) and 40.4% (95% CI 39.4%-42.1%), respectively, by 2025 (Figure 5A-B, Table 4).

Discussion

We used a provincial website tuberculosis surveillance dataset spanning 14 years to collect 976,526 PTB cases and 381,598 bacteriological-positive PTB cases beginning from 2005.

Consistent with the trend of substantially decreasing worldwide PTB incidence [9], Henan also witnessed a sharp decline in

PTB and bacteriological-positive PTB incidence rates, from $91.4/10^5$ to $58.5/10^5$ and $44.5/10^5$ to $14.7/10^5$, respectively. This result reflects the effectiveness of targeted TB control measures, such as the investment of 1.42 million US dollars to improve the cure rates of DS-TB. It is important to note that the successful treatment of a TB case translates into a reduction in infection sources, which, in turn, reduces transmission and incidence. The post-2015 global TB-related WHO targets propose a reduction of 50.0% in incidence and 75.0% in mortality by 2025 [10]. Recent studies show that, even under the most optimistic circumstances, China is unlikely to meet these global TB-related targets by intensifying its current strategies [5,6]. Our model also shows that although Henan province's decrease ranges for incidence (22.6%) and mortality (27.9%) are higher than those for China (20.1% and 23.1%, respectively), and even though all the changes and interventions considered in our analysis were implemented, Henan province will also not be able to achieve the post-2025 WHO global target. To reach this target, Henan province should actively detect and treat TB instead of adopting passive strategies [4,6].

The results indicate that in the context of TB prevention and control, more attention should be paid to males, elderly people, rural areas, and poverty-stricken regions. Specifically, the incidence in men was always higher than in women [1]. However, there are currently no prevention and control measures specifically targeted at the male population. Previous studies have shown that more frequent social communication and smoking could increase the risk of TB, especially sputum smear-positive TB [11-13]. Notably, compared with children under 15 years of age, the incidence in middle-aged and elderly people was high. Elderly people with low immune function may be particularly vulnerable to mycobacterial infections [14], and their treatment adherence is generally poor; therefore, the rate of unfavorable outcomes is higher in this sector of the population [14,15]. Moreover, up to 75.8% of residents aged ≥ 60 years in China have at least one chronic disease [16], making them more prone to complications and death when infected with TB [17]. Based on the above, the older TB cases should be proactively detected. Indeed, active TB detection and intervention in the older population may prove an effective public health strategy for reducing TB incidence [18].

The incidence of PTB in rural regions was significantly higher than in urban areas, and this result was similar to those from other regions in China. Cluster analysis also showed that regions in Henan province with high TB incidence were poverty-stricken counties (with a gross domestic product of <385.9 USD/person). In addition, our study also shows that bacteriological-positive PTB cases are concentrated in poverty-stricken counties, and the occurrence of bacteriological-positive PTB is associated with rural living. TB is associated with economic income [19], health and medical coverage, and culture and management. Regarding these aspects, rural regions lag behind urban areas. Therefore, these areas need to be designated as high-risk areas, or key areas where poverty alleviation should be strengthened, resident income should be increased, and medical service conditions should be improved.

In this study, we also observed that the proportion of bacteriological-positive PTB cases has decreased gradually. However, considering that bacteriological-positive cases are the main source of TB infection and that a single untreated patient with bacteriological-positive infection could infect 10-15 persons per year, new measures must be taken to achieve a significant reduction. At present, Henan province has carried out active surveillance on the fixed populations of rural areas to detect and treat all bacteriological-positive PTB patients, and also to design active detection strategies for key areas and key populations combined with detailed epidemiological analysis results.

The proportion of MDR-TB was high in China [9,20]; however, the dynamic compartmental model analysis revealed that improving the detection rate of, and increasing treatment success for, MDR-TB could yield the greatest reduction in the percentage of MDR-TB in all patients and new cases from 2015 to 2025. Based on this, 60% of smear-positive TB cases underwent tests for MDR-tuberculosis in Henan, including a routine drug sensitivity test and a rapid molecular drug sensitivity test, and second-line drugs, such as clofazimine and cycloserine, were provided to MDR-tuberculosis patients. However, the cure rates of clofazimine and cycloserine-containing regimens were only 68.7% and 66.0%, respectively [21]. Even if the cure rate of drugs against MDR-TB reaches 85.0% in the future, which could bring a greater reduction in the proportion of MDR-TB in all patients and new cases between 2015 and 2025, the WHO's goal would still not be reached. Nevertheless, improving cure rates of DS-TB and MDR-TB remains the keystone to reducing incidence and mortality.

Limitations

This study has some limitations. First, all the data used in this study were obtained from the TBIMS. However, drug resistance was not fully analyzed because the system contained incomplete drug resistance data. Second, no active detection and intervention data were obtained from the TBIMS. Hence, we only determined whether the WHO's goal could be achieved through passive strategies.

Conclusions

In China, from 1990 to 2020, 3 nationwide random surveys on TB were carried out to evaluate the effectiveness of the DOTS strategy. However, based on the internet-based TBIMS, we can now monitor the changes of TB epidemiological characteristics in real time while saving money, manpower, and time in the evaluation of the effect of different strategies. The local government has embarked on a series of nonfragmented TB-related intervention combinations, and the incidence of PTB and bacteriological-positive PTB has continued to decrease during the last 14 years; however, the WHO's 2025 goal will not be reached. New active—rather than passive—strategies for detection and intervention should be formulated based on epidemiological characteristics while improving the cure rates of DS-TB and MDR-TB.

Acknowledgments

This study was funded by grants from the National Key Research and Development Program (2018YFC2000300), National Science and Technology Major Project of China (2018ZX10302 302001004), National Natural Science Foundation of China (U1903118), Scientific Research Project of Beijing Educational Committee (SQKM201710025024), Beijing Natural Science Foundation（No Z20001), National Natural Science Foundation of China (No 11971478), and the Research Funds for the Major Innovation Platform of Public Health & Disease Control and Prevention, Renmin University of China.

Authors' Contributions

WX (wlxu@ruc.edu.cn) and WL (lwm_18@aliyun.com) were co-senior and co-corresponding authors on this study. HJ, GZ, JY, and DZ contributed equally to this study. WL and WX conceived, designed, and supervised the study. GZ and ZD collected the data. JY, FL, and CC cleaned the data. HJ, JY, FL, YY, FL, JX, and XL analyzed the data. HJ wrote the drafts of the manuscript. HJ and WL interpreted the findings. WL and WX commented on and revised the drafts of the manuscript. WL had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed description of the dynamic compartmental model.

[DOC File , 162 KB - publichealth_v7i1e24830_app1.doc]

References

1. Wang L, Zhang H, Ruan Y, Chin DP, Xia Y, Cheng S, et al. Tuberculosis prevalence in China, 1990–2010; a longitudinal analysis of national survey data. The Lancet 2014 Jun 14;383(9934):2057-2064. [doi: 10.1016/S0140-6736(13)62639-2] [Medline: 24650955]
2. World Health Organization. Global tuberculosis report 2016. Switzerland: World Health Organization.
3. Huang F, Cheng S, Du X, Chen W, Scano F, Falzon D, et al. Electronic recording and reporting system for tuberculosis in China: experience and opportunities. J Am Med Inform Assoc 2014;21(5):938-941 [FREE Full text] [doi: 10.1136/amiajnl-2013-002001] [Medline: 24326537]
4. Lönnroth K, Castro KG, Chakaya JM, Chauhan LS, Floyd K, Glaziou P, et al. Tuberculosis control and elimination 2010–50: cure, care, and social development. The Lancet 2010 May 22;375(9728):1814-1829. [doi: 10.1016/S0140-6736(10)60483-7] [Medline: 20488524]
5. Lin H, Wang L, Zhang H, Ruan Y, Chin DP, Dye C. Tuberculosis control in China: use of modelling to develop targets and policies. Bull World Health Organ 2015 Nov 01;93(11):790-798 [FREE Full text] [doi: 10.2471/BLT.15.154492] [Medline: 26549907]
6. Huynh GH, Klein DJ, Chin DP, Wagner BG, Eckhoff PA, Liu R, et al. Tuberculosis control strategies to reach the 2035 global targets in China: the role of changing demographics and reactivation disease. BMC Med 2015 Apr 21;13:88 [FREE Full text] [doi: 10.1186/s12916-015-0341-4] [Medline: 25896465]
7. Menzies NA, Cohen T, Lin H, Murray M, Salomon JA. Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. PLoS Med 2012;9(11):e1001347 [FREE Full text] [doi: 10.1371/journal.pmed.1001347] [Medline: 23185139]
8. Alkema L, Raftery AE, Brown T. Bayesian melding for estimating uncertainty in national HIV prevalence estimates. Sex Transm Infect 2008 Aug;84 Suppl 1:i11-i16 [FREE Full text] [doi: 10.1136/sti.2008.029991] [Medline: 18647860]
9. World Health Organization. Global tuberculosis report 2019. Switzerland: World Health Organization.
10. World Health Orgnaization. Global strategy and targets for tuberculosis prevention, care and control after 2015. Presented at: Sixty-seventh World Health Assembly; 21 May 2014; Geneva URL: http://www.who.int/tb/post2015_tbstrategy.pdf?ua=1
11. Zong Z, Huo F, Shi J, Jing W, Ma Y, Liang Q, et al. Relapse Versus Reinfection of Recurrent Tuberculosis Patients in a National Tuberculosis Specialized Hospital in Beijing, China. Front Microbiol 2018;9:1858 [FREE Full text] [doi: 10.3389/fmicb.2018.01858] [Medline: 30154770]
12. Leung CC, Yew WW, Chan CK, Chang KC, Law WS, Lee SN, et al. Smoking adversely affects treatment response, outcome and relapse in tuberculosis. Eur Respir J 2015 Mar;45(3):738-745 [FREE Full text] [doi: 10.1183/09031936.00114214] [Medline: 25359352]
13. Lee MS, Leung C, Kam K, Wong M, Leung MC, Tam C, et al. Early and late tuberculosis risks among close contacts in Hong Kong. Int J Tuberc Lung Dis 2008 Mar;12(3):281-287. [Medline: 18284833]
14. Rajagopalan S. Tuberculosis and aging: a global health problem. Clin Infect Dis 2001 Oct 01;33(7):1034-1039. [doi: 10.1086/322671] [Medline: 11528577]

15. Kwon YS, Chi SY, Oh IJ, Kim KS, Kim YI, Lim SC, et al. Clinical characteristics and treatment outcomes of tuberculosis in the elderly: a case control study. BMC Infect Dis 2013 Mar 05;13:121 [FREE Full text] [doi: 10.1186/1471-2334-13-121] [Medline: 23510403]
16. Wang LM, Chen ZH, Zhang M, Zhao ZP, Huang ZJ, Zhang X, et al. Study of the prevalence and disease burden of chronic disease in the elderly in China. Zhonghua Liu Xing Bing Xue Za Zhi 2019 Mar 10;40(3):277-283. [doi: 10.3760/cma.j.issn.0254-6450.2019.03.005] [Medline: 30884604]
17. 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]
18. Gao L, Lu W, Bai L, Wang X, Xu J, Catanzaro A, LATENTTB-NSTM study team. Latent tuberculosis infection in rural China: baseline results of a population-based, multicentre, prospective cohort study. Lancet Infect Dis 2015 Mar;15(3):310-319. [doi: 10.1016/S1473-3099(14)71085-0] [Medline: 25681063]
19. Wei W, Yuan-Yuan J, Ci Y, Ahan A, Ming-Qin C. Local spatial variations analysis of smear-positive tuberculosis in Xinjiang using Geographically Weighted Regression model. BMC Public Health 2016 Oct 06;16(1):1058 [FREE Full text] [doi: 10.1186/s12889-016-3723-4] [Medline: 27716319]
20. Du Y, Qiu C, Chen X, Wang J, Jing W, Pan H, et al. Treatment Outcome of a Shorter Regimen Containing Clofazimine for Multidrug-resistant Tuberculosis: A Randomized Control Trial in China. Clin Infect Dis 2020 Aug 14;71(4):1047-1054. [doi: 10.1093/cid/ciz915] [Medline: 31549147]
21. Wang J, Pang Y, Jing W, Chen W, Guo R, Han X, et al. Efficacy and safety of cycloserine-containing regimens in the treatment of multidrug-resistant tuberculosis: a nationwide retrospective cohort study in China. Infect Drug Resist 2019;12:763-770 [FREE Full text] [doi: 10.2147/IDR.S194484] [Medline: 31040707]

Abbreviations

AOR: adjusted odds ratio
DOTS: directly observed treatment short-course
DS-TB: drug-susceptible tuberculosis
MDR-TB: multidrug-resistant tuberculosis
OR: odds ratio
PTB: pulmonary tuberculosis
TB: tuberculosis
TBIMS: Tuberculosis Information Management System
WHO: World Health Organization

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by J Zhang; comments to author 03.11.20; revised version received 11.11.20; accepted 07.12.20; published 22.01.21.

Please cite as:
Jiang H, Zhang G, Yin J, Zhao D, Liu F, Yao Y, Cai C, Xu J, Li X, Xu W, Li W
Assessment of Strategies and Epidemiological Characteristics of Tuberculosis in Henan Province, China: Observational Study
JMIR Public Health Surveill 2021;7(1):e24830
URL: http://publichealth.jmir.org/2021/1/e24830/
doi:10.2196/24830
PMID:33480857

Original Paper

Electronic Cigarette Users' Perspective on the COVID-19 Pandemic: Observational Study Using Twitter Data

Yankun Gao[1], PhD; Zidian Xie[1], PhD; Dongmei Li[1], PhD

Department of Clinical & Translational Research, University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Dongmei Li, PhD
Department of Clinical & Translational Research
University of Rochester Medical Center
Saunders Research Building 1.303J
265 Crittenden Boulevard CU 420708
Rochester, NY, 14642-0708
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: Previous studies have shown that electronic cigarette (e-cigarette) users might be more vulnerable to COVID-19 infection and could develop more severe symptoms if they contract the disease owing to their impaired immune responses to viral infections. Social media platforms such as Twitter have been widely used by individuals worldwide to express their responses to the current COVID-19 pandemic.

Objective: In this study, we aimed to examine the longitudinal changes in the attitudes of Twitter users who used e-cigarettes toward the COVID-19 pandemic, as well as compare differences in attitudes between e-cigarette users and nonusers based on Twitter data.

Methods: The study dataset containing COVID-19–related Twitter posts (tweets) posted between March 5 and April 3, 2020, was collected using a Twitter streaming application programming interface with COVID-19–related keywords. Twitter users were classified into two groups: Ecig group, including users who did not have commercial accounts but posted e-cigarette–related tweets between May 2019 and August 2019, and non-Ecig group, including users who did not post any e-cigarette–related tweets. Sentiment analysis was performed to compare sentiment scores towards the COVID-19 pandemic between both groups and determine whether the sentiment expressed was positive, negative, or neutral. Topic modeling was performed to compare the main topics discussed between the groups.

Results: The US COVID-19 dataset consisted of 4,500,248 COVID-19–related tweets collected from 187,399 unique Twitter users in the Ecig group and 11,479,773 COVID-19–related tweets collected from 2,511,659 unique Twitter users in the non-Ecig group. Sentiment analysis showed that Ecig group users had more negative sentiment scores than non-Ecig group users. Results from topic modeling indicated that Ecig group users had more concerns about deaths due to COVID-19, whereas non-Ecig group users cared more about the government’s responses to the COVID-19 pandemic.

Conclusions: Our findings show that Twitter users who tweeted about e-cigarettes had more concerns about the COVID-19 pandemic. These findings can inform public health practitioners to use social media platforms such as Twitter for timely monitoring of public responses to the COVID-19 pandemic and educating and encouraging current e-cigarette users to quit vaping to minimize the risks associated with COVID-19.

(JMIR Public Health Surveill 2021;7(1):e24859) doi:10.2196/24859

KEYWORDS

COVID-19; Twitter; infodemiology; Electronic cigarette; perspective; observational; social media; vulnerable; sentiment analysis; topic modeling; concern

Introduction

The World health Organization declared COVID-19 as a pandemic on March 11, 2020 [1]. The United States has reported the highest number of confirmed COVID-19 cases globally [2]. With the spread of COVID-19, significant concern has been raised about the potential increased risk for electronic cigarette (e-cigarette) users to COVID-19 infection and related mortality [3,4]. Recent studies have shown that nicotine increases the expression of the angiotensin-converting enzyme 2 (ACE-2) in human bronchial epithelial cells. ACE-2 is the binding site for SARS-CoV-2, the virus that causes COVID-19 [5-8]. A national online survey study of 4351 youth and young adults showed a 5-fold increase in COVID-19 diagnoses among ever e-cigarette users compared to non-users [9]. However, no study has evaluated the attitudes of e-cigarette users toward the COVID-19 pandemic and whether their attitudes differ from those of nonusers. Therefore, it is important to characterize how e-cigarette users perceive the COVID-19 pandemic and how their perception differs from that of non-users. These findings will facilitate us to understand how e-cigarette users might respond to the COVID-19 pandemic, especially in terms of vaping.

Twitter is one of the most popular social media platforms, with an average of 330 million monthly active users sharing content on the platform, as of 2019 [10]. Twitter users can publish publicly available posts (called tweets), making Twitter a rich data source to monitor social phenomena and public health issues [11]. This study focused on understanding how Twitter users in the United States who used e-cigarettes responded to the COVID-19 pandemic by using sentiment analysis and topic modeling to extract users' subjective attitudes and to identify topics from the textual contents of their tweets. Understanding the attitudes of e-cigarette users toward the COVID-19 pandemic and topics discussed by them on Twitter could help public health workers and policymakers take appropriate actions such as encouraging e-cigarette users to quit vaping during the current COVID-19 pandemic.

Methods

Data Collection

Since the correlation between COVID-19 and e-cigarettes has been a popular topic during the current pandemic, tweets about e-cigarettes in our COVID-19 dataset were not necessarily from e-cigarette–related user accounts. Therefore, to identify e-cigarette users, we used an e-cigarette–related dataset from 2019, that is, before the COVID-19 pandemic. Tweets were collected between May 2019 and August 2019 through a Twitter streaming application programming interface (API) by using e-cigarette–related keywords (ie, "e-cig," "e-cigs," "Ecig," "Ecigs," "electroniccigarette," "Ecigarette," "Ecigarettes," "vape," "vapers," "vaping," "vapes," "e-liquid," "ejuice," "eliquid," "e-juice," "vapercon," "vapeon," "vapefam," "vapenation," and "juul") [12,13]. In addition, a list of spam-specific keywords was used to remove tweets that were unrelated to e-cigarettes [14]. In this e-cigarette–related dataset, Twitter users whose username and user screen name did not contain any e-cigarette keywords were considered as e-cigarette users. Although we intended to use tweets before the announcement of flavor ban policies in different states (starting from September 2019) to identify users who tweeted about e-cigarettes in order to avoid the potential noise, the starting point (ie, May 2019) was randomly selected.

The COVID-19 dataset was collected using a Twitter streaming API to crawl Twitter posts between March 5, 2020, and April 3, 2020, with coronavirus-related keywords ("CORONA," "corona," "COVID19," "covid19," "covid," "coronavirus," "Coronavirus," "CoronaVirus," and "NCOV"), which were identified from COVID-19–related tweets. Twitter IDs were used to identify unique Twitter users. To get a clean dataset, promotion-related Twitter IDs and posts were filtered out. In addition, tweets that mentioned "corona" (a brand name for beer) as a beverage were removed from the dataset. The keywords used to clean the COVID-19 dataset included "dealer," "deal," "supply," "beer," "drink," "drank," "drunk," "store," "promo," "promotion," "customer," "discount," "sale," "free shipping," "sell," "$," "%," "dollar," "offer," "percent off," "save," "price," and "wholesale". After filtering the data, US-based Twitter posts were selected using geolocation keywords, such as "United States," "New York," "USA," and "US." Duplicate tweets were removed, and retweets were included in the final dataset. The tweets in the US COVID-19 dataset, which were posted by the above-identified e-cigarette users, were classified as the e-cigarette (Ecig) group. The remaining COVID-19 tweets were classified as the non-e-cigarette (non-Ecig) group.

Ethical Statement

In this descriptive, observational study, we collected and analyzed user-generated content from Twitter. No intervention or interaction was made with the users who posted information on Twitter. The identifiers that could be associated with the Twitter data are usernames or Twitter handles, which are accessible by the public or anyone with internet access. All the usernames or Twitter handles in the study were randomly assigned a numerical number after the Twitter data was collected.

Data Availability Statement

The data and scripts used for the analyses and to create figures are available on request from the corresponding author.

Sentiment Analysis

Sentiment analysis refers to the contextual mining of an incoming message, which can extract the underlying attitudes and determine whether the sentiment is positive, negative, or neutral. The sentiment score for each tweet in our dataset was computed using VADER (Valence Aware Dictionary and sEntiment Reasoner), a tool used to obtain sentiments from social media data [15]. Each user's average sentiment score was calculated for each day of the study period. The mean value of the average sentiment score of users from the same group was then calculated to represent the overall group sentiment for each day. A sentiment score of +0.05 or higher denotes a positive attitude. A sentiment score −0.05 or lower denotes a negative attitude. A sentiment score between −0.05 and +0.05 denotes a neutral attitude. The mean sentiment scores were then examined longitudinally across the study period to evaluate their potential links with COVID-19 spread and government policy changes.

Topic Modeling

Topic modeling, specifically the latent Dirichlet allocation (LDA) model, was used for text content analysis. LDA is a 3-layer hierarchical Bayesian model, in which each word in the document is modeled into a specific topic, and the words in each topic are weighted based on their appearance [16]. Using the LDA model allowed us to identify the topics of conversations in both Ecig and non-Ecig groups. Next, data cleaning processes were performed. All punctuation, white spaces, stop words were removed. In addition, uppercase characters were converted to lowercase characters. Words were lemmatized to their stem form to ignore different tenses, and frequent bigrams and trigrams were identified using a Python library, Gensim. Topic modeling was applied to tweets from both Ecig and non-Ecig groups. Topic coherence was used to determine the optimal number of topics to identify the frequently discussed topics in each group [17].

Results

The US COVID-19 tweets dataset between March 5, 2020, and April 3, 2020, consisted of 10,902,142 tweets from 2,144,599 unique Twitter users. From the e-cigarette–related tweets dataset generated between May 2019 and August 2019, we identified 930,290 tweets from 902,310 unique Twitter users. From the COVID-19 tweets collected, we identified 11,479,773 tweets from 2,511,659 unique Twitter users in the non-Ecig group and 4,500,248 tweets from 187,399 unique Twitter users in the Ecig group.

Figure 1 shows the average sentiment score of COVID-19 tweets in each group from March 5 to April 3, 2020. Users in neither the Ecig group nor the Non-Ecig group showed a positive attitude. Other than on March 7, 2020, Ecig group users showed a more negative attitude towards COVID-19 than non-Ecig group users. Except in early March, the average sentiment scores of non-Ecig group users were mostly neutral. In contrast, Ecig group users had a negative sentiment for almost the entire study period. The sentiment scores from both groups showed similar trends over time.

To obtain content-wise insights from the discussions in Ecig and non-Ecig groups, the LDA topic model was applied to the tweets posted by users from both groups. Tables 1 and 2 summarize the popular topics discussed in each group, including the top 10 keywords for each topic. The top 3 topics (percentage of tokens) in the Ecig group included “Trump handling corona” (12.8%), “Death toll” (11.7%), and “Stay home” (11.3%). The top 3 topics (percentage of tokens) in the non-Ecig group included “Trump blame China” (12.9%), “Hospital caring and testing” (10.7%), and “COVID testing” (10.5%).

Figure 1. Comparison of different sentiments toward COVID-19 between US-based Twitter users who used e-cigarettes (Ecig group) and those who did not use e-cigarettes (non-Ecig group) from March 5, 2020, to April 3, 2020.

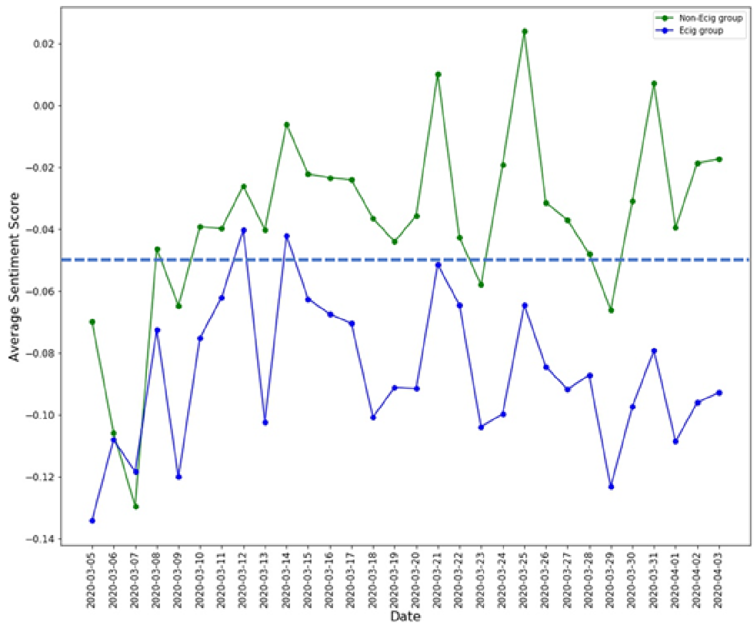


Table 1. Major topics discussed based on COVID-19 tweets posted by e-cigarette users (Ecig group) in the US, from March 5, 2020, to April 3, 2020.

Topic	Percentage of tokens (%)	Keywords
Trump handling corona	12.8	Trump, say, call, lie, take, would, president, response, medium, job
Death toll	11.7	Test, case, death, positive, new, report, number, confirm, people, day
Stay home	11.3	Health, home, stay, pay, public, worker, family, leave, need, emergency
Death and virus spread	10.7	Virus, corona, people, take, year, many, know, die, spread, time
Testing of COVID	9.3	Get, test, day, week, say, kit, testing, go, covid, make
Trump wants hospitals and doctors to help patients	9.2	Patient, trump, want, help, people, doctor, know, give, say, hospital
Virus spread	9.1	Spread, Chinese, stop, response, virus, help, make, pandemic, covid, global
Combat with COVID	7.1	Know, vote, say, hand, bill, would, may, travel, good, time
School shutdown	7	Due, close, cancel, go, concern, hear, school, people, get, say

Table 2. Major topics discussed in COVID-19 tweets posted by e-cigarette nonusers (non-Ecig group) in the US, from March 5, 2020, to April 3, 2020.

Topic	Percentage of Tokens (%)	Keywords
Trump blame China	12.9	Trump, say, call, people, lie, Chinese, medium, government, response, crisis
Hospital caring and testing	10.7	Covid, need, patient, hospital, fight, help, care, worker, testing, family
COVID testing	10.5	Virus, corona, test, get, positive, people, symptom, day, know, go
Death toll	10.4	Case, death, report, number, test, new, confirm, day, first, state
Stop work and stay home	9.6	People, work, amp, stop, hand, go, die, stay, home, get
School and business shutdown	9	Due, spread, school, cancel, close, concern, health, public, plan, business
Relief bill	8.1	Vote, watch, time, hold, people, bill, play, run, relief, help
Stay home	8	Take, home, stay, covid, go, people, order, spread, good, say
Response to COVID	7.9	Thank, response, question, covid, share, late, update, ask, release, amp

Discussion

Principal Findings

During the COVID-19 pandemic, people worldwide have widely used Twitter to follow news and express their opinions and responses to the pandemic [18]. Although Twitter users in the non-Ecig group had a neutral attitude toward COVID-19 during most of the study period (March 5 to April 3, 2020), Twitter users in the Ecig group had a negative attitude toward this pandemic. The topics most frequently discussed by Ecig group users were how the US President Donald Trump handled COVID-19, deaths due to COVID-19, and staying at home. On the other hand, the most frequently discussed topics in the non-Ecig group included Trump blames China, hospital care for patients with COVID-19, and COVID-19 testing. The differences between Ecig and non-Ecig group users' attitudes toward the COVID-19 pandemic indicated a good opportunity to educate e-cigarette users about the potential harms of vaping and encourage them to quit vaping during the COVID-19 pandemic.

The average sentiments of users from both Ecig and non-Ecig groups were relatively parallel during the study period, which suggests that the dynamic changes of the COVID-19 pandemic and other factors (such as government policies) had similar effects on the sentiments of e-cigarette users and nonusers towards the COVID-19 pandemic. We noticed, however, that some variation in the sentiment scores might be associated with the government's policies. For example, when all nonessential businesses in New York City—the worst-affected area—were closed on March 22, 2020, sentiment scores of users in both Ecig and non-Ecig groups decreased to trough on March 23. The sentiment scores of non-Ecig group users quickly reached the highest peak on March 25 when the Congress agreed on a $2 trillion virus relief package bill.

Our study findings show that Ecig group users presented a more negative attitude towards the COVID-19 pandemic than did non-Ecig group users. Moreover, Ecig group users discussed more topics related to death and virus spread. Some of the common topics discussed in both groups included how Trump responded to COVID-19, deaths due to COVID-19, and social distancing practices such as staying at home and shutting down schools. One of the top topics unique to the discussion in the Ecig group was death and virus spread, which did not feature among the top topics discussed in the non-Ecig group. The concerns in the Ecig group about the virus spread and COVID-19–related deaths might be related to the discussions that vaping may increase the risk of severe COVID-19 infection. Starting from 2019, the epidemic of vaping associated lung injury (EVALI) in the US drew significant attention among the public [19]. An early study (February 2020) showed that patients with COVID-19 had similar symptoms as EVALI, such as fever and cough, as well as characteristic lung phenotypes [20]. In addition, studies have shown that e-cigarette use can suppress the genes related to the immune and inflammatory response [21,22], which could increase the duration and severity of respiratory infections. These findings might lead to more concerns about the possible connection between vaping and the COVID-19 pandemic for e-cigarette users.

Systematic surveillance of vaping-related discussions on Twitter identified public health–related topics at the intersection of vaping and COVID-19; these topics included health concerns as well as unsubstantiated health claims [23]. Currently, there is a lack of evidence that e-cigarette users are more susceptible to COVID-19 infection and death. Although public health experts claim that vaping and smoking could increase the risk of COVID-19 infection, and multiple research studies have suggested that smoking is associated with adverse outcomes of COVID-19 [24,25], a study in Europe published contrasting conclusions that daily smokers had a lower risk of developing severe COVID-19 symptoms [26]. Future studies should further investigate the association of smoking or vaping with COVID-19 infection and death. Notably, the abovementioned European study was published in the end of April 2020, which was beyond our study period. Therefore, how the report of the negative association between smoking and COVID-19 affects the sentiments of people, especially e-cigarette users and smokers, awaits further investigation.

Limitations

This study has several limitations. First, as with many other social media studies using Twitter data, significant geographic bias exists in the sentiments expressed in tweets over the same time period [27]. Moreover, the sentiments expressed in tweets could be biased based on the geographic location—whether the user is local or visiting that area and what other activities they have completed during the course of a day [28]. Second, the generalization of the study results is limited by the representation of Twitter users in the general population. Twitter users are relatively younger and more educated than the general population [29]. Highly active Twitter users also have different behaviors than the rest of the Twitter population [29]. Third, some Twitter account types, such as information aggregators, which could also aggregate vaping discount information but were not e-cigarette users, were not removed from our dataset and could introduce some bias in the results of the analysis [30]. Furthermore, the non-Ecig group may include some e-cigarette users who were not identified from the earlier e-cigarette–related dataset, which could also introduce bias in the results. Fourth, some Twitter accounts were marked as private from the API; therefore, we were unable to retrieve tweets from those accounts. Fifth, only a small proportion of Twitter accounts provided the geolocation, and we could only select Twitter accounts that provided this information [31]. Sixth, other than human users, there are some social bots accounts on Twitter. However, those bot accounts were not excluded in this study, which may also cause some bias. Moreover, this study did not identify smokers in both groups who might have different attitudes towards the pandemic, which might lead to some additional bias in the results. As we defined the Ecig group based on Twitter data collected from May to August 2019, e-cigarette users who did not post e-cigarette–related tweets during this period might be mislabeled and subsequently misclassified into the non-Ecig group. Moreover, non-Ecig users in that period could have become e-cigarette users during the COVID-19 pandemic. This could introduce potential selection bias and misclassification in both directions given the time lag. Seventh, we could not distinguish individual accounts from institutional or group accounts based on the Twitter data; thus, the information about user attitudes toward COVID-19 might not all represent individuals. Finally, our study period was in the early stage of the pandemic with limited information available about the potential link between vaping and COVID-19, which might introduce some biases. With the rapid spread of the COVID-19 pandemic and emergence of more evidence on the link between vaping and COVID-19, the perception and responses to the COVID-19 pandemic of the public, including e-cigarette users, might evolve; however, this requires re-evaluation of the outcomes using more recent Twitter data.

Conclusions

In this study, Twitter users in the Ecig group showed a more negative attitude toward the COVID-19 pandemic than those in the non-Ecig group. This study highlights the importance of using Twitter for surveillance of public responses to the COVID-19 pandemic, which can provide early insights for public health awareness, especially among specific population groups (such as e-cigarette users). Users in the Ecig group discussed topics such as the spread of the virus and COVID-19–related deaths, which highlights these vapers' concerns about the potentially elevated risks of COVID-19. These findings may provide a useful opportunity for public health practitioners to educate current e-cigarette users and encourage them to quit vaping to reduce the risks associated with COVID-19.

Authors' Contributors

YG, ZX, and DL conceived and designed the study. YG analyzed the data. YG wrote the manuscript. YG, ZX, and DL assisted with the interpretation of analyses and edited the manuscript.

Conflicts of Interest

None declared.

Acknowledgments

This study was supported by the National Cancer Institute of the National Institutes of Health (NIH) and the Food and Drug Administration (FDA) Center for Tobacco Products under Award Number U54CA228110.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Johns Hopkins University of Medicine - Coronavirus Resource Center. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-12-31]
3. Smoke or Vape? Here's What You Need to Know About COVID-19 Risks. Healthline. 2020. URL: https://www.healthline.com/health/smoking-vaping-covid-19-risks [accessed 2020-12-31]
4. What You Need to Know About Smoking, Vaping and COVID-19. American Lung Association. 2020 Mar 27. URL: https://www.lung.org/blog/smoking-and-covid19 [accessed 2020-12-31]
5. Kaur G, Lungarella G, Rahman I. SARS-CoV-2 COVID-19 susceptibility and lung inflammatory storm by smoking and vaping. J Inflamm (Lond) 2020;17:21 [FREE Full text] [doi: 10.1186/s12950-020-00250-8] [Medline: 32528233]
6. McAlinden KD, Eapen MS, Lu W, Chia C, Haug G, Sohal SS. COVID-19 and vaping: risk for increased susceptibility to SARS-CoV-2 infection? Eur Respir J 2020 Jul;56(1) [FREE Full text] [doi: 10.1183/13993003.01645-2020] [Medline: 32430427]
7. Sharma P, Zeki AA. Does vaping increase susceptibility to COVID-19? Am J Respir Crit Care Med 2020 Oct 01;202(7):1055-1056. [doi: 10.1164/rccm.202005-2103le]
8. Zhang H, Rostami MR, Leopold PL, Mezey JG, O'Beirne SL, Strulovici-Barel Y, et al. Expression of the SARS-CoV-2 receptor in the human airway epithelium. Am J Respir Crit Care Med 2020 Jul 15;202(2):219-229. [doi: 10.1164/rccm.202003-0541oc]
9. Gaiha SM, Cheng J, Halpern-Felsher B. Association between youth smoking, electronic cigarette use, and COVID-19. J Adolesc Health 2020 Oct;67(4):519-523 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.002] [Medline: 32798097]
10. Clement J. Statista. 2020 Feb 02. URL: https://www.statista.com/topics/737/twitter/ [accessed 2020-12-31]
11. Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. J Med Internet Res 2015 Dec 06;17(11):e251 [FREE Full text] [doi: 10.2196/jmir.4466] [Medline: 26545927]
12. Lu X, Chen L, Yuan J, Luo J, Luo J, Xie Z, et al. User perceptions of different electronic cigarette flavors on social media: observational study. J Med Internet Res 2020 Jun 24;22(6):e17280 [FREE Full text] [doi: 10.2196/17280] [Medline: 32579123]
13. Chen L, Lu X, Yuan J, Luo J, Luo J, Xie Z, et al. A social media study on the associations of flavored electronic cigarettes with health symptoms: observational study. J Med Internet Res 2020 Jun 22;22(6):e17496 [FREE Full text] [doi: 10.2196/17496] [Medline: 32568093]
14. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
15. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. : AAAI Publications; 2014 May 16 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI.
16. Blei DM, Ng AY, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2013 Jan;3(7):993 [FREE Full text]
17. Korenčić D, Ristov S, Šnajder J. Document-based topic coherence measures for news media text. Expert Syst Appl 2018 Dec;114:357-373. [doi: 10.1016/j.eswa.2018.07.063]
18. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
19. Salzman G, Alqawasma M, Asad H. Vaping associated lung injury (EVALI): an explosive United States epidemic. Mo Med 2019;116(6):492-496 [FREE Full text] [Medline: 31911735]
20. Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, et al. Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology 2020 Apr;295(1):210-217 [FREE Full text] [doi: 10.1148/radiol.2020200274] [Medline: 32027573]
21. Martin EM, Clapp PW, Rebuli ME, Pawlak EA, Glista-Baker E, Benowitz NL, et al. E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke. Am J Physiol Lung Cell Mol Physiol 2016 Jul 01;311(1):L135-L144 [FREE Full text] [doi: 10.1152/ajplung.00170.2016] [Medline: 27288488]
22. Law SM, Gray RD. Neutrophil extracellular traps and the dysfunctional innate immune response of cystic fibrosis lung disease: a review. J Inflamm (Lond) 2017;14:29 [FREE Full text] [doi: 10.1186/s12950-017-0176-1] [Medline: 29299029]

23. Majmundar A, Allem JP, Cruz TB, Unger JB. Public health concerns and unsubstantiated claims at the intersection of vaping and COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1667-1668 [FREE Full text] [doi: 10.1093/ntr/ntaa064] [Medline: 32285129]
24. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
25. Vardavas C, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. Tob Induc Dis 2020 Mar 20;18(March):20 [FREE Full text] [doi: 10.18332/tid/119324] [Medline: 32206052]
26. Miyara M, Tubach F, Amoura Z. Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19 infection. Qeios 2020 Apr 19. [doi: 10.32388/wpp19w]
27. Gore RJ, Diallo S, Padilla J. You are what you Tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
28. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
29. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center [Internet & Technology]. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2020-12-31]
30. Kim A, Miano T, Chew R, Eggers M, Nonnemaker J. Classification of Twitter users who Tweet about e-cigarettes. JMIR Public Health Surveill 2017 Oct 26;3(3):e63 [FREE Full text] [doi: 10.2196/publichealth.8060] [Medline: 28951381]
31. Takhteyev Y, Gruzd A, Wellman B. Geography of Twitter networks. Social Networks 2012 Jan;34(1):73-81. [doi: 10.1016/j.socnet.2011.05.006]

Abbreviations

ACE-2: angiotensin-converting enzyme 2
API: application programming interface
Ecig users: electronic cigarette users
EVALI: epidemic of vaping associated lung injury
LDA: latent Dirichlet allocation
Non-Ecig users: non-electronic cigarette users
VADER: Valence Aware Dictionary and sEntiment Reasoner

Edited by T Sanchez; submitted 07.10.20; peer-reviewed by R Gore, K Harding - Wheeler; comments to author 17.11.20; revised version received 07.12.20; accepted 09.12.20; published 05.01.21.

Please cite as:
Gao Y, Xie Z, Li D
Electronic Cigarette Users' Perspective on the COVID-19 Pandemic: Observational Study Using Twitter Data
JMIR Public Health Surveill 2021;7(1):e24859
URL: http://publichealth.jmir.org/2021/1/e24859/
doi:10.2196/24859
PMID:33347422

Original Paper

Social Media Use, Unhealthy Lifestyles, and the Risk of Miscarriage Among Pregnant Women During the COVID-19 Pandemic: Prospective Observational Study

Xiaotong Zhang[1], BA; Jue Liu[2], PhD; Na Han[3], MD; Jing Yin[3], BA

[1]School of Medicine, Macau University of Science and Technology, Macau, China

[2]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

[3]Department of Obstetrics and Gynecology, Maternal and Child Health Hospital of Tongzhou District, Beijing, China

Corresponding Author:
Jue Liu, PhD
Department of Epidemiology and Biostatistics
School of Public Health
Peking University
No 38 Xueyuan Rd, Haidian District
Beijing, 100191
China
Phone: 86 010 8280 1528 ext 316
Email: jueliu@bjmu.edu.cn

Abstract

Background: The COVID-19 pandemic has resulted in changes to normal life and disrupted social and economic function worldwide. However, little is known about the impact of social media use, unhealthy lifestyles, and the risk of miscarriage among pregnant women during the COVID-19 pandemic.

Objective: This study aims to assess the association between social media use, unhealthy lifestyles, and the risk of miscarriage among pregnant women in the early stage of the COVID-19 pandemic in China.

Methods: In this prospective cohort study, 456 singleton pregnant women in mainland China were recruited during January and February 2020. Sociodemographic characteristics, history of previous health, social media use, and current lifestyles were collected at baseline, and we followed up about the occurrence of miscarriage. Log-binomial regression models were used to estimate the risk ratios (RRs) of miscarriage for women with different exposures to COVID-19–specific information.

Results: Among all the 456 pregnant women, there were 82 (18.0%) who did no physical activities, 82 (18.0%) with inadequate dietary diversity, 174 (38.2%) with poor sleep quality, and 54 (11.8%) spending >3 hours on reading COVID-19 news per day. Women with excessive media use (>3 hours) were more likely to be previously pregnant (P=.03), have no physical activity (P=.003), have inadequate dietary diversity (P=.03), and have poor sleep quality (P<.001). The prevalence of miscarriage was 16.0% (n=73; 95% CI 12.6%-19.4%). Compared with women who spent 0.5-2 hours (25/247, 10.1%) on reading COVID-19 news per day, miscarriage prevalence in women who spent <0.5 hours (5/23, 21.7%), 2-3 hours (26/132, 19.7%), and >3 hours (17/54, 31.5%) was higher (P<.001). Miscarriage prevalence was also higher in pregnant women with poor sleep quality (39/174, 22.4% vs 34/282, 12.1%; P=.003) and a high education level (66/368, 17.9% vs 7/88, 8.0%; P=.02). In the multivariable model, poor sleep quality (adjusted RR 2.06, 95% CI 1.24-3.44; P=.006), 2-3 hours of media use daily (adjusted RR 1.74, 95% CI 1.02-2.97; P=.04), and >3 hours of media use daily (adjusted RR 2.56, 95% CI 1.43-4.59; P=.002) were associated with miscarriage. In the sensitivity analysis, results were still stable.

Conclusions: Pregnant women with excessive media use were more likely to have no physical activity, inadequate dietary diversity, and poor sleep quality. Excessive media use and poor sleep quality were associated with a higher risk of miscarriage. Our findings highlight the importance of healthy lifestyles during the COVID-19 pandemic.

(JMIR Public Health Surveill 2021;7(1):e25241) doi:10.2196/25241

KEYWORDS

COVID-19; social media use; miscarriage; cohort study; pregnancy; pregnant women; social media; China; risk; prospective; online health information

Introduction

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has resulted in changes to normal life and has disrupted social and economic functions worldwide. It was reported that there were 52,487,476 confirmed cases and 1,290,653 deaths as of November 13, 2020 [1]. Compared with seasonal influenza, COVID-19 has a higher case-fatality ratio (0.98%-5.9% vs 0.1%) and infectivity (R_0: 2.3-6.2 vs 1.2-1.4) [2-5]. Mortality in patients with COVID-19 has been associated with age and comorbidities (eg, hypertension, diabetes, and cardiovascular diseases) and differed across countries [6,7]. To date, no specific treatment has been found for COVID-19 and supportive measures have been used for patients with COVID-19. Nonpharmacologic interventions remain the key for curbing the spread of the virus, including active case finding and management, identification and quarantine of close contacts, social distancing, and personal protection (eg, hand hygiene and face mask use) [5]. China has taken strict measures to prevention and control of the pandemic, especially on social distancing and social isolation during the early stage of the pandemic. Wuhan City suspended all transportation in and out of the city from January 23 to April 8, 2020. Intra-area and interarea transportation restrictions were applied throughout the entire country of China, from big cities to small villages, from January to February 2020.

Along with the transmission of SARS-CoV-2, the information related with COVID-19 was also spread rapidly. One of the most accessible and fastest platforms for broadcasting information is social media. It represents a conglomerate of electronic platforms used for creating and sharing information, ideas, messages, etc [8]. Social media has become the major source of information about COVID-19. It enabled rapid and widespread reach of public health communications to help individuals take timely self-protection interventions. However, the speedy spread of COVID-19 worldwide also became a source of public worry, and several unknowns regarding this new pathogen created a state of panic [9]. Previous studies have shown that media coverage of COVID-19–related news induced fear and caused psychological stress during geographical lockdowns, extended quarantines, and financial and social hardships [9].

Miscarriage is a common adverse pregnancy outcome and one of the major public health problems. Miscarriage refers to a spontaneous demise of pregnancy before the fetus reaches viability [10]. Approximately 25% of pregnancies end in miscarriage, most occurring within early pregnancy (<13 weeks) [11]. Although the causes of miscarriage have not been fully explained, previous studies have shown a negative association of advanced maternal age (≥35 years), tobacco use, psychological problems, BMI, and other unhealthy lifestyles with miscarriage [12-14]. Excessive media consumption about COVID-19 was reported to be associated with increased anxiety in the general population in a cross-sectional study conducted in Russia [15]. Currently, little is known about the association between social media use, unhealthy lifestyles, and the risk of miscarriage among pregnant women. In this prospective cohort study, we aim to assess the association between social media use, unhealthy lifestyles, and the risk of miscarriage among pregnant women in the COVID-19 pandemic.

Methods

Study Design

This was a prospective cohort study conducted in a tertiary maternal and child health hospital in Beijing, China. The hospital was responsible for the prenatal care of all pregnant women living in Tongzhou district of Beijing. The primary aim of this cohort study is to investigate the short- and long-term health effects of prenatal exposures (eg, poor sleep quality) on mothers and their children. Baseline recruitment was conducted in January and February 2020, and pregnant women who visited the outpatient clinic for the first prenatal examination at Tongzhou Maternal and Child Health Hospital were recruited when they met the following inclusion criteria: <14 gestational weeks, singleton pregnancy, plan to have antenatal care and delivery in Tongzhou Maternal and Child Health Hospital, and resided in Tongzhou during the past half year and have no plan to move out after delivery. The study was approved by the institutional review boards at Peking University (IRB00001052–18003), and all participants gave written informed consent at the enrollment.

Baseline information was collected at the first prenatal visit by trained medical workers through a standardized questionnaire, such as sociodemographic characteristics (age, educational level, region, and family income), history of previous health (cesarean section, preterm birth, miscarriage, and fist pregnancy or not), prepregnancy weight, smoking status, physical activities, dietary diversity, and sleep quality. There were 504 pregnant women that met the inclusion criteria and were recruited at baseline. By August 2020, 11 women moved out of Tongzhou, 27 women were lost to follow-up, and 10 women transferred to other hospitals. Finally, the remaining 456 participants were included in this study.

Assessment of Media Use About COVID-19

We collected the information on media use about COVID-19 by the following question: “How long did you spent on reading COVID-19 news every day from social media (official or unofficial)?” Participants were divided into five COVID-19–specific information exposure groups (<0.5 hours, 0.5 hours-1 hour, 1 hour-2 hours, 2-3 hours, and >3 hours) by their answers regarding the time spent on reading COVID-19 news. Because of the similar prevalence of miscarriage in the 0.5 hours-1 hour and 1 hour-2 hours groups in this study, we combined the two groups into one (0.5-2 hours) group as the reference group in the final analysis.

Follow-up for Pregnancy Outcomes

The follow-up of the pregnancy outcomes was conducted by local medical workers. Pregnant women took regular antenatal care and delivered in the hospital. The information on pregnancy outcomes was obtained though the medical electronic information system in the hospital, which automatically recorded information during each antenatal care and delivery. Miscarriage is defined as a pregnancy loss that occurs before 20 completed weeks of gestational age [10]. The prevalence of miscarriage was defined as the proportion of participants who had a miscarriage to all participants.

Covariates

Covariates were collected at the first prenatal visit, including age, educational level, region, family income, history of cesarean section, history of preterm birth, history of miscarriage, gravidity, prepregnancy weight, smoking status, physical activities, dietary diversity, and sleep quality. Prepregnancy BMI was calculated using weight (in kilograms) divided by the square of height (in meters). We assessed dietary diversity using nine food groups, as reported in previous studies [16]. The participants reported their consumption frequencies of various food groups, including meat, vegetables, fish, eggs, fruits, legumes, milk, rice, and nuts. The dietary diversity score (DDS) was calculated, with scores ranging from 0 to 9. Inadequate dietary diversity was defined as DDS<7 [16]. Sleep quality was evaluated by the Pittsburgh Sleep Quality Index (PSQI) [17]. The PSQI is the gold standard questionnaire for assessing subjective sleep quality and is framed in a 4-point Likert scale (0-3) analyzing seven factors, including subjective sleep quality, sleep duration, sleep latency, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction [17]. The scores from each component were added to give a sum score, also called a total score (range 0-21). Poor sleep quality was defined as a PSQI≥5 in the pregnant women [18].

Statistical Analysis

Mean (SD) values and proportions of baseline characteristics were calculated. We calculated the mean (SD) for age. We used proportions to describe baseline characteristics of pregnant women, such as age group, region, and educational levels, and the chi-square test or Fisher exact probability test was used to compare the distributions of baseline characteristics according to time spent on reading COVID-19 news. Prevalence of miscarriage and its 95% CI was calculated. Miscarriage prevalence in women with different characteristics were also compared using chi-square test or Fisher exact probability test.

Multivariable log-binomial regression models were used to estimate the adjusted risk ratios (RRs) and their 95% CIs of miscarriage for women with different exposures of COVID-19–specific information. Women were divided into four exposure groups (<0.5 hours, 0.5-2 hours, 2-3 hours, and >3 hours) by their answers regarding the time spent on reading COVID-19 news. Women who spent 0.5-2 hours on reading COVID-19 news per day were set as the reference group in the final analysis. In the multivariable model, we additionally adjusted for other potential risk factors, including age group (<35 years or ≥35 years), educational level (high school or below, or college or above), region (rural or urban), family income (<¥5000 [US $764], ¥5000-¥10,000 [US $764-$1528], or >¥10,000 [US $1528]), history of cesarean section (no or yes), history of preterm birth (no or yes), history of miscarriage (no or yes), first pregnancy (no or yes), prepregnancy BMI (underweight, normal weight, overweight, or obese), smoking (nonsmoker, previous smoker, or current smoker), physical activities (never, sometimes, usually, or every day), inadequate dietary diversity (no or yes), and poor sleep quality (no or yes) by backward methods. To examine the robustness of our findings, we did sensitivity analyses by adjusted covariates in the multivariable models as continuous variables for several variables (age and DDS), instead of categorical variables. In the subgroup analysis, we divided women into different subgroups by baseline characteristics (region, age group, and history of miscarriage). Among these baseline subgroups, we examined the associations between time spent on reading COVID-19 news and the risk of miscarriage after adjusting for other potential risk factors. All the analyses were done with SAS software, version 9.4 (SAS Institute). Two-sided *P* values less than .05 were regarded as statistically significant.

Results

Baseline Characteristics

Among all the 456 pregnant women included, 84.2% (n=384) were younger than 35 years, 54.4% (n=248) were living in an urban area, 9.2% (n=42) had a history of miscarriage, 56.6% (n=258) were having their first pregnancy, 28.1% (n=128) were overweight or obese, 7.9% (n=36) were a previous or current smoker, 18.0% (n=82) did no physical activity, 18.0% (n=82) had inadequate dietary diversity, and 38.2% (n=174) had poor sleep quality. The mean age at baseline was 30.0 (SD 4.2) years.

Time Spent on Reading COVID-19 News and Unhealthy Lifestyles

The mean time spent on reading COVID-19 news was 1.8 (SD 0.9) hours per day. Of the 456 pregnant women, only 23 (5.0%) spent less than 0.5 hours on reading COVID-19 news per day, whereas 247 (54.2%) women spent 0.5-2 hours and 54 (11.8%) women spent >3 hours on reading COVID-19 news per day. Women with excessive media use (>3 hours) were more likely to be previously pregnant (*P*=.03), have no physical activity (*P*=.003), have inadequate dietary diversity (*P*=.03), and have poor sleep quality (*P*<.001; see Table 1).

Table 1. Baseline characteristics of pregnant women by time spent on reading COVID-19 news.

Characteristics	Total (N=456)	Time spent on reading COVID-19 news (hours), n (%)				*P* value
		<0.5 (n=23)	0.5-2 (n=247)	2-3 (n=132)	>3 (n=54)	
Age group (years)						.65
<35	384 (84.2)	19 (82.6)	211 (85.4)	107 (81.1)	47 (87.0)	
≥35	72 (15.8)	4 (17.4)	36 (14.6)	25 (18.9)	7 (13.0)	
Educational level						.29
High school or below	88 (19.3)	40 (16.2)	5 (21.7)	29 (22.0)	14 (25.9)	
College or above	368 (80.7)	207 (83.8)	18 (78.3)	103 (78.0)	40 (74.1)	
Region						.48
Rural	208 (45.6)	7 (30.4)	117 (47.4)	59 (44.7)	25 (46.3)	
Urban	248 (54.4)	16 (69.6)	130 (52.6)	73 (55.3)	29 (53.7)	
Family income, ¥5000-¥10,000 (US $764-$1528)						.11
<5000	76 (16.7)	2 (8.7)	47 (19.0)	17 (12.9)	10 (18.5)	
5000-10,000	222 (48.7)	14 (60.9)	105 (42.5)	72 (54.5)	31 (57.4)	
>10,000	158 (34.6)	7 (30.4)	95 (38.5)	43 (32.6)	13 (24.1)	
History of cesarean section						.71
No	394 (86.4)	18 (78.3)	215 (87.0)	114 (86.4)	47 (87.0)	
Yes	62 (13.6)	5 (21.7)	32 (13.0)	18 (13.6)	7 (13.0)	
History of preterm birth						.09
No	444 (97.4)	23 (100.0)	243 (98.4)	128 (97.0)	50 (92.6)	
Yes	12 (2.6)	0 (0.0)	4 (1.6)	4 (3.0)	4 (7.4)	
History of miscarriage						.32
No	414 (90.8)	19 (82.6)	229 (92.7)	118 (89.4)	48 (88.9)	
Yes	42 (9.2)	4 (17.4)	18 (7.3)	14 (10.6)	6 (11.1)	
First pregnancy						*.03*[a]
No	198 (43.4)	8 (34.8)	99 (40.1)	58 (43.9)	33 (61.1)	
Yes	258 (56.6)	15 (65.2)	148 (59.9)	74 (56.1)	21 (38.9)	
Prepregnancy BMI						.62
Underweight	58 (12.7)	4 (17.4)	34 (13.8)	13 (9.8)	7 (13.0)	
Normal weight	270 (59.2)	11 (47.8)	153 (61.9)	75 (56.8)	31 (57.4)	
Overweight	82 (18.0)	6 (26.1)	36 (14.6)	28 (21.2)	12 (22.2)	
Obese	46 (10.1)	2 (8.7)	24 (9.7)	16 (12.1)	4 (7.4)	
Smoking						.48
Nonsmoker	420 (92.1)	23 (100.0)	226 (91.5)	121 (91.7)	50 (92.6)	
Previous smoker	32 (7.0)	0 (0.0)	17 (6.9)	11 (8.3)	4 (7.4)	
Current smoker	4 (0.9)	0 (0.0)	4 (1.6)	0 (0.0)	0 (0.0)	
Physical activities						*.003*
Never	82 (18.0)	0 (0.0)	39 (15.8)	24 (18.2)	19 (35.2)	
Sometimes	230 (50.4)	10 (43.5)	124 (50.2)	73 (55.3)	23 (42.6)	
Usually	124 (27.2)	10 (43.5)	73 (29.6)	31 (23.5)	10 (18.5)	
Every day	20 (4.4)	3 (13.0)	11 (4.5)	4 (3.0)	2 (3.7)	
Inadequate dietary diversity (DDS[b] score<7)						*.03*

Characteristics	Total (N=456)	Time spent on reading COVID-19 news (hours), n (%)				*P* value
		<0.5 (n=23)	0.5-2 (n=247)	2-3 (n=132)	>3 (n=54)	
No	374 (82.0)	21 (91.3)	209 (84.6)	107 (81.1)	37 (68.5)	
Yes	82 (18.0)	2 (8.7)	38 (15.4)	25 (18.9)	17 (31.5)	
Poor sleep quality (PSQI[c] score≥5)						*<.001*
No	282 (61.8)	21 (91.3)	193 (78.1)	56 (42.4)	12 (22.2)	
Yes	174 (38.2)	2 (8.7)	54 (21.9)	76 (57.6)	42 (77.8)	

[a]Italics indicate significant *P* values.

[b]DDS: dietary diversity score.

[c]PSQI: Pittsburgh Sleep Quality Index.

Prevalence of Miscarriage by Time Spent on Reading COVID-19 News and Unhealthy Lifestyles

Out of 456 pregnant women, 73 had a miscarriage. The prevalence of miscarriage was 16.0% (95% CI 12.6%-19.4%). Compared with women who spent 0.5-2 hours (25/247, 10.1%) on reading COVID-19 news per day, miscarriage prevalence in women who spent <0.5 hours (5/23, 21.7%), 2-3 hours (26/132, 19.7%), and >3 hours (17/54, 31.5%) were higher (*P*<.001; see Figure 1). Miscarriage prevalence was also higher in pregnant women with poor sleep quality (39/174, 22.4% vs 34/282, 12.1%; *P*=.003) and a high education level (66/368, 17.9% vs 7/88, 8.0%; *P*=.02; see Table 2).

Figure 1. Comparison of miscarriage prevalence in pregnant women by time spent on reading COVID-19 news.

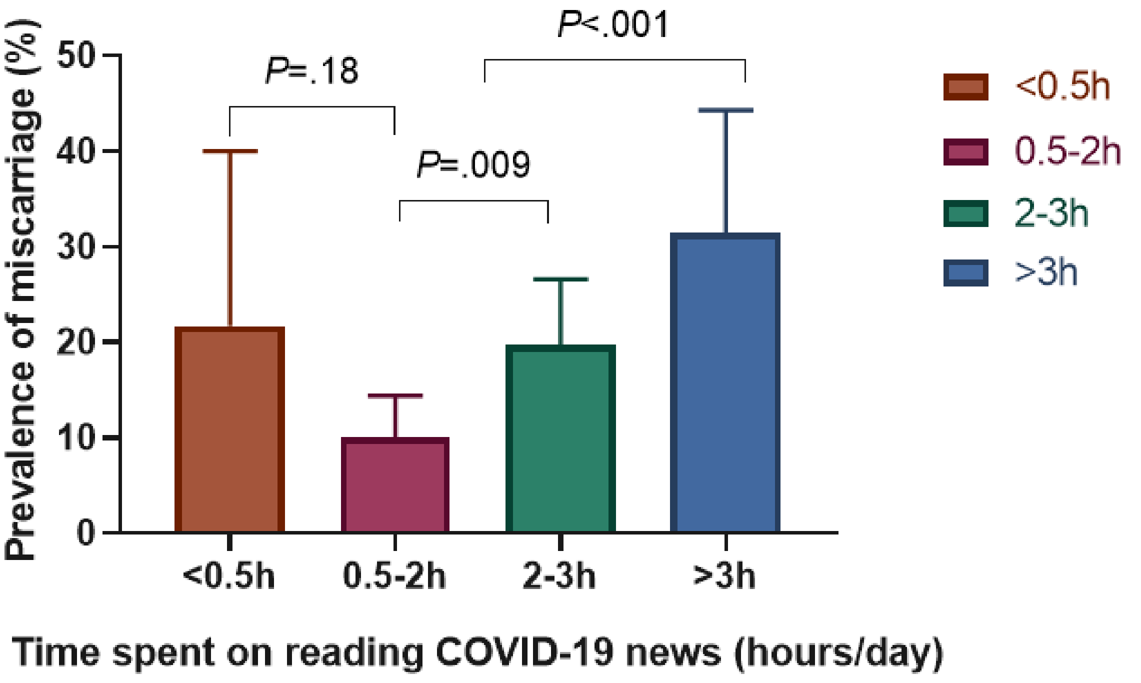


Table 2. Prevalence of miscarriage in pregnant women by different baseline characteristics.

Characteristics	Participants, n	Miscarriage prevalence, n (%)	*P* value
Total	456	73 (16.0)	N/A[a]
Age group (years)			.85
<35	384	62 (16.1)	
≥35	72	11 (15.3)	
Educational level			.02
High school or below	88	7 (8.0)	
College or above	368	66 (17.9)	
Region			.17
Rural	208	28 (13.5)	
Urban	248	45 (18.1)	
Family income, ¥5000-¥10,000 (US $764-$1528)			.06
<5000	76	7 (9.2)	
5000-10,000	222	44 (19.8)	
>10,000	158	22 (13.9)	
History of cesarean section			.73
No	394	64 (16.2)	
Yes	62	9 (14.5)	
History of preterm birth			.39
No	444	70 (15.8)	
Yes	12	3 (25.0)	
History of miscarriage			.57
No	414	65 (15.7)	
Yes	42	8 (19.0)	
First pregnancy			.14
No	198	26 (13.1)	
Yes	258	47 (18.2)	
Prepregnancy BMI			.13
Underweight	58	4 (6.9)	
Normal weight	270	51 (18.9)	
Overweight	82	12 (14.6)	
Obese	46	6 (13.0)	
Smoking			.56
Nonsmoker	420	65 (15.5)	
Previous smoker	32	7 (21.9)	
Current smoker	4	1 (25.0)	
Physical activities			.14
Never	82	19 (23.2)	
Sometimes	230	30 (13.0)	
Usually	124	22 (17.7)	
Every day	20	2 (10.0)	
Inadequate dietary diversity (DDS[b] score<7)			.17

Characteristics	Participants, n	Miscarriage prevalence, n (%)	*P* value
No	374	64 (17.1)	
Yes	82	9 (11.0)	
Poor sleep quality (PSQI[c] score>5)			.003
No	282	34 (12.1)	
Yes	174	39 (22.4)	
Time spent on reading COVID-19 news (hours)			<.001
<0.5	23	5 (21.7)	
0.5-2	247	25 (10.1)	
2-3	132	26 (19.7)	
≥3	54	17 (31.5)	

[a]N/A: not applicable.

[b]DDS: dietary diversity score.

[c]PSQI: Pittsburgh Sleep Quality Index.

Association Between Media Use, Lifestyles, and the Risk of Miscarriage

We observed a U-shape relationship between media use about COVID-19 and the risk of miscarriage (see Figure 2). In the multivariable model, poor sleep quality (adjusted RR 2.06, 95% CI 1.24-3.44; *P*=.006), 2-3 hours of media use daily (adjusted RR 1.74, 95% CI 1.02-2.97; *P*=.04), and >3 hours of media use daily (adjusted RR 2.56, 95% CI 1.43-4.59; *P*=.002) were associated with miscarriage (see Table 3).

Figure 2. The adjusted risk ratios of association between media use about COVID-19 and the risk of miscarriage by a log-binomial regression model. The 0.5-2 hours group was the reference group.

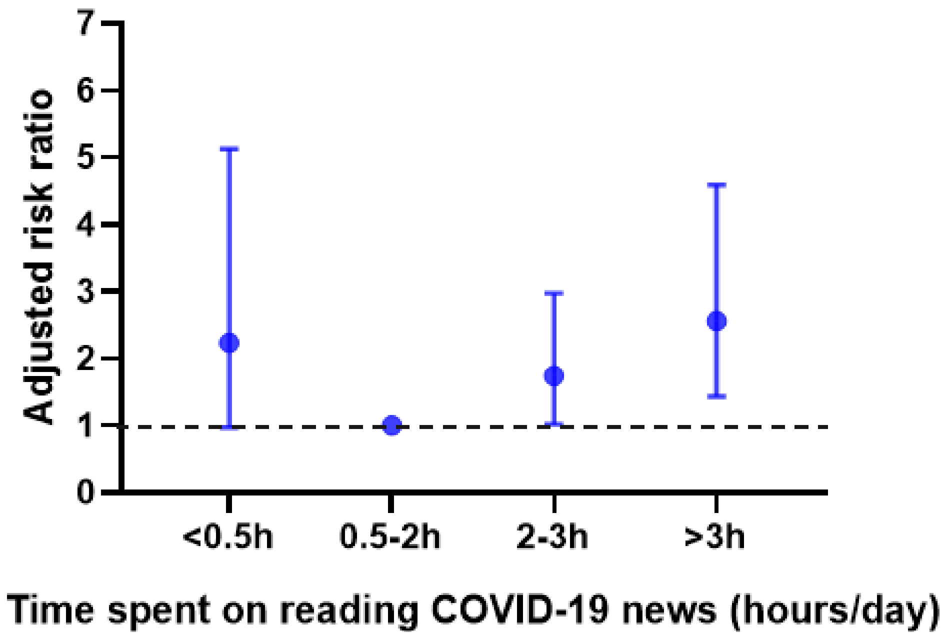


Table 3. Risk factors related with miscarriage by a log-binomial regression model.

Factors	Participants, n	Miscarriage, n (%)	Multivariable model[a]	
			Adjusted RR[b] (95% CI)	*P* value
Time spent on reading COVID-19 news (hours)				
<0.5	23	5 (21.7)	2.23 (0.97-5.13)	.06
0.5-2	247	25 (10.1)	1 (reference)	N/A[c]
2-3	132	26 (19.7)	1.74 (1.02-2.97)	.04
>3	54	17 (31.5)	2.56 (1.43-4.59)	.002
Poor sleep quality (PSQI[d] score≥5)				
No	282	34 (12.1)	1 (reference)	N/A
Yes	174	39 (22.4)	2.06 (1.24-3.44)	.006

[a]In the multivariable model, we adjusted sociodemographic characteristics (age, educational level, region, and family income), history of previous health (cesarean section, preterm birth, miscarriage, and fist pregnancy), prepregnancy BMI, smoking, physical activities, dietary diversity, and sleep quality. The covariates with *P*<.05 are shown in this table.

[b]RR: risk ratio.

[c]N/A: not applicable.

[d]PSQI: Pittsburgh Sleep Quality Index.

Sensitivity and Subgroup Analyses

In the sensitivity analysis, the association between excessive media use about COVID-19 and the risk of miscarriage was stable (see Multimedia Appendix 1). In the subgroup analysis, the risk of miscarriage was significantly higher in the <0.5 hours media use group among women living in urban areas (*P*=.04), who had a history of miscarriage (*P*=.005), and who had advanced maternal age (*P*<.001), and was significantly higher in the >3 hours media use group among women living in rural areas (*P*=.001), who had no history of miscarriage (*P*<.001), and who were younger (*P*=.004; see Multimedia Appendix 2).

Discussion

Principal Findings

To our knowledge, this is the first study exploring the status of social media use and lifestyles among pregnant women during the COVID-19 pandemic and assessing their associations with the risk of miscarriage in a prospective cohort study. Our results showed a significant association between excessive media use, unhealthy lifestyle, and the risk of miscarriage in Chinese pregnant women. No previous study has assessed the status of social media use during the COVID-19 pandemic among pregnant women. There were some studies that examined the impact of exposure to COVID-19 information on the metal health status among the nonpregnant population (eg, internet users and factory workers) [18]. Nekliudov et al [15] conducted a cross-sectional online survey in a large Russian population using multiple social media platforms and found that time spent following news on COVID-19 was strongly associated with an increased anxiety. To be specific, compared to less than 30 minutes spent reading COVID-19 news per day, the 1-2 hours group was associated with a 5.46 (95% CI 5.03-5.90) point difference, the 2-3 hours group with a 7.06 (95% CI 6.37-7.74) point difference, and the >3 hours group with an 8.65 (95% CI 7.82-9.47) point difference [15]. Pan et al [19] did a cross-sectional web-based survey of 3035 factory workers at the beginning of work resumption following the COVID-19 outbreak in Shenzhen, China. They found that higher overall information exposure to COVID-19 was associated with higher depression symptoms. Similar with the previous studies, we found that, compared with the 0.5-2 hours media use group, the risk of miscarriage was significantly higher in the 2-3 hours media use group (adjusted RR 1.74) and >3 hours media use group (adjusted RR 2.56). One possible explanation of these findings was that women who spend too much time on social media might be more likely to have unhealthy lifestyles (eg, fewer physical activities, inadequate dietary diversity, and poor sleep quality), which might be related with miscarriage. Another possible explanation of these findings was that pregnant women with exposure to the excessive media information were more likely to have psychological problems (eg, depression and anxiety), which might also increase the risk of miscarriage. Berthelot et al [20] found that women in the COVID-19 pandemic were more likely to present clinically significant levels of depression and anxiety symptoms (odds ratio 1.94) than pre–COVID-19 women in Canada. The underlying mechanism on the relationship between excessive media information and the risk of miscarriage needs to be explored in the future.

It is worth noting that a U-shape relationship between time spent on reading COVID-19 news per day and the risk of miscarriage was found in our study. The risk of miscarriage was significantly higher in the <0.5 hours media use group among women who lived in urban areas, had a history of miscarriage, and had advanced maternal age, and significantly higher in the >3 hours media use group among women who lived in rural areas, had no history of miscarriage, and were younger. Previous history of miscarriage and advanced maternal age are both risk factors for miscarriage in the literature [12]. In a multicenter European study, the risk of miscarriage was found to be higher if the woman was 35 years or older, after adjustment for various factors (eg, reproductive history and country) [12]. Inadequate information on getting essential knowledge might be the explanation for the association between inadequate media use and the risk of miscarriage. In addition, this association was more obvious in populations with other risk factors of miscarriage. Our findings highlight the importance of obtaining moderate COVID-19–related information from social media and that either inadequate (<0.5 hours) or excessive (>3 hours) exposure to COVID-19–related information was not beneficial for the individuals.

In our study, we found that the mean time spent on reading COVID-19 news was 1.8 hours per day for the pregnant women, which was slightly shorter than the general population (2.4 hours) reported in other studies [21]. The worry of radiation from phones or computers among pregnant women might be related to their relatively shorter time spent on social media. Social media has an imperative role in the world that can provide a unified platform for public health communications, comprehensive health care education guidelines, and robust social distancing strategies while still maintaining social connections [21]. Meanwhile, fake news about COVID-19 on social networks could harm public health [22]. For individuals, it is difficult to effectively identify true and false information in the mass media. Just like a "double-edged sword," social media needs to be used properly to help provide equal access to health care and end discrimination and social stigmatization. We also found that women with excessive media use were more likely to be previously pregnant, have no physical activity, have inadequate dietary diversity, and have poor sleep quality. Moreover, too much time spent on social media might also be related with unhealthy lifestyles, such as few physical activities, inadequate dietary diversity, and poor sleep. Keeping healthy lifestyles is helpful to prevent the occurrence of infectious disease and chronic disease. Our findings provide a clue for the early identification of a potentially high-risk population and miscarriage among pregnant women during the COVID-19 pandemic.

Strengths and Limitations

The prospective cohort study design, controlling various risk factors related with miscarriage, and the first insight into the association of social media use with miscarriage are the strengths of this study. However, there are several limitations in this study. First, we did not collect information on the time spent on different kinds of social media (eg, official and unofficial web-based media, newspapers, and magazines) and individual behaviors on COVID-19 prevention. Second, genetic and psychological factors associated with miscarriage were not investigated in this study. The results need to be interpreted with caution. The potential intermediation role of psychological factors on the association between time spent on reading COVID-19 news and miscarriage needs to be explored in further studies. Third, this study was a single-center cohort conducted in China. A multicenter cohort study is needed to verify the findings in this study. Despite these limitations, our findings are helpful to better understand the role of social media and lifestyle on health among pregnant women during the COVID-19 pandemic. The role of media and public health communications needs to be correctly understood and explored further, as they will be an essential tool for delivering information and combating COVID-19 and health promotion, especially for vulnerable populations such as pregnant women.

Conclusions

Pregnant women spent about 2 hours a day reading COVID-19 news in the early stage of the COVID-19 pandemic in China. Pregnant women with excessive media use were more likely to having no physical activity, inadequate dietary diversity, and poor sleep quality. Excessive media use and poor sleep quality were associated with a higher risk of miscarriage. Our findings highlight the importance of healthy lifestyles during the COVID-19 pandemic.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant number 81703240), the National Science and Technology Key Projects on Prevention and Treatment of Major Infectious Diseases of China (grant number 2020ZX10001002), and the National Key Research and Development Project of China (grant number 2019YFC1710301; 2020YFC0846300). We sincerely thank the staff of Tongzhou Maternal and Child Health Hospital for their efforts made in data collection.

Authors' Contributions

This study was designed by XZ, NH, and JL. NH, JY, and JL coordinated the acquisition of data. XZ analyzed the data and drafted the manuscript with input from NH, JY, and JL. NH, JY, and JL reviewed and revised the report. All authors gave approval for the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental Table: Sensitivity analysis on the association between media use about COVID-19 and the risk of miscarriage.
[DOCX File , 16 KB - publichealth_v7i1e25241_app1.docx]

Multimedia Appendix 2
Supplemental Figure: Subgroup analysis of the association between media use about COVID-19 and miscarriage.
[PNG File , 150 KB - publichealth_v7i1e25241_app2.png]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-11-14]
2. Reed C, Biggerstaff M, Finelli L, Koonin LM, Beauvais D, Uzicanin A, et al. Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. Emerg Infect Dis 2013 Jan;19(1):85-91. [doi: 10.3201/eid1901.120124] [Medline: 23260039]
3. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020 Jun;20(6):669-677 [FREE Full text] [doi: 10.1016/S1473-3099(20)30243-7] [Medline: 32240634]
4. Liu J, Liu M, Wan S. Progress on the basic reproduction number of SARS-CoV-2. Chin Sci Bull 2020 Jun 4;65(22):2334-2341. [doi: 10.1360/tb-2020-0413]
5. Li Z, Chen Q, Feng L, Rodewald L, Xia Y, Yu H, China CDC COVID-19 Emergency Response Strategy Team. Active case finding with case management: the key to tackling the COVID-19 pandemic. Lancet 2020 Jul 04;396(10243):63-70 [FREE Full text] [doi: 10.1016/S0140-6736(20)31278-2] [Medline: 32505220]
6. Du M, Lin Y, Yan W, Tao L, Liu M, Liu J. Prevalence and impact of diabetes in patients with COVID-19 in China. World J Diabetes 2020 Oct 15;11(10):468-480 [FREE Full text] [doi: 10.4239/wjd.v11.i10.468] [Medline: 33133394]
7. Luo L, Fu M, Li Y, Hu S, Luo J, Chen Z, et al. The potential association between common comorbidities and severity and mortality of coronavirus disease 2019: a pooled analysis. Clin Cardiol 2020 Dec;43(12):1478-1493. [doi: 10.1002/clc.23465] [Medline: 33026120]
8. Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL. Perceptions, knowledge, and behaviors related to COVID-19 among social media users: cross-sectional study. J Med Internet Res 2020 Sep 08;22(9):e19913 [FREE Full text] [doi: 10.2196/19913] [Medline: 32841153]
9. Anwar A, Malik M, Raees V, Anwar A. Role of mass media and public health communications in the COVID-19 pandemic. Cureus 2020 Sep 14;12(9):e10453 [FREE Full text] [doi: 10.7759/cureus.10453] [Medline: 33072461]
10. ESHRE Guideline Group on RPL, Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, et al. ESHRE guideline: recurrent pregnancy loss. Hum Reprod Open 2018;2018(2):hoy004 [FREE Full text] [doi: 10.1093/hropen/hoy004] [Medline: 31486805]
11. Freeman A, Neiterman E, Varathasundaram S. Women's experiences of health care utilization in cases of early pregnancy loss: a scoping review. Women Birth 2020 Aug 25. [doi: 10.1016/j.wombi.2020.07.012] [Medline: 32859562]
12. de la Rochebrochard E, Thonneau P. Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study. Hum Reprod 2002 Jun;17(6):1649-1656. [doi: 10.1093/humrep/17.6.1649] [Medline: 12042293]
13. Larsen EC, Christiansen OB, Kolte AM, Macklon N. New insights into mechanisms behind miscarriage. BMC Med 2013 Jun 26;11:154 [FREE Full text] [doi: 10.1186/1741-7015-11-154] [Medline: 23803387]
14. Gao L, Qu J, Wang AY. Anxiety, depression and social support in pregnant women with a history of recurrent miscarriage: a cross-sectional study. J Reprod Infant Psychol 2020 Nov;38(5):497-508. [doi: 10.1080/02646838.2019.1652730] [Medline: 31411054]
15. Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, et al. Excessive media consumption about COVID-19 is associated with increased state anxiety: outcomes of a large online survey in Russia. J Med Internet Res 2020 Sep 11;22(9):e20955 [FREE Full text] [doi: 10.2196/20955] [Medline: 32788143]
16. Tao L, Xie Z, Huang T. Dietary diversity and all-cause mortality among Chinese adults aged 65 or older: a community-based cohort study. Asia Pac J Clin Nutr 2020;29(1):152-160 [FREE Full text] [doi: 10.6133/apjcn.202003_29(1).0020] [Medline: 32229454]
17. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
18. Sedov ID, Cameron EE, Madigan S, Tomfohr-Madsen LM. Sleep quality during pregnancy: a meta-analysis. Sleep Med Rev 2018 Apr;38:168-176. [doi: 10.1016/j.smrv.2017.06.005] [Medline: 28866020]
19. Pan Y, Xin M, Zhang C, Dong W, Fang Y, Wu W, et al. Associations of mental health and personal preventive measure compliance with exposure to COVID-19 information during work resumption following the COVID-19 outbreak in China: cross-sectional survey study. J Med Internet Res 2020 Oct 08;22(10):e22596 [FREE Full text] [doi: 10.2196/22596] [Medline: 32936776]
20. Berthelot N, Lemieux R, Garon-Bissonnette J, Drouin-Maziade C, Martel É, Maziade M. Uptrend in distress and psychiatric symptomatology in pregnant women during the coronavirus disease 2019 pandemic. Acta Obstet Gynecol Scand 2020 Jul;99(7):848-855. [doi: 10.1111/aogs.13925] [Medline: 32449178]

21. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
22. Galhardi C, Freire N, Minayo M, Fagundes M. Fact or fake? An analysis of disinformation regarding the Covid-19 pandemic in Brazil. Cien Saude Colet 2020 Oct;25(suppl 2):4201-4210 [FREE Full text] [doi: 10.1590/1413-812320202510.2.28922020] [Medline: 33027357]

Abbreviations

DDS: dietary diversity score
PSQI: Pittsburgh Sleep Quality Index
RR: risk ratio

Edited by G Eysenbach; submitted 24.10.20; peer-reviewed by T Liyuan; comments to author 13.11.20; revised version received 13.11.20; accepted 08.12.20; published 05.01.21.

Please cite as:
Zhang X, Liu J, Han N, Yin J
Social Media Use, Unhealthy Lifestyles, and the Risk of Miscarriage Among Pregnant Women During the COVID-19 Pandemic: Prospective Observational Study
JMIR Public Health Surveill 2021;7(1):e25241
URL: https://publichealth.jmir.org/2021/1/e25241
doi:10.2196/25241
PMID:33293263

Original Paper

Implementation of Telemedicine in a Tertiary Hospital–Based Ambulatory Practice in Detroit During the COVID-19 Pandemic: Observational Study

Alpana Garg[1], MD; Sachin Goyal[1], MD; Rohit Thati[2], MSc; Neelima Thati[1], MD

[1]Wayne State University, Detroit, MI, United States

[2]Georgia State University, Atlanta, GA, United States

Corresponding Author:
Alpana Garg, MD
Wayne State University
University Health Center, 4201 St. Antoine
Detroit, MI, 48201
United States
Phone: 1 708 501 2938
Email: alpanagarg24@yahoo.com

Abstract

Background: The COVID-19 pandemic, caused by SARS-CoV-2, has forced the health care delivery structure to change rapidly. The pandemic has further widened the disparities in health care and exposed vulnerable populations. Health care services caring for such populations must not only continue to operate but create innovative methods of care delivery without compromising safety. We present our experience of incorporating telemedicine in our university hospital–based outpatient clinic in one of the worst-hit areas in the world.

Objective: Our goal is to assess the adoption of a telemedicine service in the first month of its implementation in outpatient practice during the COVID-19 pandemic. We also want to assess the need for transitioning to telemedicine, the benefits and challenges in doing so, and ongoing solutions during the initial phase of the implementation of telemedicine services for our patients.

Methods: We conducted a prospective review of clinic operations data from the first month of a telemedicine rollout in the outpatient adult ambulatory clinic from April 1, 2020, to April 30, 2020. A telemedicine visit was defined as synchronous audio-video communication between the provider and patient for clinical care longer than 5 minutes or if the video visit converted to a telephone visit after 5 minutes due to technical problems. We recorded the number of telemedicine visits scheduled, visits completed, and the time for each visit. We also noted the most frequent billing codes used based on the time spent in the patient care and the number of clinical tasks (eg, activity suggested through diagnosis or procedural code) that were addressed remotely by the physicians.

Results: During the study period, we had 110 telemedicine visits scheduled, of which 94 (85.4%) visits were completed. The average duration of the video visit was 35 minutes, with the most prolonged visit lasting 120 minutes. Of 94 patients, 24 (25.54%) patients were recently discharged from the hospital, and 70 (74.46%) patients were seen for urgent care needs. There was a 50% increase from the baseline in the number of clinical tasks that were addressed by the physicians during the pandemic.

Conclusions: There was a high acceptance of telemedicine services by the patients, which was evident by a high show rate during the COVID-19 pandemic in Detroit. With limited staffing, restricted outpatient work hours, a shortage of providers, and increased outpatient needs, telemedicine was successfully implemented in our practice.

(JMIR Public Health Surveill 2021;7(1):e21327) doi:10.2196/21327

KEYWORDS

telemedicine; telehealth; COVID-19; Detroit; ambulatory care; primary care; internal medicine; pandemic

Introduction

Background

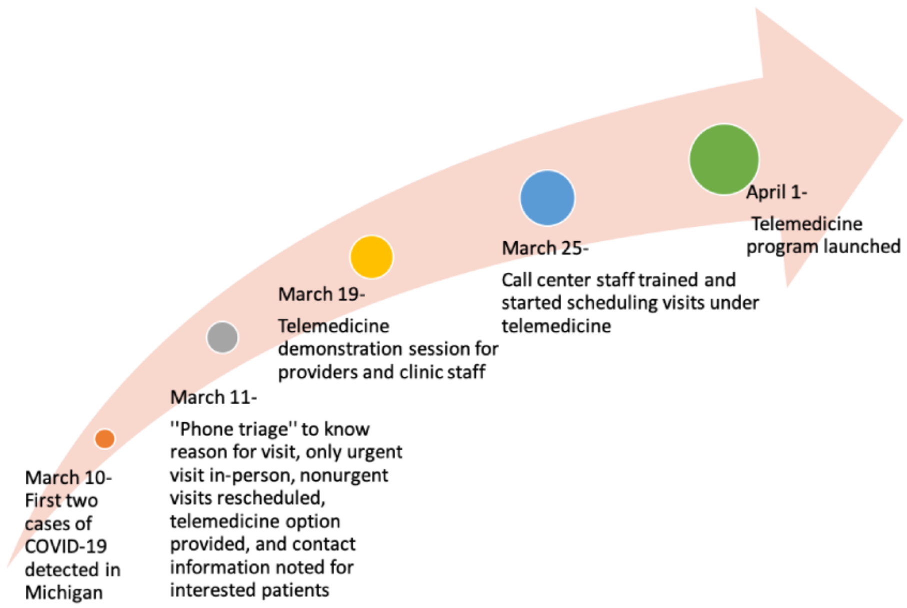
The COVID-19 pandemic has impacted daily life and led to a rapid evolution in the structure of health care delivery. As of November 19, 2020, more than 56 million people have been infected, and 1.35 million people have died worldwide [1]. The World Health Organization declared the COVID-19 outbreak a pandemic on March 11, 2020—near the time of the first confirmed case of COVID-19 in Michigan, which became one of the major epicenters of the pandemic in the United States [2,3]. In the city of Detroit, Michigan, at the peak of the pandemic when the health care system was dealing with an unprecedented surge, outpatient clinics were taking quick and decisive steps to ensure continuity of care while keeping both patients and health care workers safe [4]. As a tertiary care institute located in the heart of Detroit, our institution cares for a predominantly African American population. Michigan is one of the ten states where prevalence of multiple chronic conditions (defined as two or more of ten diagnosed chronic conditions) is higher (30.3%) than the national average (25.7%) [5]. In addition, economic challenges and socioeconomic disparities are reflected by a high unemployment rate, high poverty rate, low health insurance coverage, and a lack of transportation, which represent significant barriers to optimal medical care [6-8]. These factors make our patient population unique and vulnerable, as evident by the higher case-fatality rate of COVID-19 in Michigan compared to other parts of the country [3,9-11].

Telemedicine involves clinical care using an electronic communication medium between two different locations [12]. Even though telemedicine existed before the pandemic, its use in health care delivery has increased tremendously during the COVID-19 pandemic [13,14]. Various health systems have already adopted remote medical care and integrated telemedicine to assist in providing care to the patients [15-18]. For our purposes, adoption is the intention, initial decision, or action to try or employ an innovation or evidence-based practice [19]. We present our experience with telemedicine in our university hospital–based ambulatory care practice during the initial phase of the COVID-19 public health emergency (PHE).

Need for a Transition From the Outpatient Clinic to Telemedicine Clinic Model

Soon after the governor declared a state of emergency in Michigan, there was a sharp decline in the number of patients attending in-person visits [20]. In addition, health care providers were specially assigned tasks to "phone triage," which meant calling all patients scheduled for an in-person visit to ask about the reason for their visits. Providers reviewed electronic medical records (EMRs), and if possible, patients' concerns, such as medication refills or generating return to work letters, were addressed remotely. Otherwise, patients kept their in-person appointments with their health care providers. All nonurgent in-person visits were rescheduled, and patients were encouraged to stay at home to comply with safety guidelines issued by the Centers for Disease Control and Prevention [21]. With the immense reduction in the number of in-office patient visits, the number of medical assistants and clinical staff had to be reduced. Providers were being assigned to inpatient units to assist in care for the critically ill. The patient population at risk of infection had distinct needs (eg, necessary medications and medical supplies) that were in jeopardy of being unmet, which was a significant concern during the initial days of the stay-at-home guidelines. Telephone use by providers to answer patient questions and help with medication refills was done to the best that such resources would allow. In addition, certain patients were discharged from the inpatient units sooner than would be typical to decrease the risk of COVID-19 spread, and those patients needed a timely postdischarge follow-up. This caused an urgent need for a telemedicine service in our outpatient clinic. Patients were given the option of video visits, and the contact information of those patients interested in the video visits was noted. Figure 1 shows the timeline of telemedicine implementation in our practice.

Figure 1. Ambulatory telemedicine clinic timeline.



Transition From In-Person Clinic Structure to Telemedicine Clinic

Clinic Structure

The clinic structure before and during the COVID-19 pandemic for one half-day session is shown in Textbox 1. Before the pandemic, our ambulatory clinic was staffed with an attending physician, and all patients were seen initially by a resident physician. We had no midlevel providers in our clinic. The medical assistant and front desk staff had dedicated duties regarding patient intake and checkout, while the clinic manager took care of the overall functioning of the clinic, as established before the pandemic. In the new telemedicine setup, the medical assistant would call the patient before the telemedicine appointment to obtain additional information such as email address, confirming availability of a webcam device, the reason for visit, medication reconciliation, and pharmacy information.

Textbox 1. Clinic structure before and during the COVID-19 pandemic.

Before COVID-19

- 8-12 resident physicians per session (half day clinic)
- 10 sessions per week
- 3 faculty members per week
- 4 medical assistants
- 2 nurses
- 2 registration staff members
- 1 clinic manager
- Fully functional registration and appointment center

During COVID-19

- 4 resident physicians per week
- 3-4 sessions per week
- 2 faculty members per week
- 1 medical assistant
- 1 technical support staff member
- 1 registration staff member
- 1 clinic manager
- Reduced staff at the registration and appointment center

Patient Scheduling

The initial patients willing to conduct a video visit were contacted as the system was being established. Additionally, all the patients who were discharged from our affiliated hospitals were instructed to follow up in the telemedicine clinic. The telemedicine providers were informed about the discharged patients via a secure email platform who would then arrange the telemedicine visit with the help of ancillary staff to ensure timely follow-up for patients. Some of these patients also included those with suspected or confirmed COVID-19.

Telemedicine Platform

We adopted Zoom (Zoom Video Communications Inc) and Google Hangouts Meet (Google) platforms to conduct telemedicine visits. To use these platforms for patient care, our organization entered into business associate agreements with these platforms to secure Health Insurance Portability and Accountability Act compliance and made institutional accounts for the involved physicians. This also prevented physicians from using their personal accounts for patient care to mitigate concerns regarding the sharing of personal information with patients. We circulated detailed instructions for the providers, and on-site technical staff helped with technical issues to streamline the process. A support helpline telephone number was made available for queries too. The clinic staff, including clinic manager, medical assistants, and checkout staff, were also prepared to schedule visits and helped with calling patients before their scheduled telemedicine visits.

Health Care Providers and Our Telemedicine Process

Physicians who had chronic medical conditions and those who were pregnant or were quarantined (due to potential exposure) but otherwise healthy were assigned to telemedicine service in addition to regular primary care physicians. All physicians were required to undergo training by completing telemedicine modules available online [22]. Providers could see their telemedicine schedule in the EMR platform and would call the patients at the appointment time either from the clinic or from the provider's home. The resident physician and the attending physician would both join the virtual meeting, and after introductions, the resident physician was given time to assess the patient. After the resident physician assessed the patient, the patient was put on hold to discuss the assessment and plan with the attending physician. The plan was then discussed with the patient, and relevant orders for investigation and follow-up were placed in the EMR. Medical students were also able to participate in video visits. The students assigned telemedicine rotation were sent meeting links for the day so they could join the visits. To ensure adequate communication about the workflow and expectations, weekly virtual meetings with resident physicians and attending faculty were conducted.

Methods

Prospective data were collected from the first month of telemedicine in the adult ambulatory clinic of Detroit Medical Center, Wayne State University, Detroit, Michigan. We defined a telemedicine visit as the synchronous audio-video communication between provider and patient for clinical care for 5 minutes or more. A telemedicine visit during which the call was started as an audio-video call but was switched to a telephone call after 5 minutes or more due to a technical problem was also considered a completed visit. We collected data including the number of telemedicine visits scheduled, the number of visits completed, and the most frequent billing codes used. Information about completed visits and visit durations was obtained from the provider. However, due to emergent circumstances, patient-specific information such as demographics and any protected health information was not obtained.

In our ambulatory clinic, patients were able to contact the call center to leave a message or a task for the provider (including requests directly from pharmacies for medication refill), which were directed to the internal medicine task inbox. This generated task was completed by the provider by contacting the patient via telephone or electronically, as required. The average number of tasks in the general medicine inbox were noted during the study period. Baseline information, including the number of patients scheduled per day, show rate, and the average number of tasks in the inbox for providers before the pandemic, was also obtained from the clinic manager. The numbers were recorded on Microsoft Excel spreadsheet software (Microsoft Corporation). Percentages, medians, and averages were calculated. Data were analyzed, and graphs (including box plots) were made using Python programming language software [23].

Results

After the initial setup, 110 patients were scheduled for telemedicine visits over 4 weeks from April 1, 2020, to April 30, 2020. Of 110 visits, 16 visits were not completed (defined in this study as when the provider was unable to contact the patient at all or the visit was shorter than 5 minutes). The total number of completed telemedicine visits distributed over 1 month is shown in Figure 2. There was no specific time pattern observed, as telemedicine scheduled visits were variable. The average length of the visit was 35 minutes (range 5-120 minutes; Figure 3). Technical challenges were encountered, such as difficulty in establishing the video call with the patient or disconnection within a few minutes. These calls were billed according to the time spent in actual patient care, excluding the time in technical challenges; calls shorter than 5 minutes were excluded. Figure 4 presents the most frequently used billing codes based on the time spent in patient care, new patient or established patient visits, and the medium (telephone or video visit). Based on visit type, 25.5% (24/94) of patients who completed telemedicine visits were recently discharged from the hospital, and 74.5% (70/94) were scheduled for telemedicine visits based on urgent needs. Most of the telemedicine visits scheduled (94/110, 85.4%) were completed. At baseline, before the pandemic, 60-70 patients were scheduled per day with a 70% show rate.

Figure 2. Number of patients with completed telemedicine visits over 4 weeks.

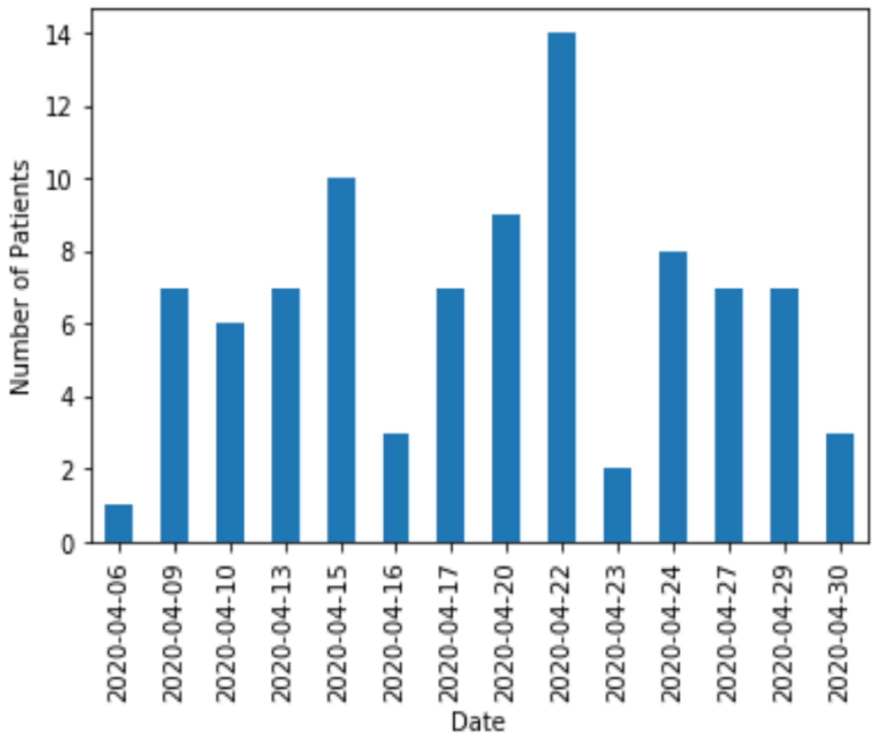


Figure 3. Box plot depicting duration of visits. Outliers are the values >80 minutes, and the average value is approximately 35 minutes.

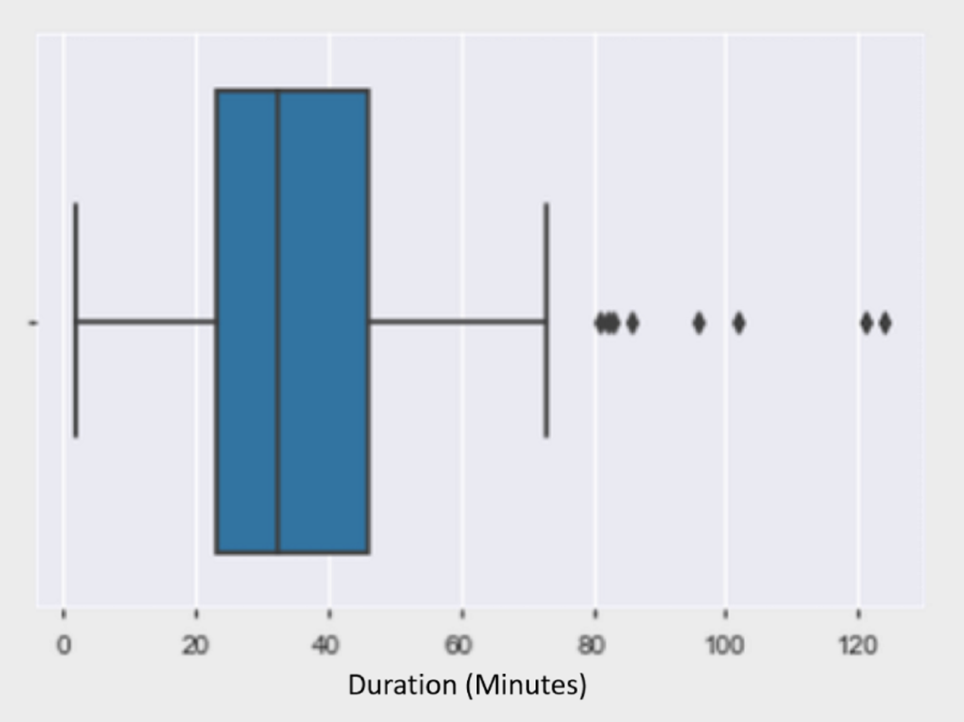
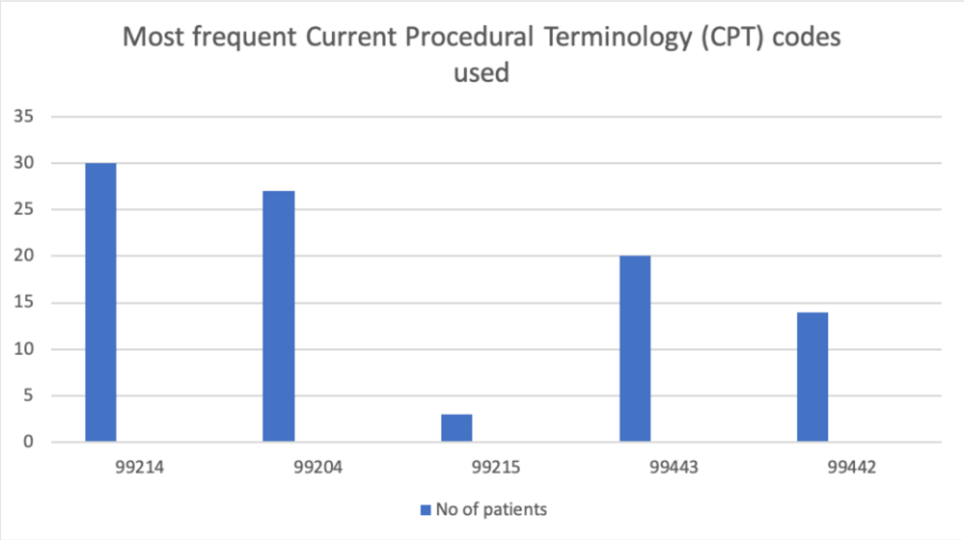


Figure 4. Histogram showing most common billing codes used. 99214 is an established patient office visit for 25 minutes (used with telemedicine code modifier); 99204 is a new patient office visit for 45 minutes (used with telemedicine code modifier); 99215 is an established patient office visit for 40 minutes (used with telemedicine code modifier); 99443 is a telephone evaluation and management for 21-30 minutes; 99442 is a telephone evaluation and management for 11-20 minutes.



On average, 300 tasks were addressed per week by a dedicated team of resident physicians and support staff of the clinic, and included medication refills, generating referrals, producing work letters, answering queries related to COVID-19–like symptoms, and triaging. The tasks were addressed electronically or, if necessary, by calling the patient. Before the pandemic, the average number of tasks in the general medicine inbox was 200 tasks per week, which means there was a 50% increase in the number of tasks during the pandemic.

Discussion

Principal Findings

The adoption of telemedicine in our ambulatory practice to adjust to the unique circumstances during the COVID-19 pandemic was successful. Although the number of telemedicine visits was less than compared to in-person visits before the pandemic, the patients who had telemedicine scheduled had a high show rate of 85.4% (94/110) compared to the 70% show rate for our traditional ambulatory visits. Various factors like calling patients 2-3 days before the telemedicine visit, posthospital discharge sign-out between the providers, no requirement of transportation, the in-home convenience of a telemedicine visit, elevated levels of patient concern related to COVID-19, and reduced availability of in-person visits may have contributed to the high show rates for telemedicine. The advantages of telemedicine that we experienced are shown in Textbox 2.

Textbox 2. Benefits of telemedicine and telephone communications during the pandemic.

Continued clinical care for patients

- Follow up on chronic medical conditions like hypertension, diabetes, asthma, congestive heart failure, and medication adjustment, and ensuring adequate number of refills

Management of patients with possible SARS-CoV-2 infection

- Patients with mild symptoms of COVID-19 were assessed remotely and given home quarantine instructions, testing information, and symptomatic treatment.
- Patients at home were seen through a video visit, preventing unnecessary emergency department visits with the potential to spread the disease and use critical resources.

Decreasing hospital admission and length of stay

- Patients who were hospitalized for conditions not related to COVID-19 and were discharged from the hospitals required a closer follow-up with primary care physicians.
- Patients with SARS-CoV-2 infection who were deemed stable for discharge but still required a follow-up after discharge (based on continued oxygen requirements or other complications) were closely monitored via a follow-up in the clinic.

Arranging durable medical equipment

- Increased number of requests (due to increase at-home monitoring for medical conditions) for medical equipment such as blood pressure sets, glucometers, nebulizers, and supplies for home oxygen, which were arranged in a timely manner

Involvement of medical students and continued resident teaching and training

- Medical students who were required to stay home during the pandemic with suspended clinical duties were still involved during the telemedicine sessions.

In the first several visits, there were technical struggles in setting up the video meeting with the patient—one visit took a total time of 120 minutes. However, we then implemented a protocol where scheduled patients were called by staff before the video visit to receive instructions to save time during the actual visit with the provider. The patients were given instructions 2-3 days before their scheduled session to allow time for downloading the applications and reviewing step-by-step instructions, which streamlined the process of setting up the video calls. If requested by the patient, written instructions were also emailed to patients. Patients and family members were informed of this secure, private platform, and they were assured that video meetings would not be recorded. Having the clinic staff call the patient before their appointment helped the providers follow the schedule and stay on time.

Through our telemedicine platform, we could follow up with patients discharged from the hospital. Discharge follow-up monitoring is important for preventing readmissions for conditions like heart failure; however, during the pandemic, it was essential to follow up with patients with confirmed or suspected COVID-19 to monitor for complications and emphasize quarantine and precautionary measures [24,25].

Weekly meetings were held with resident physicians to keep them abreast of ongoing changes and answer any specific queries. During the weekly meetings, we were able to address the challenges like limited staff available to call patients before the video visit. In addition, it was challenging for the inpatient providers to proactively set outpatient appointments through email, as initially, there was no automated system available. However, with time, staffing issues were resolved, and case managers assisted with hospital discharge follow-up telemedicine appointments. Call center staff was also urged to prioritize hospital discharge follow-up. Textbox 3 presents the challenges encountered, along with several ongoing solutions.

Textbox 3. Challenges during the first month of telemedicine with ongoing solutions.

- Timely arranging for secure platform application and setting up secure links to maintain confidentiality and security for connecting patients with providers and clinic staff
  - Organization created work accounts for providers and clinic staff including call center staff.
  - Training sessions for providers and clinic staff
- Reaching out to patients to provide information about telemedicine and patient enrollment to the clinic for telemedicine, especially hospital discharge follow-up appointments
  - Call center staff available to give options and schedule video visits for interested patients
  - For hospital discharge follow-up monitoring, telemedicine order set available in electronic medical record to alert clinic staff
- Video visit mandates the presence of a smartphone or a device with a camera
  - Long-term solutions to decrease economic disparity
  - Assistance from family members (who can help in setting up video call) when possible
- Technical assistance required for the patients, providers, and clinic staff to set up a video call
  - Patient contacted a few days before the visit by clinic staff to provide verbal and written step-by-step instructions to set up a video call before appointment
  - Dedicated person to provide technical assistance to the providers and clinic staff
- Training of the resident physicians and increasing their comfort in the video visit
  - Mandating completion of telemedicine modules before starting clinic
  - Obtaining feedback from the residents to assess challenges encountered in the video visit and addressing them
- Establish a curriculum and ensure the teaching of resident physicians
  - Dedicated teaching time to ensure learning
  - Weekly ambulatory lecture including topics pertinent to telemedicine (eg, how to conduct physical exam during video visit)
- Streamline warm handoff from physician and checkout process after video visit, including methods to arrange laboratory services, referrals, and a follow-up appointment
  - Dedicated clinic staff to contact the patient after video visit to explain the process to schedule laboratory tests, referrals, and follow-up appointments (to close the loop)
  - Provider contacts patient to discuss laboratory results

Limitations

This is a single institutional observational study, and hence, findings are not broadly generalizable. During the initial setup amid rapid changes, we were not able to obtain patient demographic information. Therefore, the adoption of telemedicine by various subsets of patients is unknown. We cannot determine how gender, race, ethnicity, age group, and socioeconomic disparity will affect a patient's likelihood to schedule telemedicine visits or the impact on their ability to complete such a visit. Finally, although we ensured completion of the mandatory online telemedicine learning modules, providers learned through a rapid, on-the-job "crash course" amid a PHE, which may have impacted our results such as time required for a telemedicine visit.

Conclusions

There was a high acceptance of telemedicine services by patients, which was evident by the high show rate during the COVID-19 pandemic in Detroit. With limited staffing, restricted outpatient work hours, a shortage of providers, and increased outpatient needs, telemedicine was successfully implemented in our practice. Quality improvement projects are needed to improve the telemedicine visit experience for both patients and providers.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-11-19]
2. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
3. Coronavirus Michigan data. Michigan.gov. 2020. URL: https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html [accessed 2020-06-07]
4. Chopra T, Sobel J. Detroit under siege, the enemy within: the impact of the COVID-19 collision. Infect Control Hosp Epidemiol 2020 Sep;41(9):1122 [FREE Full text] [doi: 10.1017/ice.2020.154] [Medline: 32312338]
5. Ward BW, Black LI. State and regional prevalence of diagnosed multiple chronic conditions among adults aged ≥18 years - United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Jul 29;65(29):735-738. [doi: 10.15585/mmwr.mm6529a3] [Medline: 27467707]
6. Critical health indicators. Michigan Department of Health and Human Services. 2012 Apr. URL: https://www.michigan.gov/mdhhs/0,5885,7-339-73970_2944_5327-17501--,00.html [accessed 2020-08-26]
7. Michigan report - 2019. Talk Poverty. URL: https://talkpoverty.org/state-year-report/michigan-2019-report [accessed 2020-11-19]
8. Artiga S, Hinton E. Beyond health care: the role of social determinants in promoting health and health equity. KFF. 2018. URL: https://www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/ [accessed 2020-11-19]
9. CDC COVID data tracker. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2020-11-19]
10. Suleyman G, Fadel RA, Malette KM, Hammond C, Abdulla H, Entz A, et al. Clinical characteristics and morbidity associated with coronavirus disease 2019 in a series of patients in Metropolitan Detroit. JAMA Netw Open 2020 Jun 01;3(6):e2012270 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12270] [Medline: 32543702]
11. Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of May 2020. J Rural Health 2020 Jun;36(3):433-445 [FREE Full text] [doi: 10.1111/jrh.12476] [Medline: 32543763]
12. Medicare telemedicine health care provider fact sheet. Centers for Medicare and Medicaid Services. 2020. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2020-06-07]
13. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
14. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
15. Vilendrer S, Patel B, Chadwick W, Hwa M, Asch S, Pageler N, et al. Rapid deployment of inpatient telemedicine in response to COVID-19 across three health systems. J Am Med Inform Assoc 2020 Jul 01;27(7):1102-1109 [FREE Full text] [doi: 10.1093/jamia/ocaa077] [Medline: 32495830]
16. Basu S, Phillips RS, Phillips R, Peterson LE, Landon BE. Primary care practice finances in the United States amid the COVID-19 pandemic. Health Aff (Millwood) 2020 Sep;39(9):1605-1614. [doi: 10.1377/hlthaff.2020.00794] [Medline: 32584605]
17. Punia V, Nasr G, Zagorski V, Lawrence G, Fesler J, Nair D, et al. Evidence of a rapid shift in outpatient practice during the COVID-19 pandemic using telemedicine. Telemed J E Health 2020 Oct;26(10):1301-1303. [doi: 10.1089/tmj.2020.0150] [Medline: 32429769]
18. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and preliminary outcomes of remote diagnosis and treatment during the COVID-19 outbreak: retrospective cohort study. JMIR mHealth uHealth 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: 10.2196/19417] [Medline: 32568722]
19. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
20. Executive Order 2020-04: declaration of state of emergency. Michigan.gov. 2020. URL: https://www.michigan.gov/whitmer/0,9309,7-387-90499_90705-521576--,00.html [accessed 2020-11-19]
21. COVID-19. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2020-11-19]
22. Telehealth guidance and resources. American College of Physicians. 2020. URL: https://www.acponline.org/practice-resources/business-resources/health-information-technology/telehealth?gclid=EAIaIQobChMIgpDK16_L6QIVCr3ACh1xdwHYEAAYASAAEgJ2HPD_BwE [accessed 2020-06-07]
23. Python. 2001. URL: https://www.python.org/ [accessed 2020-08-26]
24. Zheng Z, Yao Z, Wu K, Zheng J. Patient follow-up after discharge after COVID-19 pneumonia: considerations for infectious control. J Med Virol 2020 Nov;92(11):2412-2419 [FREE Full text] [doi: 10.1002/jmv.25994] [Medline: 32383776]

25. George PM, Barratt SL, Condliffe R, Desai SR, Devaraj A, Forrest I, et al. Respiratory follow-up of patients with COVID-19 pneumonia. Thorax 2020 Nov;75(11):1009-1016 [FREE Full text] [doi: 10.1136/thoraxjnl-2020-215314] [Medline: 32839287]

Abbreviations

EMR: electronic medical record
PHE: public health emergency

Edited by G Eysenbach, R Kukafka; submitted 11.06.20; peer-reviewed by S Vilendrer, T Aslanidis; comments to author 20.07.20; revised version received 04.09.20; accepted 12.12.20; published 08.01.21.

Please cite as:
Garg A, Goyal S, Thati R, Thati N
Implementation of Telemedicine in a Tertiary Hospital–Based Ambulatory Practice in Detroit During the COVID-19 Pandemic: Observational Study
JMIR Public Health Surveill 2021;7(1):e21327
URL: http://publichealth.jmir.org/2021/1/e21327/
doi:10.2196/21327
PMID:33400680

Original Paper

Jobs, Housing, and Mask Wearing: Cross-Sectional Study of Risk Factors for COVID-19

Eline M van den Broek-Altenburg[1], MA, MSc, PhD; Adam J Atherly[2]; Sean A Diehl[3], PhD; Kelsey M Gleason[2], DrSc; Victoria C Hart[2], PhD; Charles D MacLean[2], MD; Daniel A Barkhuff[4], MD; Mark A Levine[2], MD; Jan K Carney[2], MD

[1]Department of Radiology, Larner College of Medicine, University of Vermont, Burlington, VT, United States
[2]Department of Medicine, Larner College of Medicine, University of Vermont, Burlington, VT, United States
[3]Department of Microbiology and Molecular Genetics, Larner College of Medicine, University of Vermont, Burlington, VT, United States
[4]Division of Emergency Medicine, Larner College of Medicine, University of Vermont, Burlington, VT, United States

Corresponding Author:
Eline M van den Broek-Altenburg, MA, MSc, PhD
Department of Radiology
Larner College of Medicine
University of Vermont
89 Beaumont Avenue
Burlington, VT, 05405
United States
Phone: 1 8024613951
Email: eline.altenburg@med.uvm.edu

Abstract

Background: Many studies have focused on the characteristics of symptomatic patients with COVID-19 and clinical risk factors. This study reports the prevalence of COVID-19 in an asymptomatic population of a hospital service area (HSA) and identifies factors that affect exposure to the virus.

Objective: The aim of this study is to measure the prevalence of COVID-19 in an HSA, identify factors that may increase or decrease the risk of infection, and analyze factors that increase the number of daily contacts.

Methods: This study surveyed 1694 patients between April 30 and May 13, 2020, about their work and living situations, income, behavior, sociodemographic characteristics, and prepandemic health characteristics. This data was linked to testing data for 454 of these patients, including polymerase chain reaction test results and two different serologic assays. Positivity rate was used to calculate approximate prevalence, hospitalization rate, and infection fatality rate (IFR). Survey data was used to analyze risk factors, including the number of contacts reported by study participants. The data was also used to identify factors increasing the number of daily contacts, such as mask wearing and living environment.

Results: We found a positivity rate of 2.2%, a hospitalization rate of 1.2%, and an adjusted IFR of 0.55%. A higher number of daily contacts with adults and older adults increases the probability of becoming infected. Occupation, living in an apartment versus a house, and wearing a face mask outside work increased the number of daily contacts.

Conclusions: Studying prevalence in an asymptomatic population revealed estimates of unreported COVID-19 cases. Occupational, living situation, and behavioral data about COVID-19–protective behaviors such as wearing a mask may aid in the identification of nonclinical factors affecting the number of daily contacts, which may increase SARS-CoV-2 exposure.

(JMIR Public Health Surveill 2021;7(1):e24320) doi:10.2196/24320

KEYWORDS

COVID-19; risk of infection; community exposure; self-protecting behavior; mask wearing; infection fatality rate; infection; self-protecting; mask; fatality rate; exposure; virus; SARS-CoV-2

Introduction

Since the global outbreak of SARS-CoV-2 (and the disease it causes, COVID-19), there has been significant research interest in understanding the disease's ability to spread within populations. However, understanding the spread of COVID-19 has been particularly challenging because of asymptomatic spread [1-3]. With new rapidly developing assays capable of identifying serum antibodies, some regions and countries have launched investigations to identify the prevalence of asymptomatic infection. Studies that have collected seroprevalence data have used it to test the sensitivity and specificity of an enzyme immunoassay and microneutralization assay in Hong Kong [4], to combine and evaluate targeted testing and population screening in Iceland [5], and to compare incidence and infection fatality rates in the worst-hit towns in Germany after a superspreading event [6]. What these studies have in common is that they are aimed at improving epidemiological models of how the virus spreads and evaluating its transmission behavior. They include important indicators such as age, gender, and pre-existing conditions, as well as recent travel [5] and household size [6]. These studies help identify the proportion of the population at risk, increase understanding about hospitalization and fatality rates, and help guide decision making regarding strategies to control the pandemic. Other studies have focused on environmental and behavioral factors in the population without knowing infection rates in the same population [7-9].

Less studied is how environmental and behavioral factors such as occupation, housing situation, and COVID-19–protective behaviors affect infection rates. To date, most studies have focused on demographic risk factors among those who have tested positive for the virus [6,10]. Individual protective behaviors like wearing a mask are seldom studied in the COVID-19 literature [8]. There are two recent studies that reported important differences in mask-wearing practices between countries during COVID-19 pandemic, including countries in the East and West [11], and two neighboring countries (The Netherlands and Belgium) [12].

The objective of this study was to measure the prevalence and incidence of COVID-19 in the hospital service area (HSA) of the University of Vermont Medical Center (UVMMC), identify factors that may increase or decrease the risk of infection and exposure, and to analyze factors that increase the number of daily contacts. UVMMC is the largest hospital and most densely populated county in a rural state in the northeastern United States and its HSA is the area of the local community that is intended to be served by the hospital. We evaluated the prevalence of SARS-CoV-2 among community-dwelling adults in the most densely populated county in Vermont after the height of the COVID-19 pandemic in June 2020 and explored the environmental and behavioral factors associated with the risk of infection. At the time of this study, Vermont had a very low rate of COVID-19 infection. Active disease rates in the population were low and remained low throughout recent months. We conducted our study in Vermont, because we were able to obtain data from a representative sample of the most densely populated county in the state, accounting for approximately one-third of the total population of the state. We hope that this study serves as an example for more studies linking COVID-19 seroprevalence in the general population to behavioral data potentially affecting the spread of COVID-19.

This research combined individual survey data on COVID-19 risks and social behaviors with polymerase chain reaction (PCR) testing results from nasopharyngeal swabs and two different serologic assays. The addition of biological testing to known epidemiological data allowed for the calculation of accurate population prevalence rates, the true hospitalization and infection fatality rates, and inferences about exposure to the virus that may have more widespread implications.

Methods

Recruitment

Our sampling frame included community members from Chittenden county in the HSA of UVMMC who had an encounter with their primary care provider in the past 3 years. Using electronic health records, we randomly selected 12,000 individuals aged 18-70 years who had at least one primary care visit during the preceding 3 years, stratified by age and gender.

Individuals were contacted via email in two waves between April 30 and May 13, 2020, and asked to consent to participate in the survey.

After completing the survey, an offer was sent to these 1694 participants to receive PCR and serologic testing. To prevent recruitment bias among people who may have been motivated to obtain COVID-19 testing, participants were not aware of this optional testing component when filling out the survey.

Survey

The survey instrument was developed by an international group of researchers and previously used to collect data from different countries [7]. The information collected included work and living situations, income, COVID-19–protective behaviors (such as wearing a face mask), beliefs about the COVID-19 pandemic and exposure to the virus, sociodemographic characteristics, and prepandemic health status. The survey also gathered specific information from respondents about the type of industry in which they are employed and their precise profession within that industry. Respondent profession was linked to profession exposure data derived from data from the US Department of Labor/Employment and Training Administration's Occupational Information Network (O*NET) survey, which categorizes the level of exposure to disease/infections for a wide range of professions [13]. This O*NET measure has also been used by others linking job exposure to COVID-19 [7]. Scores range from 0 to 100, where 0 is "never," 50 is "once a month or more but not every week," and 100 is "every day." Survey data was collected and stored via REDCap.

COVID-19 Tests

COVID-19 prevalence (active infection) was tested with PCR testing on nasopharyngeal swabs, while incidence rate was tested using two different serologic assays performed on patient-matched blood samples. The PCR test detects the genetic code for the SARS-CoV-2 virus (which causes COVID-19) and identifies active COVID-19 infection. The serologic tests detect antibodies to COVID-19 and indicate whether the participant has mounted an immune response to the virus. COVID-19 prevalence (active infection) was tested at the State of Vermont Department of Health Laboratory by PCR using the TaqPathTM COVID-19 Combo Kit (ThermoFisher, catalog numbers A47813 and A47814) on ribonucleic acid (RNA) extracted from nasopharyngeal swabs. This assay was granted Emergency Use Authorization [14], and uses primer sets targeting the ORF1ab, nucleocapsid, and spike regions of the SARS-CoV-2 genome. Each assay includes a positive SARS-CoV-2 RNA control (50 copies per reaction), a negative (diluent-only) control, and an MS2 phage as an internal process control for nucleic acid extraction. Briefly, RNA was extracted from nasopharyngeal swabs, reverse transcribed using the one-step multiplex Mastermix, and assessed on an Applied Biosystems 7500-Fast Dx PCR instrument as listed in the product manual using a sample cycle threshold (CT) cutoff of ≤37 for the calling of positives.

Serologic testing was done on separated serum (BD SST catalog number 367977) using two different assays granted Emergency Use Authorization by the Food and Drug Administration: (1) the VITROS (Ortho Clinical Diagnostics) anti–SARS-CoV-2 IgG test conducted by the Mayo Clinic and (2) an open-source laboratory-developed two-step enzyme-linked immunosorbent assay (ELISA) originally developed by the Mount Sinai School of Medicine [15] and conducted at the University of Vermont Larner College of Medicine. Both assays exhibit ≥90% sensitivity and 100% specificity with ≥99.5% negative predictive value (NPV) at a prevalence of 5% [16]. The two-step IgG ELISA was recently validated to over 99% sensitivity in samples from patients with COVID-19 [17]. Serology for the receptor binding domain of the SARS-CoV-2 spike protein (RBD-S) has been shown to exhibit extremely low cross-reactivity for other non-SARS coronaviruses [18] and to correlate with neutralization activity [17,19], making it a highly specific and relevant measure of SARS-CoV-2 infection.

Statistical Analysis

The testing results were merged with the survey data. Observations that had missing values for key variables were deleted (n=19), which left us with a total sample size of 435 for the multivariate analysis. We had two outcome variables in the analysis. The first was whether or not the person tested positive for COVID-19 antibodies. The second was the number of contacts the person had on a "typical" day (<18, 18-64, and >64) during the two weeks prior to the survey.

For the dichotomous outcome variable (whether a participant had a positive COVID-19 antibody test), we performed multivariate analyses using Probit models. The count data representing the number of daily contacts for the participants followed a Poisson distribution: the number of people seen outside the household can be seen as rare events, since many respondents did not see others at all. As the Poisson distribution assumes that the mean and variance are the same, we tested the fit of a Poisson model versus negative binomial models [15]. The likelihood ratio test is a test of the overdispersion parameter α: when α is zero, the more flexible negative binomial distribution is equivalent to a Poisson distribution. In our case, α was significantly different from zero, suggesting the negative binomial distribution was appropriate, so we used nbreg in Stata 16.0 (StataCorp) to analyze the number of daily contacts. We used a Vuong test of the zero-inflated model versus the standard model [16,20] and found that the excess zeros should not be modeled independently. We ran different models for number of contacts with children, adults, and older adults. We used robust standard errors for the negative binomial models. Statistical analysis was performed in Stata, including descriptive statistics and multivariate analysis.

Key control variables included age (because of the relatively small sample size of positives, age was dichotomized to over and under 45 years), income (in categories, and dichotomous >$100,000/<$100,000), gender, education (college yes/no), and presence of chronic illnesses (yes/no from a list including conditions identified by the Centers for Disease Control and Prevention as increasing the risk of COVID-19 complications, which included diabetes, high blood pressure/hypertension, heart disease, asthma or other chronic respiratory issues, allergies, and kidney disease or other chronic illnesses that require long-term care from a doctor). We also included variables indicating whether the participant had lost their job due to COVID-19 and whether their work situation had changed (working from home instead of previous location), whether they had been tested before, what symptoms they had and whether they sought testing for those symptoms, whether they had been diagnosed, whether they had pre-existing conditions, and whether they had been in contact with others who had tested positive.

Human Subjects Research Review Statement

This study has been approved by the Institutional Review Board of the University of Vermont. We received separate approval for the survey study and the COVID-19 testing study.

Study participants signed eConsent forms for both the survey part and the testing part of the study. There was no compensation for participation in this study.

The health information of participants is protected by a federal law called the Health Information Portability and Accountability Act (HIPAA). The study team stored the data from the survey and COVID-19 tests in a safe environment. Only the research team, the UVM Institutional Review Board, and state and federal agencies that oversee research have access to this information. No identifying data was made available to any other sources.

Results

Participation

A total of 12,000 patients were invited to participate in this study. All individuals were provided with an opportunity to opt out of the survey at any time during the study. A total of three follow-up reminders were sent. Of this initial sample of 12,000 individuals, 98% had functioning email addresses (n=11,700); the response rate was 19.4% (n=2275), and 75% of these respondents both read the consent form and agreed to participate (n=1961 participants). Of these, 86.4% completed the survey, for a total of 1694 respondents (14.4% of the initial sample).

All 1694 survey respondents were invited to opt in to COVID-19 testing. A total of 26.8% (n=454) of participants provided samples between June 25 and June 28, 2020.

COVID-19 Test Results

In total, 10 of 454 participants tested positive for IgG antibodies in a two-step serologic assay in which samples with presumed IgG reactivity against the RBD-S are confirmed in an independent assay wherein the IgG endpoint titer against the full-length SARS-CoV-2 spike protein is determined. Of the 10 samples, 6 were confirmed by the VITROS SARS-CoV-2 IgG assay, which detects an undisclosed antigen from SARS-CoV-2 and provides a nonquantitative positive/negative result. These 6 samples exhibited an average anti–RBD-S optical density (OD) of 0.91 (SD 0.22; range 0.64-1.28) and anti-spike reciprocal IgG titers of 21,300 (SD 27,500; range 900-72,900) in the two-step assay performed at the University of Vermont. The remaining 4 exhibited an OD of 0.56 (SD 0.55; range 0.18-1.38) and titer of 350 (SD 380; range 100-900). There was not a statistically significant difference (P=0.2 for OD and P=0.17 for titer by paired Student t test) between the 6 UVM/VITROS-positive and 4 UVM-only positive samples. Furthermore, all positive samples by two-step IgG assay met the assay positivity cutoff requirements (Step 1: RBD-S OD ≥2-fold over background, which was ~0.08 and Step 2: titer≥80). The positivity rate for antibodies against SARS-CoV-2 in our catchment area was therefore 2.2% (95% CI 0.8%-3.6%). Only 1 participant (0.2%) tested positive for active SARS-CoV-2 replication using the nasopharyngeal swab.

Extrapolating these serology results to the 164,572 residents of the county, approximately 3621 have been infected by COVID-19 so far (95% CI 1317-5925). The State Department of Health reported a total of 662 positive cases in the same county at the time the study test samples were obtained. This implies that 18.3% of positive cases have been identified by the existing community-based testing (95% CI 11.2%-50.3%).

From the onset of the COVID-19 pandemic to the time of our data collection, 50 individuals from the county had been hospitalized at UVMMC. This implies that 1.4% of persons with COVID-19 required hospital care during the March-July 2020 time frame (95% CI 0.8%-3.8%). At the time of study completion, there have been a total of 39 deaths attributed to COVID-19 in Vermont, which implies an infection fatality rate of 1.1% (95% CI 0.7%-3.0%). Of the 39 deaths, 19 (48.7%) were in nursing homes. If these deaths are excluded, we calculate a case fatality rate of 0.55%.

We did not perform statistical analyses with the PCR results, because we only found 1 positive PCR test and therefore did not have enough statistical power for analysis.

Factors Associated With Positive SARS-CoV-2 Test

Table 1 shows the association between positive serology for SARS-CoV-2 and select sociodemographic factors. The number of contacts with both adults and older adults was statistically significantly higher for those who tested positive than those who did not (5.0 versus 31.6, P<.001 and 2.9 versus 14.8, P<.001, respectively). There was no statistically significant relationship for the number of contacts with children. Similarly, the number of contacts with people who tested positive was higher for the COVID-19 population (0.9) versus the negative subjects (0.1; P<.001). There were no statistically significant differences between those who tested positive and those who did not in average age, gender, number of reported symptoms, work exposure, urbanity, living environment, or mask wearing outside work.

Table 1. Descriptive statistics from the sample of 435 survey respondents showing frequencies and percentages in different categories of risk factors associated with COVID-19 infection in Vermont between April 20 and May 13, 2020.

Respondent characteristic	COVID-19–negative subjects (n=425), mean or proportion (SD)[a]	COVID-19–positive subjects (n=10), mean or proportion (SD)[a]	*T* statistic	*P* value
Number of contacts with children	1.5 (0.3635)	1.0 (0.4629)	0.1754	.86
Number of contacts with adults	5.0 (0.6383)	31.6 (20.1112)	–5.0571	<.001
Number of contacts with older adults	2.9 (0.6565)	14.8 (7.4276)	–2.4882	.01
Number of contacts with people who tested positive for COVID-19	0.1 (0.4247)	0.9 (1.3562)	–1.8110	.07
Age (years)	51.4 (0.6366)	51.9 (3.6028)	–0.1075	.91
Sex (female=1)	0.6 (0.0241)	0.6 (0.1830)	–0.2230	.82
High income (1=yes)	0.6 (0.0243)	0.4 (0.1830)	0.9561	.34
Any symptoms (1=yes)	0.2 (0.0180)	0.5 (0.1890)	–2.6804	.007
Diabetes (1=yes)	0.04 (0.0100)	0.13 (0.1250)	–1.1380	.26
Exposure at work, Occupational Information Network score (1-100)	26.2 (1.8328)	19.0 (12.0208)	0.5159	.61
Urban (versus suburban/rural; 1=yes)	0.3 (0.0218)	0.3 (0.1637)	0.1638	.87
Live in condominium/apartment (versus house; 1=yes)	0.2 (0.01923)	0.3 (0.1637)	–0.4041	.69
Wearing mask outside work (1=yes)	0.7 (0.230)	0.5 (0.1890)	1.1382	.26

[a]Based on a *t* test, Ha: diff!=0.

Regression Results

Table 2 presents the results of the Probit regressions examining factors associated with positive COVID-19 test results. The three columns represent the different models. The first shows the effect of the number of daily contacts with children, the second shows the effect of the daily number of contacts with adults, and the third shows the effect of the daily number of contacts with older adults (>65 years). We used generally accepted standards for children (those aged <18 years), older adults (those aged ≥65 years), and adults (those aged 18-64 years). We found that with every additional adult that participants would see on a daily basis, they had a 1.2% (*P*<.05) higher probability of getting a positive test result. For contact with older adults, this increased probability was the same (1.2%, *P*<.05). With each additional contact with a person who had tested positive for COVID-19, participants had a 44.1%-53.6% (*P*<.05) higher probability of testing positive for the virus. Those aged >45 years had a 20.4%-24.8% higher probability of infection with each additional contact. We found no other covariates to be statistically significant in our models.

Table 2. Predicted probabilities of COVID-19 infection in Vermont between April 20 and May 13, 2020.

Respondent characteristic	Model 1: probability of COVID-19 infection: children contact model[a]	*P* value	Model 2: probability of COVID-19 infection: adult contact model[a]	*P* value	Model 3: probability of COVID-19 infection: older adult contact model[a]	*P* value
Number of contacts with children (those aged 0-17 years)	–0.0104 (0.0403)	.88	N/A[b]	N/A	N/A	N/A
Number of contacts with Adults (those aged 18-64 years)	N/A	N/A	0.0118 (0.0059)	.03	N/A	N/A
Number of contacts with older adults (those aged ≥65 years)	N/A	N/A	N/A	N/A	0.0122 (0.0063)	.05
Number of contacts with people who tested positive for COVID-19	0.5359 (0.2111)	.04	0.4406 (0.2243)	.16	0.5356 (0.2123)	.42
Aged ≥45 years	0.2038 (0.3701)	.36	0.2432 (0.3854)	.27	0.2484 (0.3846)	.29
Female	–0.0766 (0.3404)	.77	0.0501 (0.3632)	.53	–0.0025 (0.3567)	.65
High income	–0.3967 (0.3402)	.15	–0.3377 (0.3480)	.18	–0.4023 (0.3484)	.14
Any symptoms	0.3969 (0.3579)	.08	0.4159 (0.3703)	.07	0.4052 (0.3661)	.07
Diabetes	0.1394 (0.6342)	.96	–0.4759 (0.9156)	.46	–0.1672 (0.7184)	.68
Observations	413	N/A	413	N/A	413	N/A

[a]Probit regression marginal effects with standard errors in parentheses. Pseudo R^2: 0.42.

[b]N/A: not applicable.

Table 3 presents the results of the negative binomial models reporting factors affecting the number of daily contacts. As expected, the more work exposure (as identified in the O*NET index), the more daily contacts participants would have. This is especially true for professions in which one sees more older adults. We also found that females saw almost one fewer adult per day than men did (β=0.88, *P*<.01) and that those living in an apartment or condominium rather than a house would see almost one adult more on a daily basis (β=0.78, *P*<.05). Interestingly, results showed that workers who wear masks outside of work also saw more adults than those who did not wear a mask outside of work (β=0.77, *P*<.01).

Table 3. Relationship between survey respondent characteristics and number of daily contacts in Vermont between April 20 and May 13, 2020.

Characteristics	Model 1: number of contacts with children[a]	*P* value	Model 2: number of contacts with adults[a]	*P* value	Model 3: number of contacts with older adults[a]	*P* value
Exposure at work	0.0214 (0.0123)	.08	0.0194 (0.0048)	<.001	0.0391 (0.0111)	.009
Female	–0.3636 (0.8061)	.70	–0.8757 (0.3395)	.004	0.1514 (0.8568)	.78
Aged ≥45 years	–0.1720 (0.7237)	.97	0.2602 (0.3877)	.63	1.3367 (0.9235)	.57
Urban (versus suburban)	0.7666 (0.9628)	.42	0.5204 (0.3646)	.16	1.0165 (1.2149)	.54
Living in condominium/apartment (versus house)	–1.6075 (0.9300)	.08	0.7765 (0.4091)	.08	0.4172 (0.9216)	.99
Wearing mask outside work	1.0897 (0.7637)	.10	0.7658 (0.2822)	.01	0.3159 (0.8215)	.85
Observations	49	N/A[b]	49	N/A	49	N/A

[a]Negative binomial estimates, standard errors in parentheses.

[b]N/A: not applicable.

Discussion

Principal Results

In this study, we evaluated the prevalence of SARS-CoV-2 among community-dwelling adults in the most densely populated county in Vermont after the height of the COVID-19 pandemic in June 2020, and explored the environmental and behavioral factors associated with the risk of infection. We found a seroprevalence rate of 2.2% and an infection fatality rate of 0.55% after excluding deaths in nursing homes. In the multivariate analysis, we found that the number of daily contacts with adults and older adults increased the probability of infection. Type of occupation, living in an apartment or condominium versus a house, and wearing a face mask outside work increased the number of daily contacts.

The main objective of this study was to identify the prevalence of COVID-19 in an asymptomatic (general) population and identify behavioral and environmental differences between the infected and the uninfected. There are some COVID-19 seroprevalence studies to date, such as one in Iceland [5], which included volunteers from the total population, and a nationwide study in Spain [21]. There are also a few studies among subpopulations, such as one among health care workers in Northern Italy [22], and regional populations in Hong Kong and China [4,23], the United States [10,24,25], and Switzerland [26]. Most of these studies selected participants randomly, but environmental factors potentially affecting the seroprevalence numbers were undetermined. Important predictors in COVID-19 predictive simulation models, such as to what extent social distancing had been practiced, were unknown in these studies. Therefore, a second goal of this study was to get a better idea of actual social distancing practices in our research area and use this data to better inform modelling efforts to predict infection and hospitalization rates. The uniqueness of this study is that it combines survey data with COVID-19 testing data, which has not been done in many other places. To our knowledge, there has been one other study linking seroprevalence data to survey data [27]; however, that study in Germany primarily focused on symptoms and did not include factors such as daily routines and behaviors. Although we acknowledge the limitations of our research study, we believe it serves as an example of how to effectively link behavioral and clinical data.

We were able to identify environmental and behavioral factors affecting the risk of contracting COVID-19. We found that seeing more children per day does not increase the probability of getting COVID-19, but having more daily contact with adults and older adults does. We further identified factors that have an increasing effect on the number of daily contacts, such as living in an apartment and wearing a mask.

Limitations

Our study does have a number of limitations. One is the assumption that the prevalence rates from our sample are representative of prevalence rates for the Chittenden county population. Our sample may be nonrepresentative because of both the inclusion criteria (those with the University of Vermont Medical Center as their primary care destination) and exclusion criteria (those aged <18 years and >70 years and pregnant people). However, because we do not anticipate that the inclusion and exclusion criteria are correlated with disease prevalence in the community, our results are likely representative of the population.

There are a number of possible mechanisms that could create differences between our sample and the population and thereby potentially create bias in the estimates of community prevalence. There may be bias based on observable characteristics, as we drew our initial sample from a population that either has registered with a primary care physician or had a health event in the past 3 years. To test this possibility, we applied sample weights using Census population data for Chittenden county and found our results to be robust to weighted regression results. However, this does not address unobserved characteristics such as wealth and travel time, which may introduce selection bias. For example, if persons who believed they were infected were more likely to participate, this would create an upward bias in our estimates. To test this possibility, we estimated a weighted regression, including all survey respondents including those who received the test invitation but declined. We found no significant difference in the estimates using both populations (tested and not tested). It is also possible the results are biased based on unobserved differences, both between the sample and the population and between the survey sample and the prevalence sample. In the absence of an appropriate instrument, we could not test this effect. Specimen collection was done a little over one month after survey responses were completed. This lag in data collection may potentially pose a temporality issue in the analyses, especially related to risk factors and PCR positivity. However, our infection analysis did not focus on PCR positivity but on a positive serologic test, which addresses infection over a larger period of time.

Compared to other US states, the hospitalization rates for COVID-19 we calculated have been at the lower end of the predicted range of COVID-19 inpatient predictive models [17]. The IFR is at the higher end of reported population rates, largely driven by a high number of nursing home deaths. The data we collected finds approximately 1 out of every 100 individuals infected with COVID-19 in the county needed inpatient care. This provides a benchmark to use to anticipate future shortages of hospital capacity.

The state of Vermont has had a very low rate of COVID-19 infection since the beginning of the pandemic. Active disease rates in the population are currently very low and have been low throughout the duration of the pandemic. Although we were able to test a large sample for COVID-19, the number of positive cases was small, which limited the multivariate analysis. To simplify models, we dichotomized some covariates, thereby losing some more detailed information about the exact effect size of individual levels of the covariates.

By testing in the general population, estimations about the total number of infections in similar demographic areas with different infection rates can be made based on the IFR and the number of deaths. This facilitates the kinds of (inter)national comparisons that could be helpful for developing effective mitigation strategies. Comparing the IFR with numbers of officially reported infections can allow for more refined estimates of unreported cases, which is another data point that is important for understanding pandemic dynamics.

Conclusions

This study has several important policy implications for contemplating different COVID-19 mitigation strategies. We found that the key factors associated with a higher probability of testing positive for COVID-19 were the number of contacts with adults and older adults, particularly contacts with people who have COVID-19. The factors that predict contacts, in turn, are working environment, living environment, and regularly wearing a mask outside of work. This study reinforces the concerns about risks for persons who have high levels of public contact during the pandemic. The finding of the increased risk associated with living in apartments/condominiums likely partially explains higher infection rates in large metropolitan areas (eg, New York City) and lower income communities.

The findings with regard to mask wearing are more concerning. With many states and governments now debating whether the use of face masks should become mandatory, more research is needed about the behavioral effects of mask wearing and other policy measures. A recent study showed that mask wearing is associated with a lower prevalence of depression, which may be explained by seeing more people [28]. Another study addressed specific measures in the work environment to prevent COVID-19 [29]. It is plausible that mandating masks could be counterproductive if the increased risk associated with an increase in contacts is larger than the decrease in risk associated with mask wearing. That is, it is possible masks may provide a false sense of security that leads to people letting their guard down and trusting the mask more than is warranted. Further research into the effectiveness of masks and behavioral responses to mask mandates is urgently needed.

Acknowledgments

We thank all of the study participants for their time and effort. We thank Christine Werneke for coordination, Denis Hudon, and the UVMMC Emergency Department for sample collection. We also Camilla Strother and Nancy Graham for sample processing and performing SARS-CoV-2 serology assays at the UVM Larner College of Medicine. We thank Helen Reid, health surveillance division director at the Vermont Department of Health Laboratory, for performing PCR tests. We thank the UVM Research Protections Office, Institutional Review Board, and Institutional Biosafety Committee for rapid turnaround of COVID-19–related projects.

Serology tests performed at UVM were funded by a pilot grant to SAD from the UVM Translational Global Infectious Disease Research Center (National Institute of Health grant P20GM125498). Additional funding was from NIH grant U01AI141997 to SAD, and the University of Vermont Larner College of Medicine Departments of Emergency Medicine and Surgery. The University of Vermont Department of Radiology funded serologic tests performed at Mayo. The Vermont Department of Health funded PCR tests performed at the state laboratory.

Conflicts of Interest

None declared.

References

1. Anguelov R, Banasiak J, Bright C, Lubuma J, Ouifki R. The big unknown: The asymptomatic spread of COVID-19. BIOMATH 2020 May 11;9(1):2005103. [doi: 10.11145/j.biomath.2020.05.103]
2. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: 10.1056/NEJMe2009758] [Medline: 32329972]
3. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin JI. Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis. Int J Infect Dis 2020 Sep;98:180-186 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.052] [Medline: 32562846]
4. To KK, Cheng VC, Cai J, Chan K, Chen L, Wong L, et al. Seroprevalence of SARS-CoV-2 in Hong Kong and in residents evacuated from Hubei province, China: a multicohort study. The Lancet Microbe 2020 Jul;1(3):e111-e118. [doi: 10.1016/s2666-5247(20)30053-7]
5. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med 2020 Jun 11;382(24):2302-2315 [FREE Full text] [doi: 10.1056/NEJMoa2006100] [Medline: 32289214]
6. Streeck H, Schulte B, Kümmerer BM, Richter E, Höller T, Fuhrmann C, et al. Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. Nat Commun 2020 Nov 17;11(1):5829 [FREE Full text] [doi: 10.1038/s41467-020-19509-y] [Medline: 33203887]
7. Belot M, Syngjoo C, Jamison J, Papageorge N, Tripodi E, van den Broek-Altenburg E. Six-Country Survey on COVID-19. SSRN. 2020. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596697 [accessed 2020-06-20] [WebCite Cache ID https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3596697]

8. Belot M, Syngjoo C, Jamison J, Papageorge N, Tripodi E, van den Broek-Altenburg E. Unequal Consequences of Covid 19 across Age and Income: Representative Evidence from Six Countries. SSRN. 2020. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3638012 [accessed 2020-06-20]
9. Papageorge N, Zahn M, Belot M, van den Broek-Altenburg E, Syngjoo C, Jamison J, et al. Socio-Demographic Factors Associated with Self-Protecting Behavior during the COVID-19 Pandemic. IZA Institute of Labor Economics. 2020. URL: https://www.iza.org/publications/dp/13333/socio-demographic-factors-associated-with-self-protecting-behavior-during-the-covid-19-pandemic [accessed 2020-06-20]
10. Bendavid E, Mulaney B, Shah S, Ling E, Bromley-Dulfano R, Lai C, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. MedRxiv. Preprint published online on April 16, 2020. [doi: 10.1101/2020.04.14.20062463]
11. Wang C, Chudzicka-Czupała A, Grabowski D, Pan R, Adamus K, Wan X, et al. The Association Between Physical and Mental Health and Face Mask Use During the COVID-19 Pandemic: A Comparison of Two Countries With Different Views and Practices. Front Psychiatry 2020;11:569981 [FREE Full text] [doi: 10.3389/fpsyt.2020.569981] [Medline: 33033485]
12. van den Broek-Altenburg E, Atherly A. Adherence to COVID-19 Policy Measures: Behavioral Insights from the Netherlands and Belgium. SSRN. 2020 Sep 14. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3692644 [accessed 2020-12-26]
13. Work Context — Exposed to Disease or Infections. O*Net OnLine. URL: https://www.onetonline.org/find/descriptor/result/4.C.2.c.1.b [accessed 2020-06-20]
14. Food and Drug Administration. TaqPath COVID-19 Combo Kit - Letter of Authorization. 2020 Oct 09. URL: https://www.fda.gov/media/136113/download [accessed 2020-12-21]
15. Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova C, McMahon M, et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat Med 2020 Jul;26(7):1033-1036. [doi: 10.1038/s41591-020-0913-5] [Medline: 32398876]
16. Food and Drug Administration. EUA Authorized Serology Test Performance. URL: https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance [accessed 2020-06-20]
17. Graham N, Whitaker A, Strother C, Miles A, Grier D, McElvany B, et al. Kinetics and isotype assessment of antibodies targeting the spike protein receptor-binding domain of severe acute respiratory syndrome-coronavirus-2 in COVID-19 patients as a function of age, biological sex and disease severity. Clin Transl Immunology 2020;9(10):e1189 [FREE Full text] [doi: 10.1002/cti2.1189] [Medline: 33072323]
18. Premkumar L, Segovia-Chumbez B, Jadi R, Martinez D, Raut R, Markmann A, et al. The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. Sci Immunol 2020 Jun 11;5(48) [FREE Full text] [doi: 10.1126/sciimmunol.abc8413] [Medline: 32527802]
19. Long Q, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020 Jun;26(6):845-848. [doi: 10.1038/s41591-020-0897-1] [Medline: 32350462]
20. Greene W. The Econometric Approach to Efficiency Analysis. In: Fried HO, Schmidt SS, editors. The Measurement of Productive Efficiency: Techniques and Applications. Oxford, UK: Oxford University Press; 1993.
21. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, ENE-COVID Study Group. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet 2020 Aug 22;396(10250):535-544 [FREE Full text] [doi: 10.1016/S0140-6736(20)31483-5] [Medline: 32645347]
22. Sotgiu G, Barassi A, Miozzo M, Saderi L, Piana A, Orfeo N, et al. SARS-CoV-2 specific serological pattern in healthcare workers of an Italian COVID-19 forefront hospital. BMC Pulm Med 2020 Jul 29;20(1):203 [FREE Full text] [doi: 10.1186/s12890-020-01237-0] [Medline: 32727446]
23. Xu X, Sun J, Nie S, Li H, Kong Y, Liang M, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. Nat Med 2020 Aug;26(8):1193-1195 [FREE Full text] [doi: 10.1038/s41591-020-0949-6] [Medline: 32504052]
24. Ng D, Goldgof G, Shy B, Levine A, Balcerek J, Bapat SP, et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco Bay Area. medRxiv. Preprint published on May 27, 2020 [FREE Full text] [doi: 10.1101/2020.05.19.20107482] [Medline: 32511477]
25. Sood N, Simon P, Ebner P, Eichner D, Reynolds J, Bendavid E, et al. Seroprevalence of SARS-CoV-2-Specific Antibodies Among Adults in Los Angeles County, California, on April 10-11, 2020. JAMA 2020 Jun 16;323(23):2425-2427 [FREE Full text] [doi: 10.1001/jama.2020.8279] [Medline: 32421144]
26. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. The Lancet 2020 Aug;396(10247):313-319. [doi: 10.1016/S0140-6736(20)31304-0]
27. Aziz N, Corman V, Echterhoff A, Richter A, Schmandke A, Schmidt M, et al. Seroprevalence and correlates of SARS-CoV-2 neutralizing antibodies: Results from a population-based study in Bonn, Germany. MedRxiv. Preprint published online on August 29, 2020 [FREE Full text] [doi: 10.1101/2020.08.24.20181206]

28. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
29. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]

Abbreviations

IFR: infection fatality rate
HSA: hospital service area
NPV: negative predictive value
O*NET: US Department of Labor/Employment and Training Administration's Occupational Information Network
OD: optical density
PCR: polymerase chain reaction
RBD-S: receptor binding domain of the SARS-CoV-2 spike protein
RNA: ribonucleic acid
UVMMC: University of Vermont Medical Center

Edited by T Sanchez; submitted 14.09.20; peer-reviewed by A Dormanesh, R Ho; comments to author 07.10.20; revised version received 02.11.20; accepted 27.11.20; published 11.01.21.

Please cite as:
van den Broek-Altenburg EM, Atherly AJ, Diehl SA, Gleason KM, Hart VC, MacLean CD, Barkhuff DA, Levine MA, Carney JK
Jobs, Housing, and Mask Wearing: Cross-Sectional Study of Risk Factors for COVID-19
JMIR Public Health Surveill 2021;7(1):e24320
URL: http://publichealth.jmir.org/2021/1/e24320/
doi:10.2196/24320
PMID:33315576

Original Paper

The Association Between Chronic Disease and Serious COVID-19 Outcomes and Its Influence on Risk Perception: Survey Study and Database Analysis

Pedro Almeida Laires[1,2], MSc, PhD; Sónia Dias[1,2], PhD; Ana Gama[1,2], PhD; Marta Moniz[1,2], MSc; Ana R Pedro[1,2], PhD; Patricia Soares[1,2], PhD; Pedro Aguiar[1,2], PhD; Carla Nunes[1,2], PhD

[1]Public Health Research Centre, NOVA National School of Public Health, Universidade NOVA de Lisboa, Lisbon, Portugal

[2]Comprehensive Health Research Center, Universidade NOVA de Lisboa, Lisboa, Portugal

Corresponding Author:
Pedro Almeida Laires, MSc, PhD
Public Health Research Centre
NOVA National School of Public Health
Universidade NOVA de Lisboa
Av Padre Cruz
Lisbon
Portugal
Phone: 351 919783234
Email: pedro.laires@ensp.unl.pt

Abstract

Background: COVID-19, a viral respiratory disease first reported in December 2019, quickly became a threat to global public health. Further understanding of the epidemiology of the SARS-CoV-2 virus and the risk perception of the community may better inform targeted interventions to reduce the impact and spread of COVID-19.

Objective: In this study, we aimed to examine the association between chronic diseases and serious outcomes following COVID-19 infection, and to explore its influence on people's self-perception of risk for worse COVID-19 outcomes.

Methods: This study draws data from two databases: (1) the nationwide database of all confirmed COVID-19 cases in Portugal, extracted on April 28, 2020 (n=20,293); and (2) the community-based COVID-19 Barometer survey, which contains data on health status, perceptions, and behaviors during the first wave of COVID-19 (n=171,087). We assessed the association between relevant chronic diseases (ie, respiratory, cardiovascular, and renal diseases; diabetes; and cancer) and death and intensive care unit (ICU) admission following COVID-19 infection. We identified determinants of self-perception of risk for severe COVID-19 outcomes using logistic regression models.

Results: Respiratory, cardiovascular, and renal diseases were associated with mortality and ICU admission among patients hospitalized due to COVID-19 infection (odds ratio [OR] 1.48, 95% CI 1.11-1.98; OR 3.39, 95% CI 1.80-6.40; and OR 2.25, 95% CI 1.66-3.06, respectively). Diabetes and cancer were associated with serious outcomes only when considering the full sample of COVID-19–infected cases in the country (OR 1.30, 95% CI 1.03-1.64; and OR 1.40, 95% CI 1.03-1.89, respectively). Older age and male sex were both associated with mortality and ICU admission. The perception of risk for severe COVID-19 disease in the study population was 23.9% (n=40,890). This was markedly higher for older adults (n=5235, 46.4%), those with at least one chronic disease (n=17,647, 51.6%), or those in both of these categories (n=3212, 67.7%). All included diseases were associated with self-perceptions of high risk in this population.

Conclusions: Our results demonstrate the association between some prevalent chronic diseases and increased risk of worse COVID-19 outcomes. It also brings forth a greater understanding of the community's risk perceptions of serious COVID-19 disease. Hence, this study may aid health authorities to better adapt measures to the real needs of the population and to identify vulnerable individuals requiring further education and awareness of preventive measures.

(JMIR Public Health Surveill 2021;7(1):e22794) doi:10.2196/22794

KEYWORDS

COVID-19; risk factors; morbidity; chronic disease; risk; perception; outcome; association

Introduction

COVID-19, a viral respiratory disease caused by SARS-CoV-2, has become a global threat to human health [1,2]. By early December 2020, more than 68 million SARS-CoV-2 infections and over 1.5 million deaths have been reported worldwide [3].

Several studies have reported that chronic conditions, such as respiratory and cardiovascular diseases, are associated with worse outcomes following infection [1,4-8]. Given the rapid spread and high mortality rate of COVID-19 among those with a vulnerable health status, it soon became necessary to expand research to elucidate the epidemiology of the novel virus, namely the identification of risk factors for severe illness or death [9].

There is ample evidence that perceived susceptibility to severe disease outcomes is an important predictor of preventive behavior [10]. In accordance with theories on health behavior decisions [11-14], engagement on preventive behaviors are shaped by the awareness and risk perception, particularly among those who are more vulnerable to severe outcomes [13,15-17]. Preventive behaviors, such as curfews, social distancing, handwashing, and mask wearing, are so far the most effective ways to fight the spread of COVID-19 and related consequences [18,19]. Therefore, it is imperative to explore that risk perceptions of the community, given that such information may inform targeted interventions, including communication and health education strategies, aimed at minimizing the impact and spread of COVID-19.

Thus, the objectives of this study were (1) to examine the association between chronic diseases and worse outcomes following COVID-19 infection (ie, death and intensive care unit [ICU] admission), and (2) to understand its role on the self-perception of risk for worse COVID-19 outcomes.

Methods

Databases

This study draws on two data sources.

COVID-19 Database

The official database of COVID-19 cases in Portugal, which contains anonymized data from the Directorate-General of Health (Direção-Geral da Saúde, DGS), including all confirmed cases of COVID-19 reported to the National Epidemiological Surveillance System (Sistema Nacional de Vigilância Epidemiológica, SINAVE). Data were extracted on April 28, 2020 (n=20,293 laboratory-confirmed cases of COVID-19). SINAVE is an electronic platform through which clinicians are obligated to notify all suspected and confirmed cases of COVID-19, and includes information on clinical findings and pre-existing conditions. Notifications trigger an epidemiological investigation by the Local Public Health Services, where a public health physician (health authority in the area of residence of the infected individual) validates the case. At a later stage, the Regional Public Health Department and finally the DGS conduct a final validation of case information. Outcome data are completed primarily at the local level but can be updated at the regional and national levels (DGS). We compared the characteristics of these COVID-19 cases against a nationwide representative sample from the National Health Survey (Inquérito Nacional de Saúde, INS) (Multimedia Appendix 1) [20].

COVID-19 Barometer

We developed a community-based survey called COVID-19 Barometer, which contains data on health (including mental health), health care utilization, perception of risk, and social experiences of over 180,000 individuals, aged ≥16 years old, in Portugal during the first wave of COVID-19. Potential participants were invited to participate through existing contact networks and mailing lists (including large databases of students, teachers, researchers, staff, and other collaborators at the National School of Public Health [ENSP-NOVA] and other institutions nationwide), digital social networks, and social media promotion. The study was also promoted to vulnerable groups through partnerships with third-sector organizations, including patient associations, public health doctors, and other health care professional groups. Data were collected using a structured, closed-ended questionnaire administered online through the Microsoft Forms software program (Microsoft Corp). The questionnaire was developed based on the Portuguese National Health Survey (INS) items (respondents' sociodemographic characteristics, health status, and health care utilization) [20]. Specific questions about COVID-19 were created by the authors and based on the COVID-19 Rapid Quantitative Assessment Tool of the World Health Organization, whenever possible [21]. The questionnaire was pretested to verify response times, ensure comprehensibility, and solve operational issues. We used the latest available responses from each participant, obtained between March 21 and May 23, 2020 (n=171,087).

Measures

In this study, we considered the following main chronic diseases, which, according to the available evidence, are potential risk factors for COVID-19: respiratory, cardiovascular, and renal diseases; diabetes; and cancer.

Regarding the case definition for the main outcome in the DGS database, we analyzed a composite COVID-19 outcome of death and ICU admission. At the time of the analysis, there was a delay in the notification of death, and thus it was considered a better choice to focus on a broader major outcome.

In the COVID-19 Barometer database, we surveyed the respondent's perception of risk for severe disease in case of COVID-19 infection with the following question: "To what extent do you consider yourself to be at risk of developing serious illness or complications, if you become infected with COVID-19?" We then created a dichotomous variable designating the high-risk category as 1 and other categories as 0 (ie, moderate risk, low risk, no risk, don't know).

Statistical Analysis

The analysis included two main steps. First, logistic regression models were used to assess the association between the selected diseases and death or ICU admission, adjusting for age (categorized into 6 groups with 0-50 years as the reference group and 10-year age intervals until >90 years), sex, region, and other relevant comorbidities available in the database (eg, HIV). Two models were developed—one with all COVID-19 cases and another with a subgroup of hospitalized patients with COVID-19—to better understand the association of morbidity with worse intrahospital COVID-19 outcomes, thereby limiting potential biases arising from a higher likelihood of hospitalization for any given COVID-19 case solely based on the decider's (ie, a health care professional at a hospital) knowledge of the pre-existence of a chronic disease (as discussed below).

Second, the sample was standardized to reflect the distribution of the Portuguese population (by sex and age group using the direct method), and a logistic regression model were used to assess self-perception levels of severe disease in the population and potential influencing factors, mainly the chronic diseases under study. We adjusted the model for age, sex, region, education (grouped into three major levels according to the highest qualification completed: basic or no education, secondary school, and university), other relevant comorbidities, smoking, self-reported health and mental status (both grouped into two major levels: very good/good/moderate and poor/very poor), high-risk professional or living with one (including health care professionals, security personnel, and customer-facing positions), living alone, and confidence in the National Health Survey response. All of these factors were chosen based on the database data availability and on the plausibility of influencing study outcomes. The model was built by means of a manual stepwise technique (backward elimination). In the descriptive analysis, additional results for those aged ≥65 years were also provided, given that it is a common cut-off age criteria for increased COVID-19 risk [22].

All statistical analyses were carried out using Stata, version 13.1 (StataCorp LLC). In the descriptive analysis, the significance of the study variables was tested using the Student *t* test or the chi-square test, where appropriate. The significance level for all analyses was fixed at 5%, and confidence intervals were set at 95%.

Ethical Considerations

Data were shared by the DGS with ENSP-NOVA under a partnership for COVID-19 research. The Ethical Committee of ENSP-NOVA approved the project (approval: CE/ENSP/CREE/2/2020). Anonymity of participants and confidentiality of data in all databases used were guaranteed. Informed consent was obtained from all participants.

Results

Associations Between Chronic Diseases and COVID-19 Outcomes

The average age of all COVID-19 infection cases was 52.1 (SD 21.3) years (men: 51.7 [SD 21.0] years, and women: 52.4 [SD 21.5] years; P=.03). In total, 14.6% (n=2963) were hospitalized, 1.3% (n=263) were admitted to the ICU, and 2.5% (n=502) died (3.6% [n=765] for ICU or death). Among those hospitalized, the average age was 68.9 (SD 18.5) years (men: 67.6 [SD 17.1] years, and women: 70.3 [SD 19.8] years; P<.001). More women were infected (n=11,912, 58.7%), both amongst those below (n=8670, 59.2%) and above 65 years of age (n=3105, 57.4%). However, male gender was more frequently found in cases requiring hospitalization (n=1557, 52.4%) and among those who died or were admitted to the ICU (n=404, 54.9%; P<.001). Male gender was associated with worse outcomes (Table 1). There was also an association between death/ICU admission and chronic diseases (ie, respiratory, cardiovascular, and renal diseases; diabetes; and cancer). When analyzing specifically those who were admitted to the hospital, only lung, cardiovascular, and kidney diseases were associated with this composite outcome (Table 1).

When comparing the COVID-19 database with a nationwide representative sample from the Health Interview Survey, it was noted that, across all the analyzed groups, there was a higher proportion of older adults (≥65 years) infected with COVID-19 compared to the overall country's population—respiratory: 60.0% (n=413) vs 48.3% (n=572); cardiovascular: 91.5% (n=43) vs 69.8% (n=746); renal: 78.8% (n=252) vs 53.2% (n=571); diabetes: 65.5% (n=671) vs 59.3% (n=1216); and at least one of these underlying health conditions: 64.7% (n=1230) vs 54.0% (n=3809), respectively. This asymmetry was particularly evident for renal and cardiovascular diseases (variation of 48.1% and 31.1%, respectively). A lower proportion of women was found in the COVID-19 database versus the country's population (group with at least one of the underlying health conditions: 48.7% [n=927] vs 57.1% [n=4028], respectively).

Table 1. Multivariable logistic regression (odds ratios [OR] and 95% CIs) to assess the association between chronic diseases and severe outcomes (death or admission to intensive care unit) following COVID-19 infection.

Characteristic	All infected (n=20,203)		Hospitalized (n=2958)	
	Univariable, OR (95% CI)	Multivariable, OR (95% CI)	Univariable, OR (95% CI)	Multivariable, OR (95% CI)
Age group (years; reference <50 years)				
50-59	5.14 (3.27-8.08)[a]	5.03 (3.20-7.93)[a]	2.48 (1.53-4.03)[a]	2.52 (1.54-4.13)[a]
60-69	14.81 (9.74-22.40)[a]	12.36 (8.15-18.76)[a]	3.31 (2.13-5.14)[a]	3.21 (2.04-5.02)[a]
70-79	35.65 (23.98-52.99)[a]	24.70 (16.48-37.01)[a]	3.85 (2.53-5.86)[a]	3.58 (2.32-5.53)[a]
80-89	50.58 (34.33-74.52)[a]	35.72 (24.04-53.08)[a]	5.20 (3.45-7.84)[a]	4.90 (3.20-7.50)[a]
>90	52.47 (34.72-79.30)[a]	41.58 (27.22-63.53)[a]	5.05 (3.18-8.03)[a]	4.78 (2.95-7.75)[a]
Gender: female	0.57 (0.49-0.66)[a]	0.56 (0.48-0.66)[a]	0.83 (0.69-0.99)[a]	0.78 (0.64-0.95)[b]
Chronic disease				
Respiratory disease	4.74 (3.76-5.97)[a]	2.42 (1.89-3.10)[a]	1.65 (1.25-2.17)[a]	1.48 (1.11-1.97)[b]
Cardiovascular disease	24.08 (13.51-42.90)[a]	8.66 (4.61-16.27)[a]	3.99 (2.16-7.36)[a]	3.39 (1.80-6.39)[a]
Renal disease	11.71 (9.06-15.12)[a]	4.19 (3.17-5.53)[a]	2.68 (2.00-3.60)[a]	2.25 (1.66-3.06)[a]
Diabetes	3.33 (2.68-4.14)[a]	1.30 (1.03-1.64)[b]	1.12 (0.87-1.45)[c]	0.95 (0.73-1.25)[c]
Cancer	3.008 (2.31-4.10)[a]	1.40 (1.03-1.89)[b]	0.95 (0.68-1.32)[c]	0.90 (0.64-1.27)[c]
Other comorbidity[d]	4.33 (3.54-5.30)[a]	2.32 (1.86-2.89)[a]	1.31 (1.03-1.66)[b]	1.24 (0.960-1.60)[c]
Any major comorbidity[e]	6.62 (5.69-7.72)[a]	—[f]	1.86 (1.54-2.24)[a]	—[f]

[a] $P<.001$

[b] $P<.05$

[c] Not significant.

[d] Other comorbidity includes other diseases collected in the official database of COVID-19 cases.

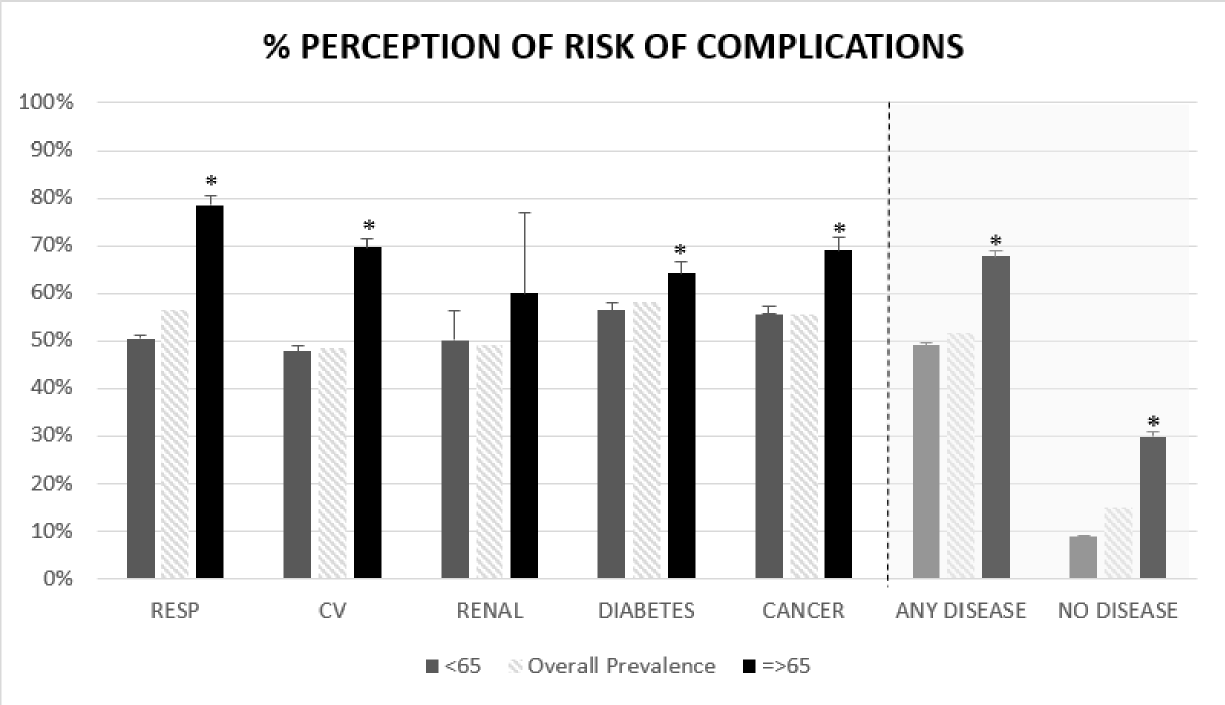
[e] Any major comorbidity: respiratory, cardiovascular, renal diseases; diabetes; or cancer.

[f] Cofactor not included in the model due to high variance inflation factor (VIF>5) to avoid multicollinearity.

Influence of Chronic Diseases on Risk Perception

We found that 23.9% (n=40,890) of the COVID-19 Barometer participants (n=171,087) considered themselves to be at high risk of developing a severe disease course in case of COVID-19 infection. This self-perception of risk was significantly higher among those aged ≥65 years (n=5235, 46.4%) and those suffering from any of the diseases under study (n=17,647, 51.6%; Figure 1). For those in both categories (ie, old age and comorbidities) that proportion rose to 67.7% (n=3212). Across all subgroups, the oldest people with respiratory diseases presented the highest self-perceived risk (n=1342, 78.5%), followed by cardiovascular disease (n=1403, 69.6%) and cancer (n=700, 69.0%) across the same age group (≥65 years). Younger individuals (<65 years) without any of the analyzed illnesses presented the lowest values (n=11,582, 8.9%; Figure 1).

Figure 1. Age-specific and standardized prevalence of self-perceived risk for developing severe disease outcomes following COVID-19 infection (n=171,087). "a" indicates a significant difference (*P*<.001) in terms of risk perception between those aged <65 years and ≥65 years. RESP: respiratory, CV: cardiovascular.



In the multivariable logistic regression, we observed a strong association between chronic diseases and self-perceived risk (Table 2), particularly for cancer (odds ratio [OR] 8.57, 95% CI 5.73-12.81), respiratory disease (OR 8.25, 95% CI 7.21-9.44), and diabetes (OR 6.17, 95% CI 4.58-8.31). Increasing age was also associated with self-perceived high risk, but it became nonsignificant for the oldest age categories in the multivariable model, likely due to the small sample size of those age groups (Multimedia Appendix 2). Females, lower education, smoking, and worse health status were also associated with self-perceived risk of severe COVID-19 disease (Table 2).

Table 2. Logistic regression (odds ratios [OR] and 95% CIs) to assess the association of chronic diseases with self-perceived risk to develop severe disease following COVID-19 infection (n=11,247).

Characteristic	Univariable, OR (95% CI)	Multivariable, OR (95% CI)
Age group (years; reference <50 years)		
50-59	1.84 (1.78 -1.90)[a]	1.74 (1.46-2.07)[a]
60-69	3.82 (3.68-3.96)[a]	2.64 (2.14-3.26)[a]
70-79	6.98 (6.57-7.41)[a]	3.85 (2.81-5.37)[a]
80-89	8.25 (6.68-10.17)[a]	1.35 (0.45-3.66)[b]
>90	9.19 (3.50-24.14)[a]	3.72 (0.34-47.06)[b]
Gender: female	0.88 (0.86-0.90)[a]	1.17 (1.02-1.33)[c]
Education (reference: basic/no education)		
Secondary school	0.68 (0.65-0.72)[a]	0.88 (0.73-1.06)[b]
University	0.53 (0.51-0.56)[a]	0.77 (0.64-0.93)[c]
Chronic disease		
Respiratory disease	7.20 (6.98-7.43)[a]	8.25 (7.21-9.44)[a]
Cardiovascular disease	5.67 (5.43-5.94)[a]	4.92 (3.69-6.55)[a]
Renal disease	5.83 (4.62-7.37)[a]	3.99 (2.67-5.98)[a]
Diabetes	6.93 (6.56-7.33)[a]	6.17 (4.58-8.31)[a]
Cancer	6.79 (6.37-7.22)[a]	8.57 (5.73-12.81)[a]
Any major comorbidity[d]	9.70 (9.43 -9.97)[a]	—[e]
Other comorbidity[f]	1.96 (1.89-2.04)[a]	3.29 (2.81-3.86)[a]
Smoking	1.17 (1.05-1.31)[c]	1.28 (1.12-1.45)[a]
Self-reported health status (worse)	8.24 (7.49-9.05)[a]	2.85 (2.09-3.90)[a]
Self-reported mental status (worse)	1.44 (1.25-1.65)[a]	—[g]
High-risk professional or living with one	0.99 (0.96-1.02)[b]	—[g]
Living alone	1.26 (1.21-1.30)[a]	—[g]
No social support	1.62 (1.56-1.69)[a]	—[g]
Lower confidence in the National Health Survey	1.13 (1.10-1.16)[a]	—[g]

[a]*P*<.001.

[b]Not significant.

[c]*P*<.05.

[d]Any major comorbidity: respiratory, cardiovascular, renal diseases; diabetes; or cancer.

[e]Cofactor not included in the model due to high VIF (>5) to avoid multicollinearity.

[f]Other comorbidity includes other diseases collected in the community-based COVID-19 Barometer survey.

[g]Cofactor excluded in the stepwise method (backward elimination with *P*>.05)

Discussion

Principal Findings

We found a significant association between chronic diseases (ie, respiratory, cardiovascular, and renal diseases; diabetes; and cancer) and COVID-19–related mortality and ICU admission. This was stronger for respiratory, cardiovascular, and renal diseases when analyzing only those COVID-19 cases requiring hospitalization. The overall self-reported prevalence of these illnesses among the country's population is 19.6%. However, these illnesses affects almost half of those aged ≥65 years (44.6%), a population vulnerable to COVID-19, namely because of age-related frailty and immune system decline (Multimedia Appendix 1) [23]. We also found that among this high-risk group, approximately two-thirds (67.7%) are self-aware of risks; this drops to about half for individuals above

65 years without any relevant chronic condition (46.4%). In fact, morbidity seems to be the strongest determinant of risk perception given the results of the multivariable regression. Furthermore, the inclusion of self-reported health status in the model did not affect these results, which suggests that this perception of risk is not so much altered by how the patient actually feels, but rather by the knowledge of having a chronic disease. This is corroborated with a low (and in some cases even absent) association of morbidity with self-perceived risk of infection (data not shown). The plausible risk of severe COVID-19 disease, and not so much the risk of infection posed by several chronic diseases, particularly among the elderly, was abundantly communicated by the media, medical societies, public health institutes, health authorities, and patient organizations. Therefore, it is not surprising that older patients with one of the analyzed illnesses were particularly concerned about the risk of developing severe outcomes following COVID-19, despite how active or controlled the disease is, or how well one feels.

Our study also found that other factors may contribute to self-perception of higher risk. Old age seems to be associated with an increased perception of risk, which is in line with other studies. For instance, a survey in the United States has shown that older adults perceive larger risks of dying if infected with COVID-19 [24]. Female gender also seems to be associated with a higher self-perception of risk. This finding aligns with other evidence showing that women tend to be more aware of their health status and seek health care more proactively than men [25-27]. Interestingly, the COVID-19 database analysis demonstrated that women were less likely to die or require the ICU in case of infection than men (despite more cases of infection among women, which is likely due to the prevalence of older women in the Portuguese population [28]). This is in line with findings reported elsewhere [29,30], and may be explained by other risk factors unequally distributed across the genders that have not been taken into consideration in this analysis. This finding is consistent with results from past surveys, which found an association between female gender and adoption of preventive behaviors during a pandemic respiratory disease [31-35]. Recently, two surveys performed in the United States showed that women are more knowledgeable about COVID-19 and engage in COVID-19 preventive behaviors more than men [17,36].

Higher education was associated with greater concern regarding the risk of severe COVID-19 disease. This supports other surveys and available data, which consistently show that education is linked with health literacy, awareness, and preventive behaviors [37,38]. On the other hand, the literature shows that lower education is associated with a greater risk of morbidity [39,40]. We thus foresee opportunities for patient education on COVID-19 targeting disadvantaged communities with a lower level of education, aggravated by lower income and reduced access to care, thereby mitigating the health inequities that are reportedly emphasized by COVID-19 [41].

It is worth noting that smokers were more likely to self-perceive high risk as well. Smokers are more susceptible to coronavirus complications, and this was thoroughly communicated in the media, thereby prompting a higher degree of concern in this group [42].

Several reports in the literature have documented the increased risks associated with comorbidities in patients infected with SARS-CoV-2–related viruses, such as the avian influenza [43-45], SARS-CoV (severe acute respiratory syndrome coronavirus) [46,47], and MERS-CoV (Middle East respiratory syndrome coronavirus) [48,49]. The most common health conditions with poorer prognosis included respiratory diseases [15,16], cardiac diseases [15,16], renal diseases [16], diabetes [18], hypertension [16], and cancer [15]. Initial reports from China suggested that these comorbidities could also play a negative role in the prognosis of COVID-19 infection [6,48,50], prompting health authorities and public health institutes, such as the Centers for Disease Control and Prevention, to act and declare these comorbidities as relevant risk factors [22]. However, some contradictory data were released that discussed alternative methodological approaches, including adjustments for potential confounders like age and gender. For instance, Wang and colleagues [51] conducted a meta-analysis, which highlighted hypertension, diabetes, chronic obstructive pulmonary disease (COPD), cardiovascular disease, and cerebrovascular disease as major risk factors for COVID-19, while ruling out cancer and renal disease. Other authors have claimed that cancer and renal disease are risk factors as well [52-54]. This inconsistency in the literature necessitates additional research on the relationship between morbidity and COVID-19 outcomes, as recently highlighted in a call for COVID-19 research [9].

Our data clearly show an independent association between respiratory, cardiovascular, and renal diseases and worse COVID-19 outcomes. Chronic diseases share several standard features with infectious disorders, such as the proinflammatory state, and the attenuation of the innate immune response, which may make individuals more susceptible to disease complications [55]. This is particularly true for cardiovascular diseases and an extensive discussion of this relationship with COVID-19 has been described elsewhere [56]. On the other hand, renal disease dysfunction causes reduced lymphocyte numbers and function, creating immunodeficiency and predisposing the individual to severe infections [57]. When it comes to underlying respiratory diseases, such as COPD, the patient's lung function is damaged and thus less resistant to viral infection and more disposed to develop serious disease [58]. This link has been presented elsewhere [27,48,59-61].

Our findings are very strong concerning diabetes and cancer since the multivariable model, which specifically focused on those hospitalized (less influenced by Berkson's bias, as discussed below), provided nonsignificant results for these pathologies. This is supported by some previous results [20], but not by others [24]. We cannot rule out that a lack of statistical power may have undermined our results.

Limitations and Strengths

There are some limitations to this study. First, the analysis is based on self-reported data, which might be subject to recall and misclassification biases (eg, chronic diseases were not clinically confirmed in the Barometer survey; differences in the case definitions across the databases used). Furthermore, it is possible that underreporting might have taken place among those who consult less and/or are less aware of their own chronic condition (eg, groups with limited education who lack health literacy and awareness). Secondly, the COVID-19 database is prone to Berkson's bias [62], given that any patient infected with COVID-19 and diagnosed with a chronic disease is more likely to be hospitalized than an infected case without a chronic disease, which might lead to spurious associations between the risk factors under study and serious COVID-19 outcomes. Furthermore, guidelines were issued recommending hospitalization of COVID-19 cases when some comorbidities were present, thereby worsening Berkson's bias [63]. This highlights the importance of analyzing the subgroup of hospitalizations that was done in this study. Thirdly, disease severity and staging were not taken into consideration, given that there was no such information in the data sets. Lastly, the Barometer survey was subjected to the volunteer bias (eg, more engaged and informed citizens completed the survey), thereby compromising the external validity of the analysis, and to social desirability bias. Although this sort of bias has been found to be lower in anonymous online surveys than in telephone or in-person surveys [64], we cannot rule out the possibility that some respondents reported more risk awareness than others due, in part, to social desirability [65]. We applied direct age and sex standardization to improve the external validity of these results.

This study has several strengths as well. It uses individual observations from two nationwide databases, including the official database with all COVID-19 cases in Portugal and a nationwide population-based survey that reached over 170,000 people, which, to our knowledge, makes it the world's largest community-based survey performed in the context of COVID-19 so far.

Policy Implications

Our results encourage authorities to protect those citizens at the highest risk to develop severe COVID-19 disease, as well as to promote knowledge and health literacy among those who, despite their increased risk, are not fully aware of it. In particular, older and uneducated men, a group with insufficient awareness, should be targeted by health policies to fight the pandemic threat effectively. Such policies should customize communication and foster preventive behaviors. Risk perception of pandemics can predict compliance with preventive measures and tendency to seek treatment or vaccination [66]. So far, social distancing and responsible behaviors have proven successful in preventing the spread of the disease, as well as its serious consequences [67]. Knowing how risk is perceived is essential for preparing an effective plan for risk communication, and may be predictive of the public's response [66,68]. As already mentioned, available literature shows that people with increased perception of risk are more likely to engage in protective behaviors [13,15-17].

Conclusions

Our study results demonstrate the association between some prevalent chronic diseases and increased risk of worse COVID-19 outcomes. It also provides further understanding on people's risk perceptions of serious COVID-19 disease. Hence, this study may aid health authorities to better adapt measures to the needs of the population and to identify those who are more vulnerable and require further education and information on preventive measures.

Acknowledgments

This study was cofunded by the Foundation for Science and Technology, under the financing program Research 4 Covid-19 (Ref FCT 608).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Prevalence of chronic diseases in the country's population.

[DOCX File , 30 KB - publichealth_v7i1e22794_app1.docx]

Multimedia Appendix 2

Patient characteristics.

[DOCX File , 16 KB - publichealth_v7i1e22794_app2.docx]

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
2. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020 Feb;91:264-266 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.009] [Medline: 31953166]
3. World Health Organization. COVID-19 Dashboard. 2020. URL: https://covid19.who.int/ [accessed 2020-12-10]

4. Guan W, Ni Z, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
5. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/s0140-6736(20)30211-7]
6. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061. [doi: 10.1001/jama.2020.1585]
7. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. The Lancet Infectious Diseases 2020 Apr;20(4):425-434. [doi: 10.1016/s1473-3099(20)30086-4]
8. Xu X, Wu X, Jiang X. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020 Feb 27;368:m792 [FREE Full text] [doi: 10.1136/bmj.m792] [Medline: 32107200]
9. Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
10. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
11. Fischhoff B. Risk perception and communication. In: Risk Analysis and Human Behavior. London, UK: Routledge; 2012:3-32.
12. Rosenstock IM. The Health Belief Model and Preventive Health Behavior. Health Education Monographs 1974 Dec 01;2(4):354-386. [doi: 10.1177/109019817400200405]
13. Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change1. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
14. Bruine de Bruin W, Bennett D. Relationships Between Initial COVID-19 Risk Perceptions and Protective Health Behaviors: A National Survey. Am J Prev Med 2020 Aug;59(2):157-167 [FREE Full text] [doi: 10.1016/j.amepre.2020.05.001] [Medline: 32576418]
15. Wise T, Zbozinek T, Michelini G, Hagan C, Mobbs D. Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. PsyArXiv. Preprint posted online March 19, 2020. [doi: 10.31234/osf.io/dz428]
16. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
17. Li S, Feng B, Liao W, Pan W. Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
18. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2020-12-10]
19. Wong CKH, Wong JYH, Tang EHM, Au CH, Lau KTK, Wai AKC. Impact of National Containment Measures on Decelerating the Increase in Daily New Cases of COVID-19 in 54 Countries and 4 Epicenters of the Pandemic: Comparative Observational Study. J Med Internet Res 2020 Jul 22;22(7):e19904 [FREE Full text] [doi: 10.2196/19904] [Medline: 32658858]
20. Inquérito Nacional de Saúde 2014 (National Health Survey - 2014). Instituto Nacional de Estatística. 2016. URL: https://www.ine.pt/xurl/pub/263714091> [accessed 2021-01-08]
21. Risk Communication and Community Engagement (RCCE) Action Plan Guidance COVID-19 Preparedness and Response. World Health Organization. 2020 Mar 16. URL: https://www.who.int/publications-detail/risk-communication-and-community-engagement-(rcce)-action-plan-guidance [accessed 2020-12-10]
22. People at Increased Risk: People at Increased Risk for Severe Illness. Centers for Disease Control and Prevention. 2020 Nov 30. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html [accessed 2020-12-10]
23. Yao X, Li H, Leng SX. Inflammation and immune system alterations in frailty. Clin Geriatr Med 2011 Feb;27(1):79-87 [FREE Full text] [doi: 10.1016/j.cger.2010.08.002] [Medline: 21093724]
24. Bruine de Bruin W. Age differences in COVID-19 risk perceptions and mental healthvidence from a national US survey conducted in March 2020. The Journals of Gerontology: Series B 2020 May:1-5 [FREE Full text] [doi: 10.1093/geronb/gbaa074]
25. Verbrugge LM. Sex differentials in health. Public Health Rep 1982;97(5):417-437 [FREE Full text] [Medline: 6750677]
26. Cleary PD, Mechanic D, Greenley JR. Sex Differences in Medical Care Utilization: An Empirical Investigation. Journal of Health and Social Behavior 1982 Jun;23(2):106. [doi: 10.2307/2136508]

27. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
28. COVID-19 - Boletim DGS. Directorate-General of Health. 2020 Nov 3. URL: https://covid19.min-saude.pt/relatorio-de-situacao/ [accessed 2020-12-11]
29. Yi Y, Lagniton PN, Ye S, Li E, Xu R. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci 2020;16(10):1753-1766 [FREE Full text] [doi: 10.7150/ijbs.45134] [Medline: 32226295]
30. Zhang J, Yu M, Tong S, Liu L, Tang L. Predictive factors for disease progression in hospitalized patients with coronavirus disease 2019 in Wuhan, China. J Clin Virol 2020 Jun;127:104392 [FREE Full text] [doi: 10.1016/j.jcv.2020.104392] [Medline: 32361327]
31. Lau JTF, Yang X, Tsui H, Kim JH. Monitoring community responses to the SARS epidemic in Hong Kong: from day 10 to day 62. J Epidemiol Community Health 2003 Nov 01;57(11):864-870 [FREE Full text] [doi: 10.1136/jech.57.11.864] [Medline: 14600111]
32. Leung G, Ho L, Chan S, Ho S, Bacon-Shone J, Choy R, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis 2005 Jun 15;40(12):1713-1720. [doi: 10.1086/429923]
33. Quah SR, Hin-Peng L. Crisis prevention and management during SARS outbreak, Singapore. Emerg Infect Dis 2004 Feb;10(2):364-368 [FREE Full text] [doi: 10.3201/eid1002.030418] [Medline: 15030714]
34. Tang CS, Wong C. Factors influencing the wearing of facemasks to prevent the severe acute respiratory syndrome among adult Chinese in Hong Kong. Prev Med 2004 Dec;39(6):1187-1193 [FREE Full text] [doi: 10.1016/j.ypmed.2004.04.032] [Medline: 15539054]
35. Chan EYY, Huang Z, Lo ESK, Hung KKC, Wong ELY, Wong SYS. Sociodemographic Predictors of Health Risk Perception, Attitude and Behavior Practices Associated with Health-Emergency Disaster Risk Management for Biological Hazards: The Case of COVID-19 Pandemic in Hong Kong, SAR China. Int J Environ Res Public Health 2020 May 29;17(11):3869 [FREE Full text] [doi: 10.3390/ijerph17113869] [Medline: 32485979]
36. Clements JM. Knowledge and Behaviors Toward COVID-19 Among US Residents During the Early Days of the Pandemic: Cross-Sectional Online Questionnaire. JMIR Public Health Surveill 2020 May 8;6(2):e19161. [doi: 10.2196/19161]
37. Feinstein L, Sabates R, Anderson TM, Sorhaindo A, Hammond C. What Are the Effects of Education on Health? Measuring the Effects of Education on Health and Civic Engagement. Proceedings of the Copenhagen Symposium / OECD 2006:117-354 [FREE Full text]
38. Abdelhafiz AS, Mohammed Z, Ibrahim ME, Ziady HH, Alorabi M, Ayyad M, et al. Knowledge, Perceptions, and Attitude of Egyptians Towards the Novel Coronavirus Disease (COVID-19). J Community Health 2020 Oct 21;45(5):881-890 [FREE Full text] [doi: 10.1007/s10900-020-00827-7] [Medline: 32318986]
39. Nagel G, Peter R, Braig S, Hermann S, Rohrmann S, Linseisen J. The impact of education on risk factors and the occurrence of multimorbidity in the EPIC-Heidelberg cohort. BMC Public Health 2008 Nov 11;8(1):384 [FREE Full text] [doi: 10.1186/1471-2458-8-384] [Medline: 19014444]
40. Laires PA, Perelman J. The current and projected burden of multimorbidity: a cross-sectional study in a Southern Europe population. Eur J Ageing 2019 Jun 1;16(2):181-192 [FREE Full text] [doi: 10.1007/s10433-018-0485-0] [Medline: 31139032]
41. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. The Lancet 2020 Apr;395(10232):1243-1244. [doi: 10.1016/s0140-6736(20)30893-x]
42. Emami A, Javanmardi F, Pirbonyeh N, Akbari A. Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis. Arch Acad Emerg Med 2020;8(1):e35 [FREE Full text] [Medline: 32232218]
43. Placzek HE, Madoff LC. Association of Age and Comorbidity on 2009 Influenza A Pandemic H1N1-Related Intensive Care Unit Stay in Massachusetts. Am J Public Health 2014 Nov;104(11):e118-e125. [doi: 10.2105/ajph.2014.302197]
44. Mauskopf J, Klesse M, Lee S, Herrera-Taracena G. The burden of influenza complications in different high-risk groups: a targeted literature review. J Med Econ 2013;16(2):264-277. [doi: 10.3111/13696998.2012.752376] [Medline: 23173567]
45. Martínez A, Soldevila N, Romero-Tamarit A, Torner N, Godoy P, Rius C, Surveillance of Hospitalized Cases of Severe Influenza in Catalonia Working Group. Risk factors associated with severe outcomes in adult hospitalized patients according to influenza type and subtype. PLoS One 2019 Jan 11;14(1):e0210353 [FREE Full text] [doi: 10.1371/journal.pone.0210353] [Medline: 30633778]
46. Booth CM. Clinical Features and Short-term Outcomes of 144 Patients With SARS in the Greater Toronto Area. JAMA 2003 Jun 04;289(21):2801. [doi: 10.1001/jama.289.21.joc30885]
47. Matsuyama R, Nishiura H, Kutsuna S, Hayakawa K, Ohmagari N. Clinical determinants of the severity of Middle East respiratory syndrome (MERS): a systematic review and meta-analysis. BMC Public Health 2016 Nov 29;16(1):1203 [FREE Full text] [doi: 10.1186/s12889-016-3881-4] [Medline: 27899100]
48. Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, China Medical Treatment Expert Group for COVID-19. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J 2020 May 26;55(5):2000547 [FREE Full text] [doi: 10.1183/13993003.00547-2020] [Medline: 32217650]
49. Garbati MA, Fagbo SF, Fang VJ, Skakni L, Joseph M, Wani TA, et al. A Comparative Study of Clinical Presentation and Risk Factors for Adverse Outcome in Patients Hospitalised with Acute Respiratory Disease Due to MERS Coronavirus or

Other Causes. PLoS One 2016 Nov 3;11(11):e0165978 [FREE Full text] [doi: 10.1371/journal.pone.0165978] [Medline: 27812197]
50. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020 May;109(5):531-538 [FREE Full text] [doi: 10.1007/s00392-020-01626-9] [Medline: 32161990]
51. Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY) 2020 Apr 08;12(7):6049-6057 [FREE Full text] [doi: 10.18632/aging.103000] [Medline: 32267833]
52. Tian J, Yuan X, Xiao J, Zhong Q, Yang C, Liu B, et al. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. The Lancet Oncology 2020 Jul;21(7):893-903. [doi: 10.1016/s1470-2045(20)30309-0]
53. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020 May 22;369:m1985 [FREE Full text] [doi: 10.1136/bmj.m1985] [Medline: 32444460]
54. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, STOP-COVID Investigators. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med 2020 Jul 15. [doi: 10.1001/jamainternmed.2020.3596] [Medline: 32667668]
55. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
56. Driggin E, Madhavan M, Bikdeli B, Chuich T, Laracy J, Biondi-Zoccai G, et al. Cardiovascular Considerations for Patients, Health Care Workers, and Health Systems During the COVID-19 Pandemic. J Am Coll Cardiol 2020 May 12;75(18):2352-2371 [FREE Full text] [doi: 10.1016/j.jacc.2020.03.031] [Medline: 32201335]
57. Kato S, Chmielewski M, Honda H, Pecoits-Filho R, Matsuo S, Yuzawa Y, et al. Aspects of Immune Dysfunction in End-stage Renal Disease. CJASN 2008 Aug 13;3(5):1526-1533. [doi: 10.2215/cjn.00950208]
58. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical and mortal COVID-19 cases: A systematic literature review and meta-analysis. J Infect 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.021] [Medline: 32335169]
59. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ 2020 Mar 26;368:m1198. [doi: 10.1136/bmj.m1198] [Medline: 32217618]
60. Liu W, Tao Z, Wang L, Yuan M, Liu K, Zhou L, et al. Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. Chinese Medical Journal 2020;133(9):1032-1038. [doi: 10.1097/cm9.0000000000000775]
61. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020 May;97(5):829-838 [FREE Full text] [doi: 10.1016/j.kint.2020.03.005] [Medline: 32247631]
62. Berkson J. Limitations of the Application of Fourfold Table Analysis to Hospital Data. Biometrics Bulletin 1946 Jun;2(3):47. [doi: 10.2307/3002000]
63. Norma no 004/2020 de 23/03/2020 atualizada a 25/04/2020. Direção-Geral da Saúde. 2020. URL: https://covid19.min-saude.pt/wp-content/uploads/2020/12/Norma-004_2020.pdf [accessed 2021-01-08]
64. Keeter S. From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls. Pew Research Center. 2015 May 13. URL: https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/ [accessed 2021-01-08]
65. Phillips DL, Clancy KJ. Some Effects of "Social Desirability" in Survey Studies. American Journal of Sociology 1972 Mar;77(5):921-940. [doi: 10.1086/225231]
66. Leung GM, Lam TH, Ho LM, Ho SY, Chan BHY, Wong IOL, et al. The impact of community psychological responses on outbreak control for severe acute respiratory syndrome in Hong Kong. J Epidemiol Community Health 2003 Nov 01;57(11):857-863 [FREE Full text] [doi: 10.1136/jech.57.11.857] [Medline: 14600110]
67. Davies N, Kucharski A, Eggo R, Gimma A, Edmunds W, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. The Lancet Public Health 2020 Jul 02;5(7):e375-e385. [doi: 10.1016/S2468-2667(20)30133-X]
68. Karasneh R, Al-Azzam S, Muflih S, Soudah O, Hawamdeh S, Khader Y. Media's effect on shaping knowledge, awareness risk perceptions and communication practices of pandemic COVID-19 among pharmacists. Res Social Adm Pharm 2021 Jan;17(1):1897-1902 [FREE Full text] [doi: 10.1016/j.sapharm.2020.04.027] [Medline: 32340892]

Abbreviations

COPD: chronic obstructive pulmonary disease
DGS: Direção-Geral da Saúde
ENSP-NOVA: National School of Public Health

ICU: intensive care unit
INS: Inquérito Nacional de Saúde
MERS-CoV: Middle East respiratory syndrome coronavirus
OR: odds ratio
SARS-CoV: severe acute respiratory syndrome coronavirus
SINAVE: Sistema Nacional de Vigilância Epidemiológica

Edited by G Eysenbach; submitted 28.07.20; peer-reviewed by W Zhang, P Banik; comments to author 21.11.20; revised version received 13.12.20; accepted 14.12.20; published 12.01.21.

Please cite as:
Laires PA, Dias S, Gama A, Moniz M, Pedro AR, Soares P, Aguiar P, Nunes C
The Association Between Chronic Disease and Serious COVID-19 Outcomes and Its Influence on Risk Perception: Survey Study and Database Analysis
JMIR Public Health Surveill 2021;7(1):e22794
URL: http://publichealth.jmir.org/2021/1/e22794/
doi:10.2196/22794
PMID:33433397

Original Paper

YouTube Videos Demonstrating the Nasopharyngeal Swab Technique for SARS-CoV-2 Specimen Collection: Content Analysis

Kyohei Itamura[1], MD; Arthur Wu[1,2], MD; Elisa Illing[2], MD; Jonathan Ting[2], MSc, MD, MBA; Thomas Higgins[3,4], MD, MSPH

[1]Division of Otolaryngology - Head and Neck Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[2]Department of Otolaryngology - Head and Neck Surgery, Indiana University, Indianapolis, IN, United States

[3]Department of Otolaryngology - Head and Neck Surgery and Communicative Disorders, University of Louisville, Louisville, KY, United States

[4]Rhinology, Sinus & Skull Base, Kentuckiana Ear, Nose, and Throat, Louisville, KY, United States

Corresponding Author:
Thomas Higgins, MD, MSPH
Department of Otolaryngology - Head and Neck Surgery and Communicative Disorders
University of Louisville
6420 Dutchman's Parkway, STE 380
Louisville, KY, 40205
United States
Phone: 1 502 894 8441
Email: thomas.higgins@louisville.edu

Abstract

Background: Real-time polymerase chain reaction using nasopharyngeal swabs is currently the most widely used diagnostic test for SARS-CoV-2 detection. However, false negatives and the sensitivity of this mode of testing have posed challenges in the accurate estimation of the prevalence of SARS-CoV-2 infection rates.

Objective: The purpose of this study was to evaluate whether technical and, therefore, correctable errors were being made with regard to nasopharyngeal swab procedures.

Methods: We searched a web-based video database (YouTube) for videos demonstrating SARS-CoV-2 nasopharyngeal swab tests, posted from January 1 to May 15, 2020. Videos were rated by 3 blinded rhinologists for accuracy of swab angle and depth. The overall score for swab angle and swab depth for each nasopharyngeal swab demonstration video was determined based on the majority score with agreement between at least 2 of the 3 reviewers. We then comparatively evaluated video data collected from YouTube videos demonstrating the correct nasopharyngeal swab technique with data from videos demonstrating an incorrect nasopharyngeal swab technique. Multiple linear regression analysis with statistical significance set at P=.05 was performed to determine video data variables associated with the correct nasopharyngeal swab technique.

Results: In all, 126 videos met the study inclusion and exclusion criteria. Of these, 52.3% (66/126) of all videos demonstrated the correct swab angle, and 46% (58/126) of the videos demonstrated an appropriate swab depth. Moreover, 45.2% (57/126) of the videos demonstrated both correct nasopharyngeal swab angle and appropriate depth, whereas 46.8% (59/126) of the videos demonstrated both incorrect nasopharyngeal swab angle and inappropriate depth. Videos with correct nasopharyngeal swab technique were associated with the swab operators identifying themselves as a medical professional or as an Ear, Nose, Throat–related medical professional. We also found an association between correct nasopharyngeal swab techniques and recency of video publication date (relative to May 15, 2020).

Conclusions: Our findings show that over half of the videos documenting the nasopharyngeal swab test showed an incorrect technique, which could elevate false-negative test rates. Therefore, greater attention needs to be provided toward educating frontline health care workers who routinely perform nasopharyngeal swab procedures.

(JMIR Public Health Surveill 2021;7(1):e24220) doi:10.2196/24220

KEYWORDS

COVID-19; coronavirus; SARS-coV-2; nasopharyngeal swab; viral testing; PCR; YouTube; infodemiology; digital epidemiology; testing; diagnostic; content analysis; video; error

Introduction

Qualitative real-time polymerase chain reaction (RT-PCR) of nasopharyngeal secretions is the gold standard for testing respiratory viruses, including SARS-CoV-2 [1]. However, major concerns have been raised regarding false-negative rates of RT-PCR tests in community testing locations [2]. An early retrospective review of community hospital testing performed in China reported a test sensitivity of only 71% [3]. Although the false-negative results could be attributed to various reasons, including laboratory errors, patient misidentification, and inadequate collection of specimens, improper technique resulting in the swabs not reaching the target site of the nasopharynx is a potentially pervasive but modifiable error.

The trajectory from the nostril to the nasopharynx is often presumed to be along the dorsum of the nose, likely because of the visual appearance of the external nose. In reality, the correct trajectory is along the floor of the nose in the direction back toward the ear. Deviating from this trajectory can lead to pain from contacting a deviated septum or nasal turbinate or failure to reach the nasopharynx. Although the Center for Disease Control and Prevention has provided guidance regarding the proper nasopharyngeal swab (NPS) technique, vivid descriptions of painful patient experiences are currently commonplace in the media [4].

Although many health centers around the world are likely providing proper training to frontline health care workers, there is concern that improper NPS techniques for specimen collection may lead to false-negative results in RT-PCR tests [5]. This is a significant concern, as false-negative test results underestimate the prevalence of COVID-19 and give a false sense of security to patients as well as the health care workers caring for them [6]. Moreover, the use of improper NPS techniques limits public health efforts in identifying and contact tracing the spread of the virus. Thus, with the widely established use of NPS as a large-scale screening tool for COVID-19 and other respiratory viral diseases, ensuring a proper collection technique is used is essential in yielding sensitive test results [7],

Accordingly, the purpose of this study was to determine how the NPS technique for SARS-CoV-2 is instructed or demonstrated and how the NPS test is administered in real life by reviewing videos hosted on the web-based video-sharing platform YouTube (Google, LLC) [8].

Methods

Sample Size Determination

The sample size calculation was performed using the Kelsey methodology for cross-sectional study with a power of 80% and 2-sided confidence level of 0.05%; the exposure ratio was estimated to be 1:1 and odds ratio was estimated to be 3. The total estimated sample size was 64.

YouTube Database Search

YouTube is a widely used social media database of videos uploaded by the general public. Due to the broad accessibility of this database, there was no requirement for research approval by the Institutional Review Board–Human Subjects. The terms "nasopharyngeal swab," "nasopharyngeal test," "nasal swab," "coronavirus swab," "coronavirus test,""covid swab," and "covid test" were used to query the YouTube video database [8]. The query was filtered by setting the "sort by" filter to "by upload date" to compile all videos published from January 1, 2020, to May 15, 2020. Inclusion and exclusion criteria were defined to screen all search results. The inclusion criterion was that the NPS test is performed on screen with visualization of swab insertion into either naris. The exclusion criteria were duplicate videos, non–COVID-19 swab indication, and swab testing intended for anatomical regions other than the nasopharynx (eg, anterior nasal swabbing).

Video Evaluation and Data Collection

Three faculty rhinologists individually reviewed the selected NPS demonstration videos for swab angle and swab depth. Swab angle was assessed as either "along the nasal floor" or "not along the nasal floor." Swab depth was assessed as either "appropriate depth" or "inappropriate depth." All reviewers were blinded to each other's assessments. The following ancillary video data were collected: video author type ("medical," including physician, registered nurse, physician's assistant, or nurse practitioner, vs "nonmedical"), operator type ("medical" vs "nonmedical"), type of video ("instructional vs "noninstructional"), specialty ("otolaryngology" vs "other"), country of origin ("United States [USA]" vs "other"), number of likes, number of author subscribers, time in nasal cavity, time at nasopharynx, and video post date relative to May 15, 2020.

Statistical Analysis

Interrater reliability among the 3 reviewers was assessed using Fleiss' Kappa. The overall score for swab angle and swab depth for each NPS demonstration video was determined based on the majority score with agreement between at least 2 of the 3 reviewers. Video data were also compared between YouTube videos demonstrating the correct NPS technique and those demonstrating an incorrect NPS technique. Multiple linear regression analysis was performed to determine predictive variables among video data for videos demonstrating the correct NPS technique. Statistical significance was set at P=.05. All statistical analyses were performed on Microsoft Excel (Microsoft Corp.).

Results

The final qualitative analysis included 126 independent, unique videos. The video selection process, including screening for inclusion and exclusion criteria, is summarized in Figure 1.

The κ value indicating interrater reliability for the 3 reviewers was 0.66 for swab angle (P<.001; 95% CI=0.56-0.76) and 0.68 for swab depth (P<.001; 95% CI=0.58-0.78). For the assessment of swab angle, there was complete agreement among all reviewers, with all 3 scores consistent for 74.6% (94/126) of all videos. For the assessment of swab depth, there was complete agreement among all 3 reviewers for 76.1% (96/126) of all videos.

Moreover, we found that 52.3% (66/126) of all NPS demonstration videos had the correct angle, and 46% (58/126) showed appropriate depth (Figure 2). In addition, 45.2% (57/126) of all videos had both correct NPS angle and appropriate depth, whereas 46.8% (59/126) of the videos had both incorrect NPS angle and inappropriate depth. We observed concordance between the swab angle and depth (ie, correct swab angle with appropriate swab depth or incorrect swab angle with inappropriate swab depth) in 92% (116/126) of the videos. The agreement between these measures was nearly equivalent with regard to both measures being correct compared with both measures being incorrect. In the remaining approximately 8% (10/126) of the videos, 8 videos demonstrated correct swab angle but inappropriate swab depth, and the remaining 2 videos demonstrated incorrect swab angle but appropriate swab depth.

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

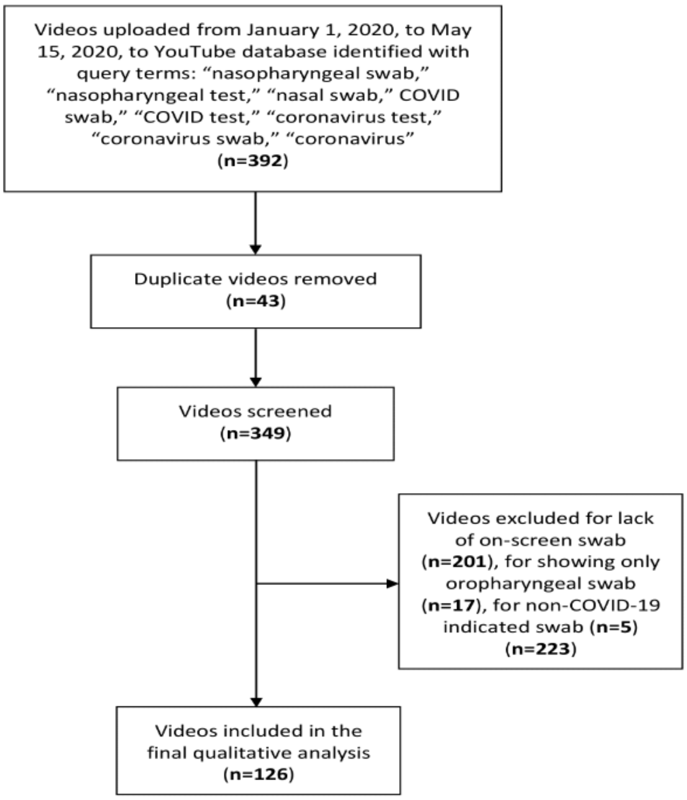


Figure 2. Percentage of YouTube videos demonstrating correct or incorrect nasopharyngeal swab angle and/or appropriate or inappropriate nasopharyngeal swab depth. NPS: nasopharyngeal swab.

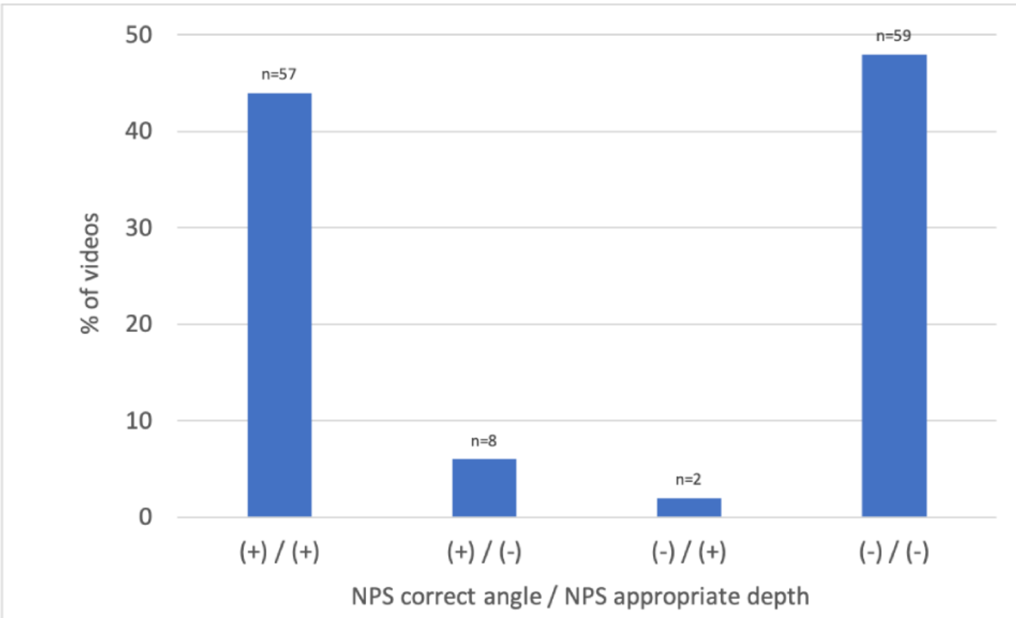


Table 1 compares the video data between videos demonstrating NPS with correct technique (both correct swab angle and appropriate swab depth) and those demonstrating NPS with incorrect technique (incorrect swab angle and/or inappropriate swab depth). We found that approximately half of the videos with the correct NPS technique were from a medical video author, instructional in nature, and/or of US origin. Videos demonstrating the correct NPS technique also were posted onto YouTube closer to May 15, 2020, compared to those demonstrating an incorrect technique. ENT-related providers were only found in videos demonstrating the correct NPS technique. Video viewership metrics, including number of views, likes, and author subscriber count, varied widely between videos demonstrating correct and incorrect NPS techniques. For correctly performed NPS techniques, the median time at the nasopharynx was 4 seconds. By definition, no time was spent at the nasopharynx in incorrectly performed NPS techniques, with regard to both incorrect depth and/or angle.

Table 1. Comparison of video data of YouTube videos demonstrating correct nasopharyngeal swab technique with those demonstrating incorrect nasopharyngeal swab technique

Data type	Correct NPS[a] technique (n=57)	Incorrect NPS technique (n=69)
Video author type, n (%)		
Medical (academic/nonacademic physician, nonphysician health care worker, medical group/entity)	29 (51)	27 (39)
Nonmedical (media, layperson)	28 (49)	42 (61)
Swab operator type, n (%)		
Medical (explicit identification as MD, physician's assistant, nurse practitioner, registered nurse)	55 (96)	59 (86)
Unidentified	2 (4)	10 (14)
Video type, n (%)		
Instructional (for teaching purposes, demonstrational)	32 (56)	33 (48)
Real-world test	25 (44)	36 (52)
Swab operator specialty, n (%)		
ENT (ear, nose, throat)	9 (16)	0 (0)
Non-ENT (emergency medicine, primary care, unidentified)	48 (84)	69 (100)
Swab collection location, n (%)		
Drive-through	15 (26)	26 (38)
Non drive-through (clinic, urgent/emergency care, walk-in testing site)	42 (74)	43 (62)
Video country of origin, n (%)		
USA	29 (51)	41 (60)
Rest of the world	28 (49)	28 (40)
Time for which the swab was inserted in the nose (seconds), median (IQR)	11 (7, 16)	8 (6, 13)
Number of views, median (IQR)	111.5 (883, 3502)	863 (185, 36618)
Number of likes, median (IQR)	11 (0, 38.5)	9 (1.5, 106)
Percentage of likes over total views, median (IQR)	0.6 (0, 1.7)	0.5 (0.2, 1.3)
Number of channel subscribers, median (IQR)	309 (19.5, 2285)	1690 (24.5, 20,700)
Video post date (number of days before May 15, 2020), median (IQR)	22 (7, 38)	32 (16, 42)
Time at nasopharynx (seconds), median (IQR)	4 (1.5, 7)	N/A[b]

[a]NPS: nasopharyngeal swab.

[b]N/A: not applicable.

Multiple linear regression analysis was performed on video data for videos demonstrating the correct NPS technique as the reference dependent variable (Table 2). The correct NPS technique was associated with the NPS operator identifying as a medical professional and, additionally, as a provider within the ENT specialty. There was also a significant association between correct NPS technique and recency of video post date relative to May 15, 2020.

Table 2. Multiple linear regression analysis of video data (reference dependent variable: correct nasopharyngeal technique). Italicized values indicate statistical significance.

Data type	Standardized β coefficient	95% confidence interval	*P* value
Author type (ref: medical)	−.079	−0.34 to 0.19	.55
Swab operator (ref: medical)	.23	0.040-0.41	*.018*
Video type (ref: instructional)	.046	−0.24 to 0.34	.75
Specialty (ref: ENT)	.28	0.11-0.46	*.002*
Testing location (ref: drive-through)	−.094	−0.30 to 0.11	.37
Country (ref: USA)	−.032	−0.22 to 0.15	.74
Longer time in nose	−.0074	−0.26 to 0.11	.42
Number of video views	.026	−0.58 to 0.63	.93
Number of video likes	−.072	−0.68 to 0.53	.82
Ratio of likes over views	.024	−0.15 to 0.20	.80
Number of video channel subscribers	.16	−0.011 to 0.33	.07
Video post date (Number of days before May 15, 2020)	.27	0.076-0.46	*.007*

Discussion

Principal Findings

This study demonstrates that almost half (46.8%) of all NPS demonstration videos reviewed showed incorrect swab angle and inappropriate depth, as judged by 3 rhinologists. The noninstructional videos provide a broad vantage point of how testing is actually performed in the real world, so it is alarming to find that the proper technique may be so infrequently used. Moreover, 51% of all instructional videos demonstrated an improper technique. Although it is unknown how many of these instructional videos are actually being used by viewers for the purposes of learning the NPS technique, it still highlights the fact that those claiming to be experts, whether local, national or otherwise, may not have a complete understanding of the NPS technique. Furthermore, there were no statistically significant differences between the viewership of videos demonstrating the correct technique and those demonstrating incorrect NPS techniques. This finding is consistent with previously published work, especially in the early months of the COVID-19 pandemic showing the pervasiveness of uncredentialed, low-quality media publicly available on the internet [9]. These results are not totally unexpected given the lack of sinonasal anatomic knowledge most NPS operators have and the inherent difficulty of navigating the septum and turbinates to reach the nasopharynx. It is noteworthy that all videos of otolaryngologists performing or instructing the NPS were done correctly. These findings emphasize the onus of the otolaryngology field to educate our colleagues on sinonasal anatomy and proper NPS technique during the ongoing COVID-19 pandemic.

Although lower respiratory samples such as bronchoalveolar lavage and sputum samples have higher viral loads in patients with SARS-CoV-2, NPS testing is considered the best alternative over other minimally invasive specimen collection options such as oropharyngeal swabs, blood samples, or stool samples [1]. NPS is widely used to test for other respiratory viral infections and has supplanted nasopharyngeal aspirates for its accuracy and convenience in this setting [10]. However, poor techniques used for the NPS method may convert this test into a simple nasal swab. The NPS test is inherently uncomfortable for the patient even with good technique, and patients or the NPS operator may retract prematurely before the swab reaches the correct location and saturates with mucus. There has been limited attention paid towards the impact proper technique has on the accuracy of results in NPS testing even with regard to influenza or other respiratory viral testing.

Although a comparison of viral loads between the nasal cavity and the nasopharynx has not been reported for SARS-CoV-2, significant differences in viral loads have been demonstrated for other viruses; nevertheless, this is not direct evidence that NPS technique may affect testing accuracy [11,12]. However, NPS sampling technique has been shown to potentially affect false-negative rates in SARS-CoV-2 in a single study [13]. Ma et al [14] demonstrated significantly improved test accuracy and patient comfort when patients were tested in a supine position compared to in a sitting position. Of note, an increasing number of tests are performed via drive-through testing to provide convenience, increase throughput, and adhere to social distancing recommendations [14]. Despite the rapid adoption of this modality, there has not been substantial review of its effect on testing accuracy, and patient and operator positioning may not be optimized for proper NPS technique.

Because of these multiple unknowns in the current SARS-CoV-2 testing climate, we feel that specimen collection technique is at least one aspect that could be easily remedied. Three points need to be emphasized to frontline health care workers performing the NPS technique: trajectory angle, depth, and patient expectations. The swab should be angled to follow the floor of the nose, and the depth required to reach the nasopharynx, approximately 9-10 cm in adults, is often surprising to non-otolaryngologists [7]. In many cases, this means that almost the entire length of the swab is inserted into the nasal cavity with only a small portion left to hold outside the nose. Finally, both the patient and the operator should set proper expectations for the procedure: the NPS is uncomfortable but should not cause sharp or severe pain. Such discomfort should indicate to the operator that anatomic obstruction such as a deviated septum is occluding the pathway, and a modified trajectory or contralateral approach should then be attempted.

Study Limitations

This study has some limitations, including the lack of consistent video angle and quality. The types of videos ranged from professionally produced instructional videos to "selfies" posted by patients themselves. Although this inconsistency could potentially introduce difficulty in the judging NPS technique being demonstrated, we excluded videos that were clearly difficult to be analyzed. Nevertheless, although a moderate level of interrater agreement was demonstrated by Fleiss' Kappa analysis, the remaining discordance may be attributable to the large variety in video quality. We believe that the inclusion of a wide array of video types would allow for a more complete view of real-life NPS testing across the globe. Additionally, there was no way to correlate proper or improper NPS technique with false-positive or false-negative testing rates. Finally, reproducibility of the results may be limited due to the involvement of only 3 reviewers, although all of them were board-certified otolaryngologists with fellowship training in rhinology.

Conclusions

The majority of NPS demonstration videos evaluated in this study used an improper technique. This technical deficiency may affect false-negative rates for SARS-CoV-2 testing. Therefore, based on these findings, we suggest that otolaryngologists should work to educate their medical colleagues and frontline health care workers who perform NPS techniques about relevant anatomy and technical considerations.

Conflicts of Interest

None declared.

References

1. Torretta S, Zuccotti G, Cristofaro V, Ettori J, Solimeno L, Battilocchi L, et al. Diagnosis of SARS-CoV-2 by RT-PCR using different sample sources: review of the literature. Ear Nose Throat J 2020 Aug 31:014556132095323 [FREE Full text] [doi: 10.1177/0145561320953231] [Medline: 32865458]
2. West CP, Montori VM, Sampathkumar P. COVID-19 testing: the threat of false-negative results. Mayo Clin Proc 2020 Jun;95(6):1127-1129 [FREE Full text] [doi: 10.1016/j.mayocp.2020.04.004] [Medline: 32376102]
3. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 2020 Aug;296(2):E115-E117 [FREE Full text] [doi: 10.1148/radiol.2020200432] [Medline: 32073353]
4. Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens for COVID-19. Centers for Disease Control and Prevention - COVID-19. URL: https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html [accessed 2020-12-16]
5. Tang Y, Schmitz JE, Persing DH, Stratton CW. Laboratory diagnosis of COVID-19: current issues and challenges. J Clin Microbiol 2020 May 26;58(6) [FREE Full text] [doi: 10.1128/JCM.00512-20] [Medline: 32245835]
6. Lippi G, Simundic A, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). Clin Chem Lab Med 2020 Jun 25;58(7):1070-1076. [doi: 10.1515/cclm-2020-0285] [Medline: 32172228]
7. Marty FM, Chen K, Verrill KA. How to Obtain a Nasopharyngeal Swab Specimen. N Engl J Med 2020 May 28;382(22):e76. [doi: 10.1056/nejmvcm2010260]
8. YouTube. URL: https://www.youtube.com/ [accessed 2021-01-11]
9. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
10. Tunsjø HS, Berg AS, Inchley CS, Røberg IK, Leegaard TM. Comparison of nasopharyngeal aspirate with flocked swab for PCR-detection of respiratory viruses in children. APMIS 2015 Jun;123(6):473-477. [doi: 10.1111/apm.12375] [Medline: 25904242]
11. Ohrmalm L, Wong M, Rotzén-Östlund M, Norbeck O, Broliden K, Tolfvenstam T. Flocked nasal swab versus nasopharyngeal aspirate for detection of respiratory tract viruses in immunocompromised adults: a matched comparative study. BMC Infect Dis 2010 Nov 26;10:340 [FREE Full text] [doi: 10.1186/1471-2334-10-340] [Medline: 21110854]
12. Coghill AE, Wang C, Verkuijilen SA, Yu KJ, Hsu W, Middeldorp JM, et al. Evaluation of nasal and nasopharyngeal swab collection for the detection of Epstein-Barr virus in nasopharyngeal carcinoma. J Med Virol 2018 Jan;90(1):191-195. [doi: 10.1002/jmv.24918] [Medline: 28833336]
13. Ma SY, Luo YM, Hu TY, You ZC, Sun JG, Yu SY, et al. Clinical application effect of modified nasopharyngeal swab sampling for 2019 novel coronavirus nucleic acid detection (Article in Chinese). Zhonghua Shao Shang Za Zhi 2020 Aug 20;36(8):679-685. [doi: 10.3760/cma.j.cn501120-20200312-00153] [Medline: 32268456]
14. Nundy S, Patel KK. Self-service diagnosis of COVID-19—ready for prime time? JAMA Health Forum 2020 Mar 16;1(3):e200333. [doi: 10.1001/jamahealthforum.2020.0333]

Abbreviations

ENT: ear, nose, throat
NPS: nasopharyngeal swab
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
RT-PCR: reverse transcription–polymerase chain reaction

Edited by T Sanchez; submitted 09.09.20; peer-reviewed by K Zeraatkar, A Rovetta; comments to author 29.10.20; revised version received 30.12.20; accepted 04.01.21; published 14.01.21.

Please cite as:
Itamura K, Wu A, Illing E, Ting J, Higgins T
YouTube Videos Demonstrating the Nasopharyngeal Swab Technique for SARS-CoV-2 Specimen Collection: Content Analysis
JMIR Public Health Surveill 2021;7(1):e24220
URL: http://publichealth.jmir.org/2021/1/e24220/
doi:10.2196/24220
PMID:33406478

Original Paper

Nowcasting for Real-Time COVID-19 Tracking in New York City: An Evaluation Using Reportable Disease Data From Early in the Pandemic

Sharon K Greene[1], PhD, MPH; Sarah F McGough[2,3], PhD; Gretchen M Culp[4], PhD; Laura E Graf[1], MPH; Marc Lipsitch[2], DPhil; Nicolas A Menzies[5], PhD; Rebecca Kahn[2], MS, PhD

[1]Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States
[2]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, United States
[3]Genentech, Inc, South San Francisco, CA, United States
[4]Bureau of Epidemiology Services, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States
[5]Department of Global Health and Population, Harvard TH Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Sharon K Greene, PhD, MPH
Bureau of Communicable Disease
New York City Department of Health and Mental Hygiene
42-09 28th Street
CN 22A, WS 06-154
Long Island City, NY, 11101
United States
Phone: 1 347 396 2679
Email: sgreene4@health.nyc.gov

Abstract

Background: Nowcasting approaches enhance the utility of reportable disease data for trend monitoring by correcting for delays, but implementation details affect accuracy.

Objective: To support real-time COVID-19 situational awareness, the New York City Department of Health and Mental Hygiene used nowcasting to account for testing and reporting delays. We conducted an evaluation to determine which implementation details would yield the most accurate estimated case counts.

Methods: A time-correlated Bayesian approach called Nowcasting by Bayesian Smoothing (NobBS) was applied in real time to line lists of reportable disease surveillance data, accounting for the delay from diagnosis to reporting and the shape of the epidemic curve. We retrospectively evaluated nowcasting performance for confirmed case counts among residents diagnosed during the period from March to May 2020, a period when the median reporting delay was 2 days.

Results: Nowcasts with a 2-week moving window and a negative binomial distribution had lower mean absolute error, lower relative root mean square error, and higher 95% prediction interval coverage than nowcasts conducted with a 3-week moving window or with a Poisson distribution. Nowcasts conducted toward the end of the week outperformed nowcasts performed earlier in the week, given fewer patients diagnosed on weekends and lack of day-of-week adjustments. When estimating case counts for weekdays only, metrics were similar across days when the nowcasts were conducted, with Mondays having the lowest mean absolute error of 183 cases in the context of an average daily weekday case count of 2914.

Conclusions: Nowcasting using NobBS can effectively support COVID-19 trend monitoring. Accounting for overdispersion, shortening the moving window, and suppressing diagnoses on weekends—when fewer patients submitted specimens for testing—improved the accuracy of estimated case counts. Nowcasting ensured that recent decreases in observed case counts were not overinterpreted as true declines and supported officials in anticipating the magnitude and timing of hospitalizations and deaths and allocating resources geographically.

(JMIR Public Health Surveill 2021;7(1):e25538) doi:10.2196/25538

KEYWORDS

COVID-19; data quality; epidemiology; forecasting; infectious disease; morbidity and mortality trends; public health practice; surveillance

Introduction

Timeliness is a key attribute of surveillance systems for reportable infectious diseases [1,2]. Timely surveillance data for COVID-19 are used by governments and communities to allocate resources and to decide when to tighten or loosen physical distancing and other prevention measures [3,4]. However, public health authorities track reportable diseases at a lag, given delays from infection to symptom onset, care seeking, specimen collection, laboratory testing, and reporting [5]. Monitoring prediagnostic data sources (eg, emergency department syndromic surveillance [6], internet searches and social media [7], participatory surveillance of self-reported symptoms [8], smart thermometers [9], etc) can improve timeliness at the expense of specificity, such as an inability to distinguish increases in respiratory illness attributable to influenza from COVID-19. Another approach that preserves specificity when monitoring COVID-19 disease trends is to leverage partially reported disease data, formally accounting for data lags.

The terms *nowcasting*, or predicting the present, and *hindcasting*, or predicting through the day prior to the present, describe a wide range of statistical adjustments used to fill in cases that are not yet reported, offering health officials a more up-to-date picture for situational awareness [10]. For example, researchers have assessed the potential to nowcast COVID-19 cases and deaths using Google Trends data available in near-real time [11], and have applied a range of modeling approaches that leverage reporting delays to estimate the number of not-yet-reported cases and deaths [12,13]. Using mathematical models to exploit COVID-19 transmission dynamics, nowcasting also has been extended to COVID-19 forecasting systems [14,15]. In a majority of these approaches, the nowcasting mechanism relies on accurately estimating the distribution of reporting delays; however, infectious disease transmission contains an important temporal component, in that incidence is correlated from one time point to the next, which has also been shown to improve nowcasting performance, including in COVID-19 applications [10,16].

We describe the use and evaluation of a time-correlated Bayesian nowcasting approach at the New York City (NYC) Department of Health and Mental Hygiene (DOHMH) during the first epidemic wave of COVID-19 to support real-time situational awareness and resource allocation. During the period from March to May 2020, approximately 203,000 laboratory-confirmed COVID-19 cases were reported to NYC DOHMH, peaking during the week of March 29, with approximately 5100 cases diagnosed per day [17]. Testing rates increased during this period as testing criteria at public health laboratories were relaxed, commercial and hospital laboratories developed testing capacity, and additional testing sites were opened and promoted [17].

Methods

Reportable Disease Surveillance Data

Persons Tested

Clinical and commercial laboratories are required to report all results, including positive, negative, and indeterminate results, for SARS-CoV-2 tests for New York State residents to the New York State Electronic Clinical Laboratory Reporting System (ECLRS) [18,19]. For NYC residents, ECLRS transmits reports to NYC DOHMH. These laboratory reports include specimen collection date and patient demographic information, including residential address.

For nowcasting persons newly tested, NYC DOHMH deduplicated laboratory reports, retaining the first report received (ie, report date) in ECLRS per person of a SARS-CoV-2 polymerase chain reaction (PCR) test. We retained the first specimen collection date for that associated test report date and the patient's ZIP Code of residence at time of report.

ZIP Codes are collections of points constituting a mail delivery route. The United States Census Bureau developed ZIP Code Tabulation Areas (ZCTAs), which are aggregates of census blocks, to provide an areal representation of ZIP Codes. NYC DOHMH created a custom geography referred to as a modified ZCTA (modZCTA) by merging ZCTAs with populations of less than 3000 to an adjacent ZCTA with a larger population and merging interior ZCTAs with smaller populations to the surrounding ZCTA [20,21]. There are 177 modZCTAs within NYC.

Confirmed Cases

At NYC DOHMH, electronic laboratory reports are automatically standardized, and positive results indicating a confirmed case (ie, detection of SARS-CoV-2 RNA in a clinical specimen using a molecular amplification detection test) [22] are transmitted to the NYC DOHMH's communicable disease surveillance database known as Maven (Conduent Public Health Solutions). For confirmed cases, the *diagnosis date* was defined as the specimen collection date of the first positive test. The *report date* was defined as the date the case was created in the disease surveillance database, which typically corresponded to the date the first positive test was reported to ECLRS.

Hospitalization status was ascertained by routinely matching patient identifiers for confirmed COVID-19 cases with hospitalized patients in supplemental data systems, including regional health information organizations, the New York State Hospital Emergency Response Data System, and NYC public hospitals [17]. For each hospitalized patient with a confirmed COVID-19 diagnosis, the hospital name for the most recent hospitalization in NYC was standardized to the name of a fully operational medical center. Patients with hospital discharge dates greater than 14 days prior to the collection date of their first positive PCR result were not considered hospitalized for

COVID-19. The date of hospitalization ascertainment was not retained.

Real-Time Nowcasting

NYC DOHMH nowcasted three outcomes (ie, confirmed cases, ever-hospitalized cases, and persons tested) among NYC residents at weekly increments; outcomes were nowcasted in real time through May 2020 on Mondays using reports received through the prior day on Sunday. Starting on March 24, 2020, nowcasts were conducted for all confirmed COVID-19 cases and restricted to the subset of confirmed COVID-19 cases among patients ever hospitalized. Starting on May 2, 2020, as testing became more widely available [23], nowcasts were conducted for persons newly tested by PCR for SARS-CoV-2. Each outcome was nowcasted citywide and also stratified by modZCTA of patient residence, to support targeting of community-based resources. Hospitalized cases were also nowcasted stratifying by health care facility, to support allocating resources to hospitals.

To account for reporting delays and the shape of the outcome-specific epidemic curve, we applied the R package Nowcasting by Bayesian Smoothing (NobBS), version 0.1.0 [10,24] (The R Foundation), to data for specimens collected or diagnoses during the 3 weeks prior to the nowcast through the date prior to the nowcast. Briefly, this approach corrects for underestimation of cases in real time caused by delays in reporting, learning the historical distribution of delays and relationship between cases in sequential time points to estimate the number of cases not yet reported. In performing stratified nowcasts, NobBS estimated the delay distribution citywide and the epidemic curve uniquely by stratum. Reports visualizing nowcast results were distributed weekly to DOHMH leadership for situational awareness.

We assumed an underlying Poisson distribution for case occurrence because this was the default setting in NobBS. The 3-week moving window was selected under the assumption that this length would adequately balance recency with stability. Although the optimal moving-window length was unknown in real time, given competing priorities during a pandemic, busy DOHMH officials would not have had adequate time to consider multiple nowcast versions with different window lengths as sensitivity analyses. The potential of the choice of moving-window length to considerably change nowcast estimates motivated a retrospective performance evaluation.

Retrospective Nowcasting Evaluation

For the outcome of confirmed COVID-19 cases, we characterized the delay distribution between diagnosis and report, overall during the study period and by month of report, by median number of days, IQR, and 90th percentile. We assessed the sensitivity of nowcasting results for patients diagnosed citywide during the period from March 22 to May 31, 2020—excluding cases diagnosed from March 1 to 21, given limited testing—to several choices: (1) day of week when the nowcast was performed, given outpatients with milder illness sought care and were diagnosed less frequently on weekends, when health care provider offices were typically closed or had more limited hours; (2) window length, given time-varying SARS-CoV-2 testing availability and uptake in NYC; and (3) assumed underlying distribution (ie, Poisson or negative binomial) for case occurrence. We generated Poisson regression models for the daily count by diagnosis date, separately for the entire study period and for every overlapping and nonoverlapping 2- and 3-week period, with and without weekends, used in the nowcasting evaluation. We checked the dispersion ratio for these Poisson regression models; dispersion ratios that were greater than 1 and statistically significant would indicate overdispersion and support instead using a negative binomial distribution. In addition, for nowcasting the number of cases stratified by modZCTA, we compared results using (1) the *strata* option in NobBS, which estimated the delay distribution citywide and epidemic curve separately for each modZCTA, versus estimating both the delay distribution and epidemic curve separately for each modZCTA and (2) 10,000 versus 3000 adaptations when optimizing the nowcasting algorithm [10].

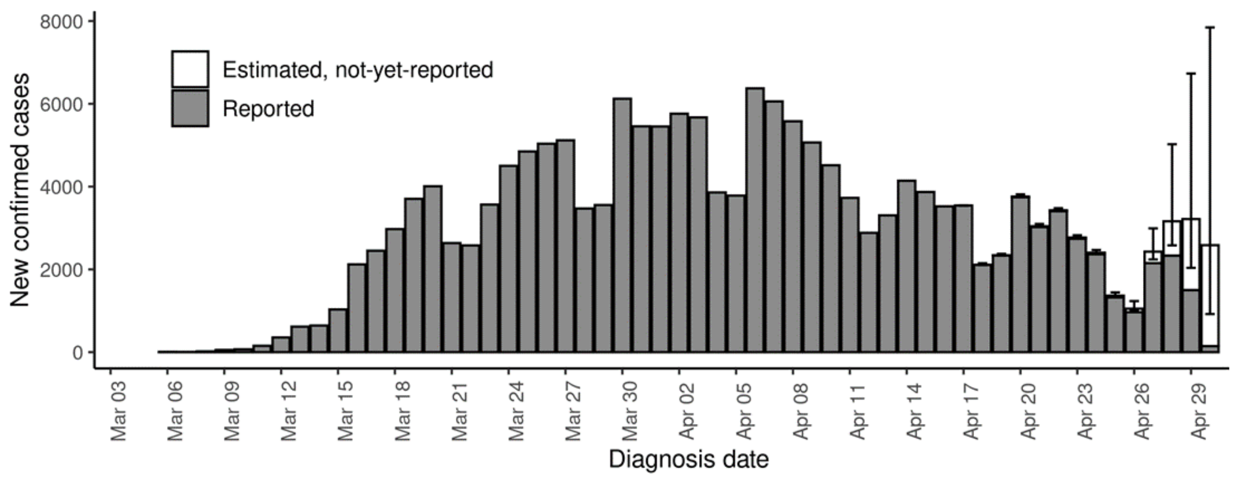
Data for the evaluation were frozen as of June 30, 2020, capturing reports received through 1 month after the end of the assessment period. We mimicked prospective surveillance at weekly intervals and daily temporal resolution, retaining the number of estimated cases for each of the prior 7 days (ie, 1-7-day hindcasts). We used the mean absolute error and the average daily relative root mean square error across all days evaluated to compare the point estimate of the number of daily hindcasted cases over the time series with the true number of cases reported. For each of these metrics, lower numbers indicate better performance of the hindcast. We also assessed the 95% prediction interval coverage (ie, the proportion of days during the study period when the 95% prediction interval included the true number of cases) [10], which should ideally be 95%.

This work was reviewed and deemed as public health surveillance that is nonresearch by the DOHMH Institutional Review Board. Line-level data, as required for nowcasting using NobBS, are not publicly available in accordance with patient confidentiality and privacy laws.

Results

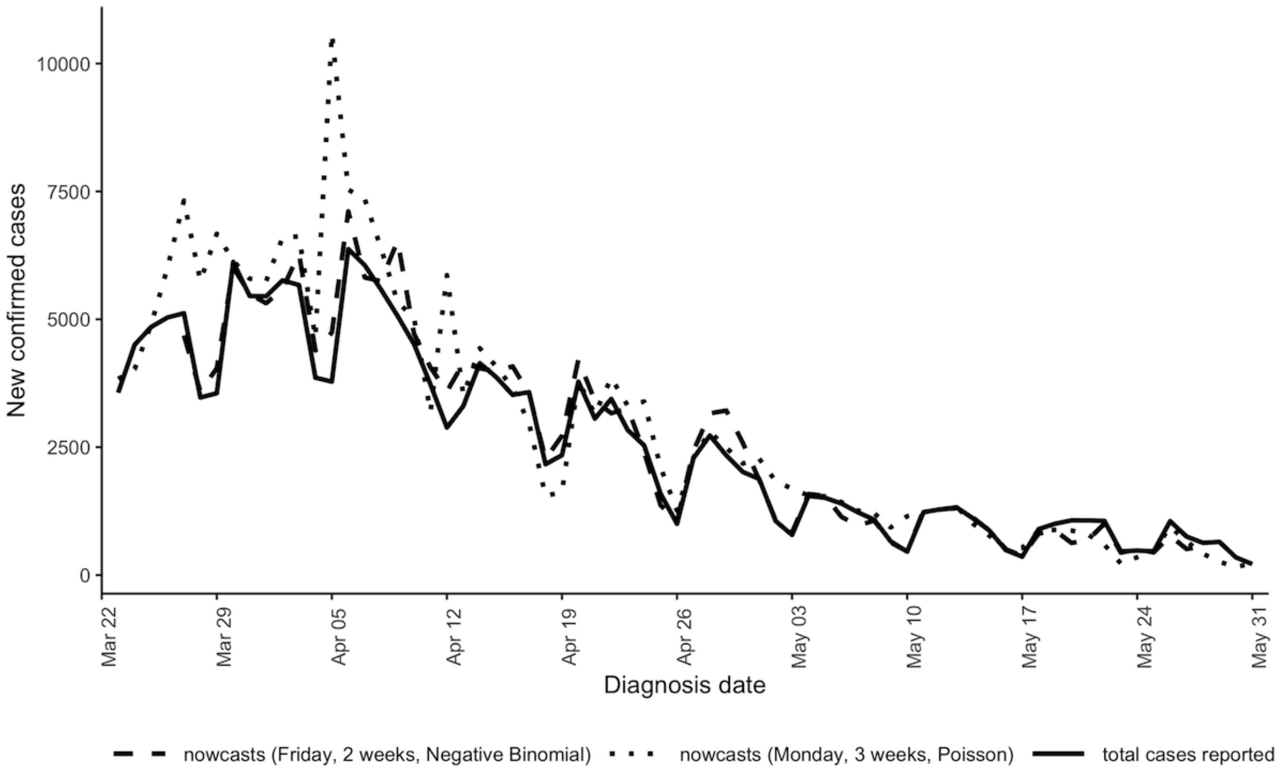
Among confirmed COVID-19 cases residing in NYC and diagnosed during the period from March to May 2020, the median delay between specimen collection and report was 2 days (IQR 1-4; 90th percentile 7). By month of report for diagnoses during the period of March to May 2020, the median number of days for this delay for reports received in March 2020 was 2 (IQR 1-4; 90th percentile 7), in April was also 2 (IQR 1-4; 90th percentile 7), in May was 2 (IQR 1-3; 90th percentile 5), and in June, given the study period included cases diagnosed through May, extended to 7 (IQR 4-19; 90th percentile 62). Hindcasts were performed weekly on Mondays in real time, with results visualized for DOHMH leadership (eg, see Figure 1).

Figure 1. Example hindcast visualization of epidemic curve of reported and estimated but not-yet-reported number of confirmed cases among New York City residents diagnosed with COVID-19, from March 1 to April 30, 2020. Illustrative hindcast performed using cases reported through April 30, 2020 (ie, a Thursday), a 2-week moving window, and a negative binomial distribution.



However, the retrospective performance evaluation determined that real-time hindcasts on Mondays using a 3-week window and an assumed Poisson distribution more often overestimated than underestimated the number of not-yet-reported cases and resulted in overly narrow 95% prediction intervals (see Figure 2 and Figure S1 in Multimedia Appendix 1). Subsequent results focus on two scenarios: the scenario that was used in real time (ie, a 3-week moving window and Poisson distribution) and the scenario that would have performed best had it been used in real time (ie, a 2-week moving window and negative binomial distribution).

Figure 2. Comparison of 7-day hindcasts conducted on Fridays with a 2-week window and negative binomial distribution, and 7-day hindcasts conducted on Mondays with a 3-week window and Poisson distribution. Total cases reported as of June 30, 2020, are shown with a black line.



We found that citywide hindcasts with a 2-week moving window and a negative binomial distribution had a 44% lower mean absolute error, a 31% lower relative root mean square error, and 0.65 higher 95% prediction interval coverage than hindcasts conducted with a 3-week moving window or with a Poisson distribution (see Table 1 as well as Table S1 and Figures S1 and S2 in Multimedia Appendix 1). Poisson regression models for daily count data for the entire study period and for each 2- and 3-week period evaluated were overdispersed (median dispersion ratio 97.5, all P<.05), which explains the better performance of the negative binomial distribution. While dispersion ratios were lower for analyses restricted to weekdays (median ratio of 32.5 vs 150 for all days), all were greater than 1, indicating overdispersion.

Table 1. Performance measures for hindcasting approaches applied to citywide case counts of New York City residents diagnosed with COVID-19, from March 22 to May 31, 2020.

Approach and sensitivity analyses	All days			Weekdays only		
	Mean absolute error	Relative root mean square error	95% prediction interval coverage	Mean absolute error	Relative root mean square error	95% prediction interval coverage
Base scenario used in near-real time by NYC DOHMH[a], using 3-week window with Poisson distribution						
All days	544	0.20	0.16	559	0.19	0.16
Hindcasting each Monday for the previous Monday-Sunday	556	0.25	0.14	338	0.12	0.20
Day-of-week hindcasting was performed for previous 7-day period, using 2-week window with negative binomial distribution						
All days	306	0.14	0.81	258	0.10	0.84
Monday	336	0.20	0.86	183	0.07	0.82
Tuesday	335	0.16	0.83	233	0.08	0.84
Wednesday	307	0.14	0.81	275	0.11	0.87
Thursday	271	0.11	0.81	257	0.11	0.84
Friday	255	0.10	0.75	267	0.11	0.84
Saturday	260	0.11	0.73	267	0.11	0.80
Sunday	372	0.16	0.87	273	0.10	0.88

[a]NYC DOHMH: New York City Department of Health and Mental Hygiene.

Hindcasts conducted toward the end of the week (ie, Thursday to Saturday) performed better than hindcasts performed earlier in the week, presumably as they had the furthest distance from the weekends. Weekends had lower overall case counts than weekdays (see Figure 1). Until mid-May, hindcasts more often overestimated than underestimated true case counts, whereas at the end of May hindcasts more often underestimated case counts, reflecting changes in the delay distribution over time (see Figure 2 and Figure S3 in Multimedia Appendix 1).

To minimize day-of-week effects that were most prominent on weekends, we also restricted performance analysis to hindcasts of cases on weekdays only, which resulted in better metrics, as expected (see Table 1 and Table S1 in Multimedia Appendix 1). The hindcasts restricted to estimating case counts for weekdays with a 2-week moving window and negative binomial distribution also performed better than the hindcasts with a 3-week moving window and Poisson distribution, with 54% lower mean absolute error, 46% lower relative root mean square error, and 0.69 higher 95% prediction interval coverage (see Table 1 and Table S1 in Multimedia Appendix 1). Performance metrics were similar across days the hindcasts were conducted, with Mondays having the lowest mean average error and relative root mean square error, as expected given the 2 additional days between the last day reported (ie, Friday) and the day the hindcast was conducted (ie, Monday). On weekdays during the study period, the average daily case count after data lags resolved was 2914, the average hindcasted case count with a 2-week window and negative binomial distribution conducted on Mondays was 2878, and the mean absolute error was 183. A combination of the window length and underlying distribution influenced the performance of the mean absolute error and relative root mean square error metrics, with larger differences occurring between different windows with the same distribution than between different distributions with the same window. On the other hand, the distribution was the primary driver for differences in the 95% prediction interval coverage (ie, differences were larger between analyses with different distributions than between analyses with the same distribution and different windows).

For hindcasts at the modZCTA level, a 2-week moving window and negative binomial distribution performed best across all metrics evaluated (see Table 2 and Table S1 in Multimedia Appendix 1), although the prediction interval coverage for the nowcasts with a Poisson distribution was higher than for citywide hindcasts. The hindcasts that assumed a citywide delay distribution performed slightly better than hindcasts that assumed different distributions by modZCTA. Metrics for 3000 versus 10,000 adaptations were essentially the same.

Table 2. Performance measures for hindcasting approaches in Nowcasting by Bayesian Smoothing (NobBS), applied to case counts of New York City residents diagnosed with COVID-19 from March 22 to May 31, 2020, stratified by modified ZIP Code Tabulation Area (modZCTA) of residence.

Approach and sensitivity analyses	All days			Weekdays only		
	Mean absolute error	Relative root mean square error	95% prediction interval coverage	Mean absolute error	Relative root mean square error	95% prediction interval coverage
Base scenario used in near-real time by NYC DOHMH[a,b]						
3-week Poisson (10,000 adaptations)	3.82	0.37	0.84	2.75	0.18	0.84
3-week Poisson (3000 adaptations)	3.83	0.37	0.84	2.76	0.18	0.84
2-week negative binomial (10,000 adaptations)	2.92	0.33	0.93	2.09	0.15	0.93
2-week negative binomial (3000 adaptations)	2.93	0.34	0.93	2.08	0.15	0.93
Conducting hindcasts on Fridays[c]						
2-week negative binomial	2.62	0.22	0.94	2.98	0.25	0.95
Estimate delay distribution separately by modZCTA[d]						
2-week negative binomial	3.55	0.36	0.94	2.57	0.21	0.95

[a]NYC DOHMH: New York City Department of Health and Mental Hygiene.

[b]The approach used the *strata* option in NobBS, which estimated the delay distribution citywide and epidemic curve separately for each modZCTA, conducted on Mondays

[c]The approach used the *strata* option in NobBS, which estimated the delay distribution citywide and epidemic curve separately for each modZCTA, conducted on Fridays.

[d]The approach involved estimating both the delay distribution and epidemic curve separately for each modZCTA conducted on Mondays.

Discussion

Principal Findings

NYC DOHMH improved situational awareness of COVID-19 testing and cases during the first epidemic wave in near-real time by applying NobBS, a readily accessible nowcasting and hindcasting method. As a result of the retrospective performance evaluation, to improve nowcast accuracy prospectively effective August 2020, we implemented the following changes to the nowcasting approach: (1) we used a negative binomial case distribution instead of a Poisson; (2) we linked the determination of the moving-window length (ie, 2 or 3 weeks) to the 90th percentile of the lag between specimen collection and report for reports received in the most recent week, choosing 3 weeks if the 90th percentile of the lag distribution is more than 14 days; and (3) we suppressed nowcasting results for specimens collected on weekends, given lack of adjustment for day-of-week effects. The evaluation supported the results of nowcasting conducted on any weekday.

Despite a mature electronic laboratory reporting system and strong informatics infrastructure and data cleaning procedures at NYC DOHMH, input data available for nowcasting had several limitations. First, for records with long lags between specimen collection and report, as long as the specimen was reported to have been collected during the pandemic period, it was not possible to distinguish long lags attributable to true delays in testing or reporting—and, thus, informative to the delay distribution—from long lags attributable to laboratory data entry errors in specimen collection dates. Second, nowcasting by patient modZCTA of residence relied on accurate laboratory reporting of patient address. For example, 1 week of real-time nowcasting results were biased when, for a batch of reports, one commercial laboratory misreported its own address as the residential address of all patients tested. Third, patient hospitalization status was largely ascertained by matching administrative records. To allow time for record matching, hospitalization nowcasts were conducted at a 3-day lag, limiting the real-time availability of results. Furthermore, records from certain facilities were unavailable in near-real time, so nowcasts of hospitalizations by patient residence and by facility were subject to spatial bias, although still considered by DOHMH leadership to be useful for situational awareness.

This version of NobBS (ie, version 0.1.0) also had several limitations when applied for nowcasting COVID-19 in NYC. First, there was no built-in functionality in NobBS to account for observable factors influencing data lags, including day-of-week and holiday effects in outpatient testing, and time-varying testing backlogs at specific laboratories differentially processing specimens for residents across neighborhoods. A recent COVID-19 nowcasting study in Bavaria, which adapted certain modeling elements from NobBS, found that modeling a weekday effect improved nowcast performance [16]. Given the substantial differences in diagnoses on weekdays compared with weekends, similar adjustments would likely benefit NYC nowcasts but were unavailable in NobBS. Similarly, there was no functionality to account for temporal trends in testing (eg, the time-varying ratio of number of tests performed to number of cases detected). Third, while 95% prediction intervals reflected uncertainty in the nowcasts themselves—encompassing uncertainty in the estimation of the delay distribution as well as in the time evolution of the epidemic curve—they did not reflect uncertainty introduced by the user-specified window length. Fourth, in generating geographically stratified nowcasts, the *strata* option in NobBS estimated the delay distribution citywide and epidemic curve separately for each modZCTA or health care facility stratum. For a highly transmissible infectious disease, nowcasting performance might be improved by considering spatial relationships across geographic strata, including spatial autocorrelation. Finally, although government officials have demonstrated interest in publicizing test percent positivity by report date [25,26], which can be biased by data lags, NobBS did not have functionality to nowcast percentages as an outcome. NobBS could be used to separately nowcast persons testing positive and negative and then to calculate test percent positivity, but there is no functionality to appropriately account for the separate uncertainties in the numerator and denominator of this percentage.

Practice Implications

When tracking ongoing outbreaks using epidemic curves, public health officials recognize that data for recent days are incomplete because of reporting delays. Data lags can make it difficult for policy makers to discern in near-real time whether apparent decreases in recent case counts are the result of public health interventions, such as social distancing guidelines. NYC DOHMH filled in COVID-19 epidemic curves using NobBS, which helped ensure that recent decreases in observed case counts were not overinterpreted as true declines in disease and supported the continuation of policies to reduce transmission. Nowcasted citywide case counts supported situational awareness and assisted DOHMH leadership in anticipating the magnitude and timing of hospitalizations and deaths. Nowcasting hospitalizations by health care facility was useful in helping to route patient transports and avoid overburdening facilities.

As the COVID-19 pandemic continues, state and local health departments should incorporate nowcasting into their workflows. This performance evaluation led to analytic improvements in place for the second wave of COVID-19 in NYC, including the use of a more suitable underlying distribution for case occurrence, a dynamic window length to account for periods with an extended lag distribution, and suppression of diagnoses on weekends to avoid biased trend estimates. Nowcasted case counts can also be used as inputs for near-real time estimates of other outbreak monitoring metrics, including the time-varying reproduction number [27] and doubling times [28]. Further evaluations are warranted to assess nowcasting performance during different COVID-19 epidemic phases and across jurisdictions experiencing a variety of data lag distributions, including more extensive reporting delays [29], and for additional outcomes, such as deaths.

Acknowledgments

The authors thank the NYC DOHMH Incident Command System Surveillance and Epidemiology Section, including Jennifer Baumgartner, Eric R Peterson, and Miranda S Moore for data preparation; Samia Baig for visualization; and Dr Annie D Fine for proposing nowcasting by health care facility. The authors also thank Angel Aponte for administering the NYC DOHMH R server. SG was supported by the Public Health Emergency Preparedness Cooperative Agreement (grant No. NU90TP922035-01), funded by the US Centers for Disease Control and Prevention. RK was supported by the US National Institute of General Medical Sciences (award No. U54GM088558). ML was supported by the Morris-Singer Fund and by a subcontract from Carnegie Mellon University under an award from the US Centers for Disease Control and Prevention (award No. U01IP001121). This article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention, the National Institutes of Health, or the Department of Health and Human Services.

Authors' Contributions

SG oversaw design and implementation of nowcasting for COVID-19 at NYC DOHMH and conceived of the evaluation. SM, ML, and NM provided critical input on design and interpretation of nowcasting analyses and evaluation. GC contributed to data interpretation and led geographic visualization of nowcasting results. LG contributed to hospitalization data standardization and analysis. RK led the nowcasting evaluation. SG and RK drafted the article. SM, GC, LG, ML, and NM reviewed and revised the article critically for important intellectual content. All authors gave final approval of the submitted version.

Conflicts of Interest

ML discloses honoraria and consulting work from Merck, Affinivax, Sanofi-Pasteur, Bristol Myers-Squibb, and Antigen Discovery; institutional research funding from Pfizer; and unpaid scientific advice to Janssen, Astra-Zeneca, One Day Sooner, and Covaxx (United Biomedical). All other authors declare no conflicts.

Multimedia Appendix 1
Supplemental table and figures.
[DOCX File , 3461 KB - publichealth_v7i1e25538_app1.docx]

References

1. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. BMC Public Health 2004 Jul 26;4:29 [FREE Full text] [doi: 10.1186/1471-2458-4-29] [Medline: 15274746]
2. Groseclose SL, Buckeridge DL. Public health surveillance systems: Recent advances in their use and evaluation. Annu Rev Public Health 2017 Mar 20;38:57-79. [doi: 10.1146/annurev-publhealth-031816-044348] [Medline: 27992726]
3. Prevent Epidemics. Tracking COVID-19 in the United States: From Information Catastrophe to Empowered Communities. New York, NY: Vital Strategies; 2020 Jul 21. URL: https://preventepidemics.org/wp-content/uploads/2020/07/RTSL_Tracking-COVID-19-in-the-United-States_-7-23-2020.pdf [accessed 2021-01-08]
4. COVID-19: Data. Public health milestones. Long Island City, NY: New York City Department of Health and Mental Hygiene; 2021. URL: https://www1.nyc.gov/site/doh/covid/covid-19-goals.page
5. Bonačić Marinović A, Swaan C, van Steenbergen J, Kretzschmar M. Quantifying reporting timeliness to improve outbreak control. Emerg Infect Dis 2015 Feb;21(2):209-216 [FREE Full text] [doi: 10.3201/eid2102.130504] [Medline: 25625374]
6. Elliot AJ, Harcourt SE, Hughes HE, Loveridge P, Morbey RA, Smith S, et al. The COVID-19 pandemic: A new challenge for syndromic surveillance. Epidemiol Infect 2020 Jun 18;148:e122 [FREE Full text] [doi: 10.1017/S0950268820001314] [Medline: 32614283]
7. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):1-5 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
8. Chan AT, Brownstein JS. Putting the public back in public health - Surveying symptoms of Covid-19. N Engl J Med 2020 Aug 13;383(7):e45. [doi: 10.1056/NEJMp2016259] [Medline: 32501663]
9. Kogan N, Clemente L, Liautaud P, Kaashoek J, Link N, Nguyen A, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real-time. ArXiv Preprint posted online on July 3, 2020. [FREE Full text]
10. McGough SF, Johansson MA, Lipsitch M, Menzies NA. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. PLoS Comput Biol 2020 Apr;16(4):e1007735. [doi: 10.1371/journal.pcbi.1007735] [Medline: 32251464]
11. Mavragani A. Tracking COVID-19 in Europe: Infodemiology approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
12. Bird S, Nielsen B. Now-casting of COVID-19 deaths in English hospitals. University of Oxford. 2020 Jul 07. URL: http://users.ox.ac.uk/~nuff0078/Covid/ [accessed 2021-01-08]
13. Schneble M, De Nicola G, Kauermann G, Berger U. Nowcasting fatal COVID-19 infections on a regional level in Germany. Biom J 2020 Nov 20:1-19 [FREE Full text] [doi: 10.1002/bimj.202000143] [Medline: 33215765]
14. Masjedi H, Rabajante JF, Bahranizadd F, Zare MH. Nowcasting and forecasting the spread of COVID-19 in Iran. medRxiv Preprint posted online on April 27, 2020. [FREE Full text] [doi: 10.1101/2020.04.22.20076281]
15. Annan JD, Hargreaves JC. Model calibration, nowcasting, and operational prediction of the COVID-19 pandemic. medRxiv Preprint posted online on May 27, 2020. [FREE Full text] [doi: 10.1101/2020.04.14.20065227]
16. Günther F, Bender A, Katz K, Küchenhoff H, Höhle M. Nowcasting the COVID-19 pandemic in Bavaria. Biom J 2020 Dec 01:1-13 [FREE Full text] [doi: 10.1002/bimj.202000112] [Medline: 33258177]
17. Thompson CN, Baumgartner J, Pichardo C, Toro B, Li L, Arciuolo R, et al. COVID-19 outbreak - New York City, February 29-June 1, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1725-1729 [FREE Full text] [doi: 10.15585/mmwr.mm6946a2] [Medline: 33211680]
18. Nguyen TQ, Thorpe L, Makki HA, Mostashari F. Benefits and barriers to electronic laboratory results reporting for notifiable diseases: The New York City Department of Health and Mental Hygiene experience. Am J Public Health 2007 Apr;97 Suppl 1:S142-S145. [doi: 10.2105/AJPH.2006.098996] [Medline: 17413058]
19. Health Advisory: Reporting Requirements for ALL Laboratory Results for SARS-CoV-2, Including all Molecular, Antigen, and Serological Tests (including “Rapid” Tests) and Ensuring Complete Reporting of Patient Demographics. Albany, NY: New York State Department of Health; 2020 Apr 30. URL: https://coronavirus.health.ny.gov/system/files/documents/2020/04/doh_covid19_reportingtestresults_rev_043020.pdf [accessed 2021-01-08]
20. ZIP Code Tabulation Areas (ZCTAs). United States Census Bureau. 2020. URL: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html [accessed 2021-01-08]
21. Modified Zip Code Tabulation Areas (MODZCTA). NYC OpenData. 2020. URL: https://data.cityofnewyork.us/Health/Modified-Zip-Code-Tabulation-Areas-MODZCTA-/pri4-ifjk [accessed 2021-01-08]
22. Turner K, Davidson S, Collins J, Park S, Pedati C. Standardized Surveillance Case Definition and National Notification for 2019 Novel Coronavirus Disease (COVID-19). Atlanta, GA: Council of State and Territorial Epidemiologists (CSTE); 2020. URL: https://cdn.ymaws.com/www.cste.org/resource/resmgr/2020ps/Interim-20-ID-01_COVID-19.pdf [accessed 2021-01-08]
23. 2020 Health Advisory #15: Updated NYC Health Department Recommendations for Identifying and Testing Patients with Suspected COVID-19. Long Island City, NY: New York City Department of Health and Mental Hygiene; 2020 May 15. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/han/advisory/2020/covid-19-provider-id-testing.pdf [accessed 2021-01-08]

24. McGough S, Menzies N, Lipsitch M, Johansson M. NobBS: Nowcasting by Bayesian Smoothing, version 0.1.0. The Comprehensive R Archive Network. 2020 Mar 03. URL: https://CRAN.R-project.org/package=NobBS [accessed 2021-01-08]
25. Governor Cuomo announces new record-high number of COVID-19 tests reported to New York State. Office of the Governor of New York State. 2020 Sep 19. URL: https://www.governor.ny.gov/news/governor-cuomo-announces-new-record-high-number-covid-19-tests-reported-new-york-state-1 [accessed 2021-01-08]
26. Walters E. Gov Greg Abbott says Texas is investigating its high proportion of coronavirus tests coming back positive. The Texas Tribune. 2020 Aug 13. URL: https://www.texastribune.org/2020/08/13/texas-positivity-rate-coronavirus/ [accessed 2021-01-08]
27. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013 Nov 01;178(9):1505-1512 [FREE Full text] [doi: 10.1093/aje/kwt133] [Medline: 24043437]
28. Jombart T, Kamvar ZN. Overview of the incidence package. The Comprehensive R Archive Network. 2020 Nov 03. URL: https://cran.r-project.org/web/packages/incidence/vignettes/overview.html [accessed 2021-01-08]
29. Goldstein J, McKinley J. Testing bottlenecks threaten NYC's ability to contain virus. The New York Times. 2020 Jul 23. URL: https://www.nytimes.com/2020/07/23/nyregion/coronavirus-testing-nyc.html [accessed 2021-01-08]

Abbreviations

DOHMH: Department of Health and Mental Hygiene
ECLRS: Electronic Clinical Laboratory Reporting System
modZCTA: modified ZIP Code Tabulation Area
NobBS: Nowcasting by Bayesian Smoothing
NYC: New York City
PCR: polymerase chain reaction
ZCTA: ZIP Code Tabulation Area

Edited by T Sanchez; submitted 05.11.20; peer-reviewed by E Hall, A Rovetta; comments to author 14.12.20; revised version received 31.12.20; accepted 04.01.21; published 15.01.21.

Please cite as:
Greene SK, McGough SF, Culp GM, Graf LE, Lipsitch M, Menzies NA, Kahn R
Nowcasting for Real-Time COVID-19 Tracking in New York City: An Evaluation Using Reportable Disease Data From Early in the Pandemic
JMIR Public Health Surveill 2021;7(1):e25538
URL: http://publichealth.jmir.org/2021/1/e25538/
doi:10.2196/25538
PMID:33406053

Original Paper

Telemedicine and the Use of Korean Medicine for Patients With COVID-19 in South Korea: Observational Study

Soobin Jang[1*], PhD; Dongsu Kim[2,3*], PhD, MPH; Eunhee Yi[2], MPA; Gunhee Choi[4]; Mideok Song[5], PhD; Eun-Kyoung Lee[6], PhD

[1]Clinical Medicine Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea
[2]KM Policy Research Center, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea
[3]College of Oriental Medicine, Dongshin University, Jeollanam-do, Republic of Korea
[4]Medical/Information and Communications Affairs, The Association of Korean Medicine, Seoul, Republic of Korea
[5]Department of Academic Affairs, The Association of Korean Medicine, Seoul, Republic of Korea
[6]Research Institute of Korean Medicine Policy, The Association of Korean Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Eun-Kyoung Lee, PhD
Research Institute of Korean Medicine Policy
The Association of Korean Medicine
91 Heojun-ro, Gangseo-go
Seoul, 07525
Republic of Korea
Phone: 82 2 5657 5000
Email: eundust@hotmail.com

Abstract

Background: COVID-19 was first reported in Wuhan, China, in December 2019, and it has since spread worldwide. The Association of Korean Medicine (AKOM) established the COVID-19 telemedicine center of Korean medicine (KM telemedicine center) in Daegu and Seoul.

Objective: The aim of this study was to describe the results of the KM telemedicine center and the clinical possibility of using herbal medicines for COVID-19.

Methods: All procedures were conducted by voice call following standardized guidelines. The students in the reception group obtained informed consent from participants and they collected basic information. Subsequently, Korean Medicine doctors assessed COVID-19–related symptoms and prescribed the appropriate herbal medicine according to the KM telemedicine guidelines. The data of patients who completed the program by June 30, 2020, were analyzed.

Results: From March 9 to June 30, 2020, 2324 patients participated in and completed the KM telemedicine program. Kyung-Ok-Ko (n=2285) was the most prescribed herbal medicine, and Qingfei Paidu decoction (I and II, n=2053) was the second most prescribed. All COVID-19–related symptoms (headache, chills, sputum, dry cough, sore throat, fatigue, muscle pain, rhinorrhea, nasal congestion, dyspnea, chest tightness, diarrhea, and loss of appetite) improved after treatment (P<.001).

Conclusions: The KM telemedicine center has provided medical service to 10.8% of all patients with COVID-19 in South Korea (as of June 30, 2020), and it is still in operation. We hope that this study will help to establish a better health care system to overcome COVID-19.

(JMIR Public Health Surveill 2021;7(1):e20236) doi:10.2196/20236

KEYWORDS

telemedicine; telehealth; herbal medicine; Korean medicine; COVID-19; Korea; pandemic; guideline; infectious disease

Introduction

Declared a pandemic in early 2020, COVID-19 has spread beyond China, reaching all around the world. The World Health Organization (WHO) increased its assessment of the risk of spread and impact of COVID-19 to "very high" at the global level on February 28 [1]. On November 15, 2020, the number of confirmed cases of COVID-19 exceeded 53.7 million, and the death toll was approximately 1.3 million [2]. In South Korea, since the day the first patient was confirmed on January 20, 2020, the total number of confirmed cases on September 1 exceeded 20,000. There was an outbreak of COVID-19 in Daegu and Gyeongbuk (Daegu-Gyeongbuk) due to religious gatherings attended by many people; consequently, the total number of cases rapidly increased from 30 on February 17 to 8086 on March 14. This health care surge led to a shortage of hospital beds, medical institutions, and medical personnel, as well as a gap in the management of patients.

The Association of Korean Medicine (AKOM) created a COVID-19 telemedicine center of Korean Medicine (KM telemedicine center) at the Daegu Korean Medicine Hospital to provide medical services via telephone to marginalized patients, on March 9, 2020. After the outbreak in Daegu-Gyeongbuk was stabilized, AKOM also established an additional telemedicine center in Seoul (Figures 1 and 2). Telemedicine, in principal, is prohibited in South Korea; however, the government temporarily allowed telephone counseling or prescriptions due to COVID-19. At the KM telemedicine center, herbal medicines were provided to patients with suspected and confirmed COVID-19 for the management of COVID-19–related symptoms following the guideline on COVID-19 telemedicine service of Korean Medicine (KM telemedicine guideline). The KM telemedicine guideline was based on three previously published guidelines in South Korea and China, "Recommendations on COVID-19 Korean Medicine (AKOM)" [3], "Clinical Guidelines on COVID-19 Korean Medicine" (the KM Professor Council of Internal Medicine of the Respiratory System) [4], and the Chinese government guidelines titled "Notice on the Issuance of the New 7th Version of the COVID-19 Diagnosis and Treatment Guidelines" [5].

Figure 1. Locations of the COVID-19 telemedicine centers of Korean Medicine. KM: Korean Medicine.

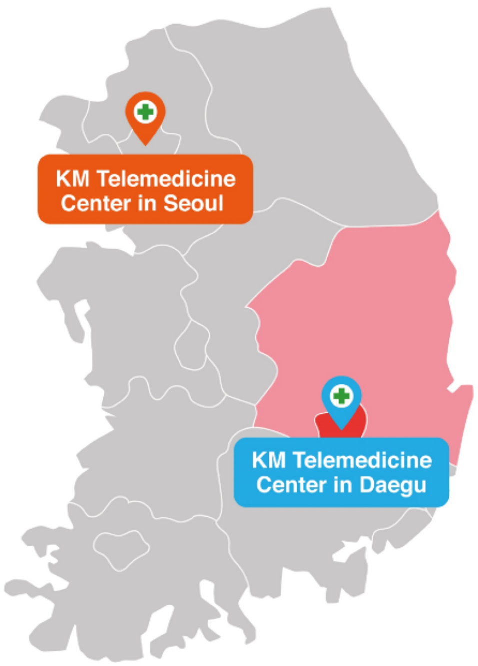
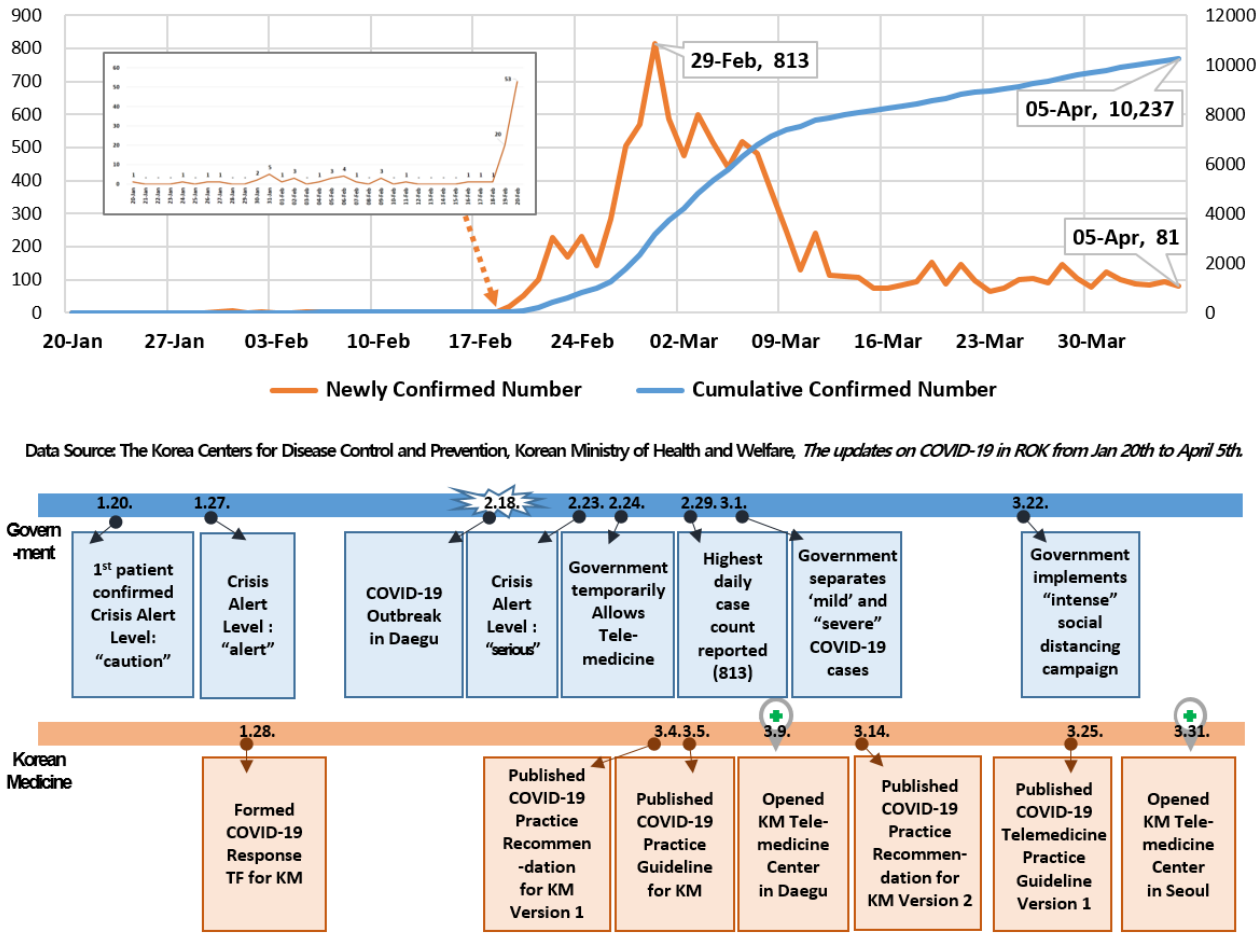


Figure 2. COVID-19 confirmed cases in South Korea over time. KM: Korean Medicine; ROK: Republic of Korea; TF: task force.



It has been reported that herbal medicines have been effective in reducing COVID-19 symptoms [6,7] in patients with mild symptoms and may be an alternative method to prevent the worsening of COVID-19 symptoms [8]. In particular, Qingfei Paidu decoction, which has been used for fever in Traditional Chinese Medicine (TCM), is strongly recommended for patients with confirmed COVID-19 [5,9]. This is the first study to report clinical cases of using herbal medicines for the treatment of COVID-19 in South Korea. This study aimed to describe patient symptoms and the use of medication provided by the KM telemedicine center and examine the clinical possibility of herbal medicines for COVID-19.

Methods

KM Telemedicine Center Operations

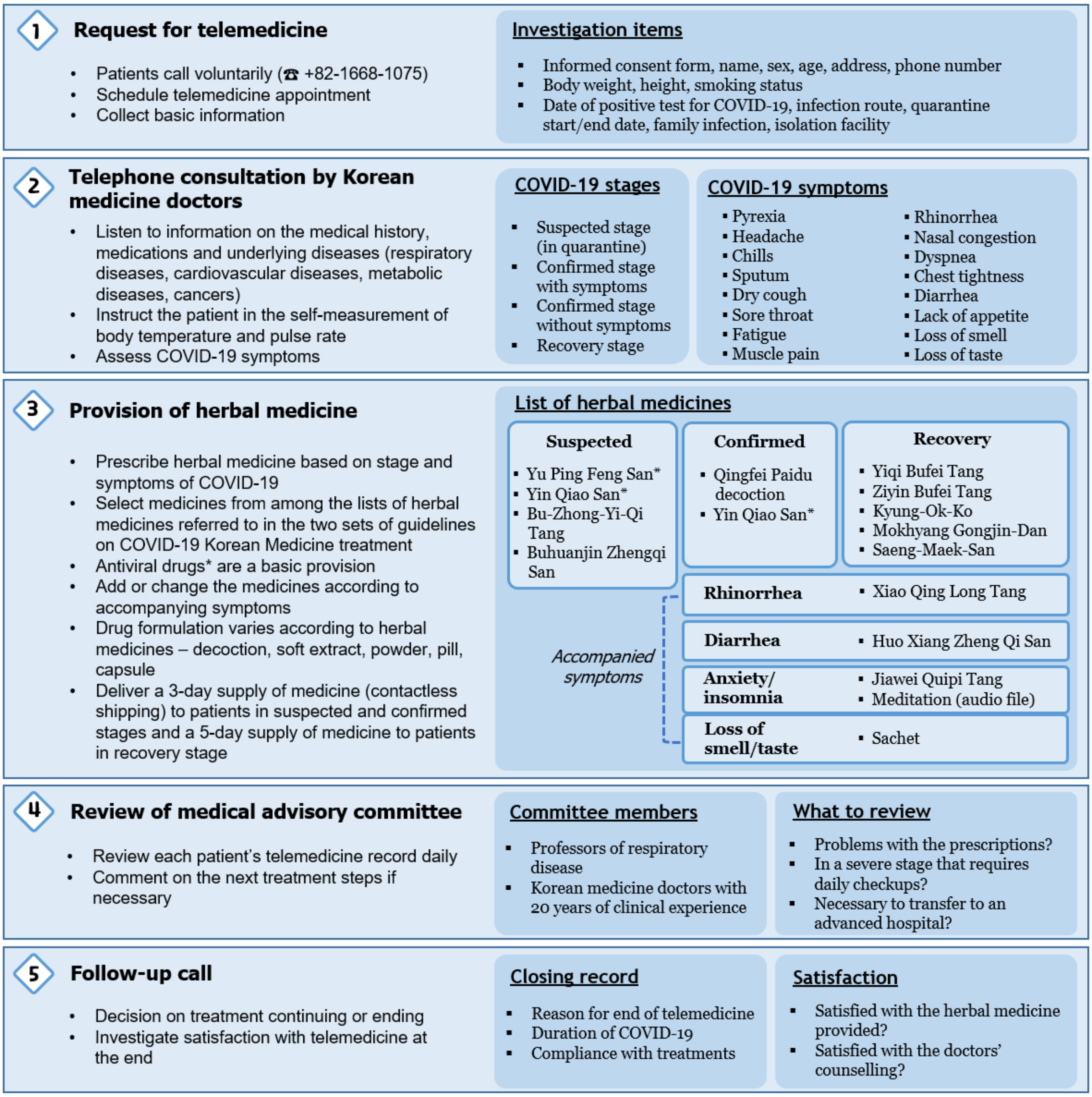
The KM telemedicine center consisted of an operation group, reception group, medical group, and advisory group. The operation group included AKOM executives and staff who managed the overall operation of the telephone medical center, including financial arrangements, the publicization of the KM telemedicine program, the recruitment of volunteers, and the delivery of herbal medicines. The reception group consisted of students from colleges of Korean Medicine; the medical group included Korean Medicine doctors (KMDs). Both groups were recruited voluntarily. KMDs who worked at the telemedicine center were educated in advance on how to make diagnoses, prescribe medicines, and write electronic charts. The advisory group consisted of professors from colleges of Korean Medicine, experts from academic societies, and experienced clinicians. In addition, several herbal pharmaceutical companies provided herbal medicines to the KM telemedicine center, and the rest of the expenditures, including delivery, meals, and electronic devices, were fundraised by KMDs.

KM Telemedicine Center Procedures

All KM telemedicine procedures are conducted by voice call following standardized guidelines. Patients who want to receive KM telemedicine service voluntarily call the representative number +82-1688-1075. The students in the reception group explain the overall process and obtain the informed consent of those participating in the KM telemedicine program. They collect basic information, including the patient's name, sex, age, address, smoking habits, date of positive COVID-19 test, infection route, and quarantine start/end date. They then schedule a telephone consultation with a KMD. KMDs then call the patients to investigate their medical history, medications, and underlying diseases, and instruct them to measure their body temperature and pulse rate. The KMDs assess COVID-19–related symptoms including pyrexia, rhinorrhea, headache, sputum, sore throat, and diarrhea, and consider pattern identification of Korean Medicine theory. They then prescribe the appropriate herbal medicine according to those presented in the KM telemedicine guidelines.

According to the guidelines, patients with COVID-19 were first classified as uncertain suspected patients, confirmed patients (mild), asymptomatic patients, and recovery patients. When the KM telemedicine center first began operating, confirmed patients who were not hospitalized because of mild symptoms and suspected patients who had been in close contact with patients with confirmed COVID-19 were targeted by the program. As the number of cured people has gradually increased, recovering patients who are within 14 days of having tested negative for COVID-19 have been included. After the first telephone consultation, follow-up telemedicine consultations take place every 4-5 days until patients are fully recovered. The suspected patients and confirmed patients are provided with a 3-day supply of herbal medicine by contactless shipping, and recovering patients receive a 5-day supply of herbal medicine. The advisory group reviews every medical record and identifies critical patients who need intensive care, recommending transfer to advanced hospitals if necessary. In addition, the advisory group may intervene during telephone consultation in real time to discuss treatment directions with the medical team. When a patient has recovered from COVID-19 and has no other symptoms, the telemedicine ends following a decision by the KMDs (Figure 3).

Figure 3. Flow diagram of steps undertaken by the COVID-19 telemedicine center of Korean Medicine.



KM Telemedicine Center Interventions

For patients who have tested positive for COVID-19, the basic herbal medicine provided at the KM telemedicine center is Qingfei Baidu decoction. There are two kinds of Qingfei Baidu decoction: I is the original, and its components are the same as the Chinese Qingfei Baidu decoction, while II is the same as I but without Ephedra Herba. Qingfei Baidu decoction (II) is prescribed to those with high potential for palpitation, a side effect of Ephedra Herba. Yu Ping Feng San and Yin Qiao San have antiviral effects; therefore, they are provided to patients with suspected COVID-19. Yin Qiao San (instead of Qingfei Baidu decoction) can be considered for confirmed patients with sore throats and headaches.

Since many patients with COVID-19 complain of digestive symptoms such as diarrhea and lack of appetite, Huo Xiang Zheng Qi San, Bu-Zhong-Yi-Qi-Tang, and Buhuanjin Zhengqi San are included in the herbal medicine list for the treatment of COVID-19. KM doctors may consider other herbal medicines according to patients' accompanying symptoms, such as Xiao Qing Long Tang for rhinorrhea and Jiawei Quipi Tang for anxiety. In particular, even if the respiratory symptoms are mild, many cases of emotional problems have been reported due to isolation or anxiety about the disease's exacerbation. It has been reported that quarantined patients with COVID-19 suffer from insomnia, anxiety, anger, overthinking, decreased concentration, and a loss of energy [10]. Emotional symptoms such as insomnia and anxiety can continue after quarantine release; therefore, Jiawei Quipi Tang and a meditation audio file are provided in those cases. If patients have lost their sense of smell or taste, a sachet consisting of 7 herbs is delivered along with the herbal medicine. In the KM telemedicine program, managing recovering patients is important because many patients still complain about physical or emotional problems even after testing negative for COVID-19. We found that Kyung-Ok-Ko is the second most frequently prescribed medicine. Yiqi Bufei Tang is for patients with digestive symptoms, and Ziyin Bufei Tang is for patients with respiratory symptoms. Saeng-Maek San targets those who have dry mouth and sweating. Kyung-Ok-Ko and Mokhyang Gongjin-Dan are used to enhance stamina after the onset of viral infection. The formulation of the herbal medicines has varied because they were donated by herbal pharmaceutical companies and KMDs (Multimedia Appendix 1).

Statistical Analyses

A statistician performed the statistical analyses for intention to treat (ITT) using SAS software (Version 9.4, SAS Institute). Data of patients who completed the program by June 30 were analyzed. Continuous variables (patient satisfaction) were displayed as mean (SD), median, and interquartile range. Categorical variables (sex, age, region, diagnosis, number of calls, residence of patients, underlying disease, prescriptions, and COVID-19 symptom severity) were shown as frequencies. The Wilcoxon signed-rank test was used for 13 COVID-19 symptoms to examine significant differences before and after treatment.

Ethics Approval

This study could not receive original individual medical charts. However, we did receive the medical activities data reported by AKOM. An institutional review board at the Korea Institute of Oriental Medicine permitted this study's use of medical records (Number I-2006/005-001-02).

Results

Basic Information About the KM Telemedicine Center's Operation and Patient Characteristics

The KM telemedicine center for COVID-19 was set up in Daegu on March 9, 2020, to provide medical services, and another center was opened in Seoul on March 31. By June 30, the total number of patients was 2324. The number of confirmed cases (excluding the suspected cases) was 1392, which was 10.9% of the nationwide confirmed cases of COVID-19 as of June 30. The number of participating KMDs and students was 546 and 332, respectively. Detailed information about patients is described in Table 1.

Table 1. Characteristics of patients enrolled in the Korean Medicine telemedicine program for COVID-19 (N=2324).

Characteristics	Patients, n (%)
Sex	
Male	569 (24.5)
Female	1755 (75.5)
Age (years)	
0-19	97 (4.2)
20-39	555 (23.9)
40-59	1121 (48.2)
≥60	550 (23.7)
Unknown	1 (0.0)
Region	
Daegu and Gyeongbuk	2196 (94.5)
Seoul	32 (1.4)
Other	94 (4.1)
Diagnosis (based on first consultation)	
Suspected cases	45 (1.9)
Confirmed cases	211 (9.1)
Recovering cases after discharge	2009 (86.4)
Other, unknown	59 (2.6)
Number of calls (per patient)	
1	166 (7.1)
2	329 (14.2)
3	432 (18.6)
4	338 (14.5)
≥5	1059 (45.6)
Residence of patients (based on first consultation)	
Home	2165 (93.2)
Facility	69 (2.9)
Hospital	43 (1.9)
Other, unknown	47 (2.0)
Underlying diseases	
Hypertension (n=1912)	
No	1548 (81.0)
Yes, not taking medication	44 (2.3)
Yes, taking medication	320 (16.7)
Diabetes mellitus (n=1859)	
No	1690 (90.9)
Yes, not taking medication	25 (1.3)
Yes, taking medication	144 (7.7)
Hyperlipidemia (n=1844)	
No	1626 (88.2)
Yes, not taking medication	72 (3.9)
Yes, taking medication	146 (7.9)

Characteristics	Patients, n (%)
Cancer (n=1734)	
No	1661 (95.8)
Yes, not taking medication	49 (2.8)
Yes, taking medication	24 (1.4)
Chronic respiratory diseases (n=1726)	
No	1611 (93.3)
Yes, not taking medication	82 (4.8)
Yes, taking medication	33 (1.9)

Results of the KM Telemedicine Center's Operation

Kyung-Ok-Ko (n=2285) was the most prescribed herbal medicine, with Qingfei Paidu decoction (I and II, n=2053) following as the second most prescribed medicine. Ziyin Bufei Tang (n=1780) and Yiqi Bufei Tang (n=1499), which are herbal medicines for the recovery stage, were the third and fourth most prescribed medicines (Table 2).

Table 2. Frequency of prescriptions by the Korean Medicine telemedicine program for COVID-19.

Prescription	Uses	First call, n (%)	Others, n (%)	Total, n (%)
Qingfei Paidu decoction I	Exogenous fever and influenza	448 (13.4)	758 (6.9)	1206 (8.4)
Kyung-Ok-Ko	Fatigue, weak condition, and dry cough	439 (13.2)	1846 (16.7)	2285 (15.9)
Ziyin Bufei Tang	Short breath, fatigue, decreased appetite, dry mouth, and dry cough	369 (11.1)	1411 (12.8)	1780 (12.4)
Yiqi Bufei Tang	Short breath, fatigue, decreased appetite, and loose stool	348 (10.4)	1151 (10.4)	1499 (10.4)
Qingfei Paidu decoction II	Exogenous fever and influenza	263 (7.9)	584 (5.3)	847 (5.9)
Jiawei Quipi Tang	Anxiety, overthinking, and insomnia	234 (7.0)	840 (7.6)	1074 (7.5)
Mokhyang Gongjin-Dan	Fatigue and weak condition	176 (5.3)	747 (6.8)	923 (6.4)
Huo Xiang Zheng Qi San	Loose stool or diarrhea, chest tightness, and lack of vigor	139 (4.2)	420 (3.8)	559 (3.9)
Saeng-Maek-San	Thirst, dry mouth, cold sweat, and weak condition	82 (2.5)	259 (2.3)	341 (2.4)
Xiao Qing Long Tang	Nasal symptoms, watery rhinorrhea, and allergic rhinitis	94 (2.8)	347 (3.1)	441 (3.1)
Yin Qiao San	Sore throats and headaches with inner heat	89 (2.7)	348 (3.2)	437 (3.0)
Yu Ping Feng San	Cold, influenza, and sweating	40 (1.2)	71 (0.6)	111 (0.8)
Buhuanjin Zhengqi San	Cold, headache, and fatigue	26 (0.8)	124 (1.1)	150 (1.0)
Other	N/A[a]	584 (17.5)	2129 (19.3)	2713 (18.9)
Total[b]	N/A	3331 (100.0)	11,035 (100.0)	14,366 (100.0)

[a]N/A: not applicable.

[b]Due to duplicate medication orders, the total number of prescriptions exceeded the number of patients.

All variables were analyzed with data from the patients who completed the KM telemedicine center's program by June 30. The difference in clinical symptoms related to COVID-19 before and after treatment are presented in Table 3. All 13 symptoms (headache, chills, sputum, dry cough, sore throat, fatigue, muscle pain, rhinorrhea, nasal congestion, dyspnea, chest tightness, diarrhea, and loss of appetite) improved after treatment (all $P<.001$). However, due to missing data, there were large differences in the analyzed numbers between the first and last call. The patient satisfaction score for treatment was 8.3 (SD 1.78) out of 10 and the convenience of the KM telemedicine center system was rated 9.3 (SD 1.27) out of 10. Patients gave high scores when asked about their willingness to recommend the KM telemedicine center to acquaintances (9.2, SD 1.58) and use Korean Medicine treatments (9.1, SD 1.54; Table 4).

Table 3. Changes in COVID-19 symptoms before and after accessing the Korean Medicine telemedicine program.

COVID-19 symptoms and severity	At first call, n (%)	At last call, n (%)	P value[a]
Headache			<.001
Severe	3 (0.1)	0 (0.0)	N/A[b]
Moderate	49 (2.4)	3 (0.2)	N/A
Mild	361 (17.4)	112 (9.3)	N/A
None	1658 (80.1)	1092 (90.5)	N/A
Total	2071 (100.0)	1207 (100.0)	N/A
Chills			<.001
Severe	2 (0.1)	0 (0.0)	N/A
Moderate	11 (0.5)	0 (0.0)	N/A
Mild	207 (10.0)	57 (4.7)	N/A
None	1843 (89.3)	1158 (95.3)	N/A
Total	2063 (100.0)	1215 (100.0)	N/A
Sputum			<.001
Severe	12 (0.6)	2 (0.2)	N/A
Moderate	43 (2.0)	10 (0.8)	N/A
Mild	707 (33.4)	308 (24.0)	N/A
None	1357 (64.0)	963 (75.1)	N/A
Total	2119 (100.0)	1283 (100.0)	N/A
Dry cough			<.001
Severe	3 (0.1)	0 (0.0)	N/A
Moderate	116 (5.5)	16 (1.3)	N/A
Mild	640 (30.6)	233 (18.7)	N/A
None	1335 (63.8)	994 (80.0)	N/A
Total	2094 (100.0)	1243 (100.0)	N/A
Sore throat			<.001
Severe	2 (0.1)	0 (0.0)	N/A
Moderate	43 (2.1)	2 (0.2)	N/A
Mild	355 (17.0)	117 (9.5)	N/A
None	1684 (80.8)	1113 (90.3)	N/A
Total	2084 (100.0)	1232 (100.0)	N/A
Fatigue			<.001
Severe	15 (0.7)	3 (0.2)	N/A
Moderate	164 (8.1)	27 (2.2)	N/A
Mild	616 (30.5)	325 (26.7)	N/A
None	1223 (60.6)	860 (70.8)	N/A
Total	2018 (100.0)	1215 (100.0)	N/A
Muscle pain			<.001
Severe	7 (0.3)	0 (0.0)	N/A
Moderate	60 (2.9)	8 (0.7)	N/A
Mild	334 (16.2)	108 (9.1)	N/A
None	1661 (80.6)	1066 (90.2)	N/A

COVID-19 symptoms and severity	At first call, n (%)	At last call, n (%)	P value[a]
Total	2062 (100.0)	1182 (100.0)	N/A
Rhinorrhea			<.001
Severe	9 (0.4)	0 (0.0)	N/A
Moderate	43 (2.1)	4 (0.3)	N/A
Mild	316 (15.3)	164 (13.6)	N/A
None	1694 (82.2)	1042 (86.1)	N/A
Total	2062 (100.0)	1210 (100.0)	N/A
Nasal congestion			<.001
Severe	5 (0.2)	1 (0.1)	N/A
Moderate	42 (2.0)	4 (0.3)	N/A
Mild	386 (18.7)	118 (9.8)	N/A
None	1627 (79.0)	1078 (89.8)	N/A
Total	2060 (100.0)	1201 (100.0)	N/A
Dyspnea			<.001
Severe	17 (0.9)	0 (0.0)	N/A
Moderate	36 (1.9)	8 (0.7)	N/A
Mild	413 (21.7)	137 (12.8)	N/A
None	1439 (75.5)	924 (86.4)	N/A
Total	1905 (100.0)	1069 (100.0)	N/A
Chest tightness			<.001
Severe	0 (0.0)	0 (0.0)	N/A
Moderate	51 (2.5)	4 (0.3)	N/A
Mild	142 (7.0)	29 (2.5)	N/A
None	1844 (90.5)	1144 (97.2)	N/A
Total	2037 (100.0)	1171 (100.0)	N/A
Diarrhea			<.001
Severe	6 (0.3)	1 (0.1)	N/A
Moderate	13 (0.6)	0 (0.0)	N/A
Mild	181 (8.8)	87 (7.4)	N/A
None	1853 (90.3)	1095 (92.6)	N/A
Total	2053 (100.0)	1183 (100.0)	N/A
Loss of appetite			<.001
Severe	10 (0.5)	2 (0.2)	N/A
Moderate	134 (6.7)	25 (2.1)	N/A
Mild	449 (22.5)	154 (12.9)	N/A
None	1407 (70.4)	1009 (84.8)	N/A
Total	2000 (100.0)	1190 (100.0)	N/A

[a]The Wilcoxon signed-rank test was performed.

[b]N/A: not applicable.

Table 4. Patient satisfaction with Korean Medicine telemedicine program for COVID-19.

Category	Patients, n	Mean score (SD) out of 10	Median score (IQR) out of 10
Satisfaction in terms of treatment	1570	8.3 (1.78)	8.0 (7.0-10.0)
Convenience of KM telemedicine center system	1568	9.3 (1.27)	10.0 (9.0-10.0)
Satisfaction with using telephone	1570	8.7 (1.59)	9.0 (8.0-10.0)
Willingness to recommend to acquaintances	1570	9.2 (1.58)	10.0 (9.0-10.0)
Willingness to use Korean Medicine treatments	1569	9.1 (1.54)	10.0 (8.0-10.0)

Discussion

Principal Findings

This study introduced the KM telemedicine center, which used Korean Medicine treatments for COVID-19 and reported the clinical symptoms of patients with COVID-19 before and after treatment. Among the confirmed cases enrolled in the KM telemedicine center, 75.5% (n=1755) Table 1) were women, while 60% of the total cumulative patients with confirmed COVID-19 in South Korea were women [11]. This difference may be because COVID-19 spread rapidly through certain religious groups that had a higher proportion of women. In addition, women's preferences regarding health care services and Korean Medicine treatment may have had an impact. The majority of patients that visited the KM telemedicine center were in Daegu-Gyeongbuk because the number of patients in that region was high until June 2020. Patients with mild illness could not be hospitalized and were isolated in their homes or other facilities; therefore, the demand for telemedicine increased.

Of the total 2324 cases, Qingfei Paidu decoction, Yiqi Bufei Tang, Ziyin Bufei Tang, and Kyung-Ok-Ko were most frequently provided to patients. As 86.4% (n=2009) of patients were in the recovery stage, Kyung-Ok-Ko, Yiqi Bufei Tang, and Ziyin Bufei Tang may be prescribed in large proportions (Figure 3). It has recently been elucidated that Qingfei Paidu decoction, which is the most recommended herbal medicine for treating COVID-19, regulates cytokine storms during the viral infection [12,13]. In addition, a clinical study in China of Qingfei Paidu decoction in addition to Western medicine was recently published [14]. This study showed that in addition to respiratory symptoms, protocols of herbal medicines have been prepared for digestive symptoms, muscle pain, and loss of taste, which are all known clinical symptoms of COVID-19. Although the results showed a statistical difference before and after treatment, it is hard to guarantee the effectiveness of herbal medicines for COVID-19 symptoms because many patients did not have symptoms at baseline. Any patient currently positive for, recovering from, or suspected of having COVID-19 could be enrolled in this study, whether symptoms were present or not. For that reason, there were many patients that responded "No" to each symptom, but many others noted having more than one symptom. It is significant that patient satisfaction with the KM telemedicine center was high. This may be because the center treated patients' symptoms and provided emotional support through counselling.

The average score for patient satisfaction at the KM telemedicine center was 8.3. The reason the score was high may be because the center supported the treatment of patients with acute symptoms. In South Korea, there were a large number of confirmed patients in Daegu-Gyeongbuk, such that the health system's capacity was exceeded. Patients with mild COVID-19 cases were not provided with medical management, and 5 self-quarantined patients in South Korea died at home [15]. The KM telemedicine center was established to fill the gap in health care that arises from these health care surges for patients with mild cases. However, patients gave lower scores for treatment satisfaction at the KM telemedicine center, compared with the results of other satisfaction score criteria. Although patients indicated it was convenient to use the KM telemedicine center (9.3), their satisfaction with telemedicine treatment (8.7) and satisfaction in terms of treatment were relatively low (8.3). This is because non–face-to-face treatments such as telephone treatment are limited compared to face-to-face treatment, which may lead to less satisfaction from patients. This shows that there are many challenges to solve in telemedicine, such as patient access to technology, the possibility of missing test results, and the lack of technology to replace medical devices [16,17].

There are several limitations in this study. First, only patient-reported outcomes were evaluated at the KM telemedicine center. Temperature and pulse rate were also self-measured; however, these could not be reported because of missing data and errors. In addition, there were a lot of missing data points regarding each symptom because this telemedicine center had not been designed for a study and this research is based on retrospective chart review. Second, this study showed the results of the entire program, so it could not show the effect of individual herbal medicines. Since the prescription was changed according to the clinical situation of the patients, we could not compare clinical symptoms before and after each herbal medicine was administered. Third, it was difficult to coordinate care with conventional treatment methods to manage patients with underlying diseases because the KM telemedicine center did not officially belong to the national quarantine system. Fourth, it was not possible to determine the color and shape of the tongue or face color over the telephone, all of which are meaningful in Korean Medicine theory.

Nonetheless, the KM telemedicine center attempted to test the weaknesses of telemedicine and improve the quality of medical care offered via telemedicine. It has developed guidelines to help medical staff perform standardized health care. The second edition of the guideline was published on April 1, 2020, and was complemented with feedback from the medical and advisory groups, containing considerations from real telemedicine cases. Above all, Korean Medicine is traditionally compatible with telemedicine because it collects information for diagnosis by listening to the patient's overall symptoms. In addition, an advisory committee composed of professors and experienced clinicians reviewed all patients' medical charts to verify telephone consultations of KMDs. These may have led people to feel more satisfied with the KM telemedicine center system despite the limitations of non–face-to-face treatment.

Conclusion

This study described the background and operational result of the KM telemedicine center for COVID-19 in South Korea. The KM telemedicine center attempted to overcome the limitations of telemedicine by providing standardized guidelines and expert advice. We hope that this study will help to establish a better health care system to overcome COVID-19.

Acknowledgments

The authors thank all patients who voluntarily called for telemedicine, and we sincerely hope for their full recovery. The authors are grateful to all participating Korean Medicine doctors, students, and staff, and the advisory group. The authors are also grateful for the support and cooperation between the COVID-19 telemedicine center of Korean Medicine and The Association of Korean Medicine. This article was supported by the Korea Institute of Oriental Medicine (KSN2021210 and KSN2021220).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Composition of herbal medicines provided by the COVID-19 telemedicine center of Korean Medicine.

[DOCX File , 19 KB - publichealth_v7i1e20236_app1.docx]

References

1. Coronavirus disease 2019 (COVID-19) Situation Report - 39. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-12-30]
2. Coronavirus disease 2019 (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. World Health Organization. URL: https://reliefweb.int/report/world/coronavirus-disease-covid-19-weekly-epidemiological-update-17-november-2020 [accessed 2020-12-30]
3. The Association of Korean medicine. Recommendations on COVID-19 Korean medicine. 2020. URL: https://tinyurl.com/y55fde97 [accessed 2020-12-30]
4. The KM Professor Council of Internal Medicine of the Respiratory System. Clinical Guidelines on COVID-19 Korean Medicine. 2020. URL: https://tinyurl.com/y4yv8j7u [accessed 2020-12-30]
5. National Health Commission of the People's Republic of China. [Notice on the Issuance of the New 7th Version of the COVID-19 Diagnosis and Treatment Guidelines]. 2020. URL: http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml [accessed 2020-12-30]
6. Shen A, Zhang W, We Z, Wang W, Hua J. [TCM Theory Analysis of Qingfei Paidu Decoction Treating COVID-19]. Liaoning Journal of Traditional Chinese Medicine 2020;47(3):106-108.
7. Lu ZZ, Lu XS. [Qingfei Paidu Decoction Demonstrates the Anti-Epidemic Effects and Self-Confidence of Traditional Chinese Medicine]. Journal of Traditional Chinese Medicine 2020;61(10):833-834.
8. Ang L, Lee HW, Choi JY, Zhang J, Soo Lee M. Herbal medicine and pattern identification for treating COVID-19: a rapid review of guidelines. Integr Med Res 2020 Jun;9(2):100407 [FREE Full text] [doi: 10.1016/j.imr.2020.100407] [Medline: 32289016]
9. National Health Commission of the People's Republic of China. [Novel Coronavirus Pneumonia Diagnosis and Treatment Plan, provisional 6th Edition]. 2020. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml [accessed 2020-12-30]
10. Park S, Park YC. Mental Health Care Measures in Response to the 2019 Novel Coronavirus Outbreak in Korea. Psychiatry Investig 2020 Feb;17(2):85-86 [FREE Full text] [doi: 10.30773/pi.2020.0058] [Medline: 32093458]
11. Korea Centers for Disease Control Prevention. [COVID-19 Response Guideline (for local governments), 9th edition]. 2020. URL: http://www.kdca.go.kr/board/board.es?mid=a20507020000&bid=0019 [accessed 2020-12-30]
12. Chen J, Wang Y, Gao Y, Hu L, Yang J, Wang J, et al. Protection against COVID-19 injury by qingfei paidu decoction via anti-viral, anti-inflammatory activity and metabolic programming. Biomed Pharmacother 2020 Oct;129:110281 [FREE Full text] [doi: 10.1016/j.biopha.2020.110281] [Medline: 32554251]
13. Yang R, Liu H, Bai C, Wang Y, Zhang X, Guo R, et al. Chemical composition and pharmacological mechanism of Qingfei Paidu Decoction and Ma Xing Shi Gan Decoction against Coronavirus Disease 2019 (COVID-19): In silico and experimental study. Pharmacol Res 2020 Jul;157:104820 [FREE Full text] [doi: 10.1016/j.phrs.2020.104820] [Medline: 32360484]

14. Xin S, Cheng X, Zhu B, Liao X, Yang F, Song L, et al. Clinical retrospective study on the efficacy of Qingfei Paidu decoction combined with Western medicine for COVID-19 treatment. Biomed Pharmacother 2020 Sep;129:110500 [FREE Full text] [doi: 10.1016/j.biopha.2020.110500] [Medline: 32768975]
15. Press Release. SBS News. 2020. URL: https://news.sbs.co.kr/news/endPage.do?news_id=N1005681219 [accessed 2020-12-30]
16. Elkbuli A, Ehrlich H, McKenney M. The effective use of telemedicine to save lives and maintain structure in a healthcare system: Current response to COVID-19. Am J Emerg Med 2020 Apr 07 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.003] [Medline: 32303410]
17. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]

Abbreviations

AKOM: The Association of Korean Medicine
KM: Korean Medicine
KMD: Korean Medicine doctor
TCM: Traditional Chinese Medicine
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 14.05.20; peer-reviewed by SG Ko, B Lim, B Smith; comments to author 19.08.20; revised version received 24.09.20; accepted 16.12.20; published 19.01.21.

Please cite as:
Jang S, Kim D, Yi E, Choi G, Song M, Lee EK
Telemedicine and the Use of Korean Medicine for Patients With COVID-19 in South Korea: Observational Study
JMIR Public Health Surveill 2021;7(1):e20236
URL: http://publichealth.jmir.org/2021/1/e20236/
doi:10.2196/20236
PMID:33342765

Original Paper

Knowledge About COVID-19 in Brazil: Cross-Sectional Web-Based Study

Vinícius Henrique Almeida Guimarães[1*], BMedSci; Maísa de Oliveira-Leandro[2*], BMedBiol; Carolina Cassiano[1], BNSc; Anna Laura Piantino Marques[3], BSE; Clara Motta[3], BSE; Ana Letícia Freitas-Silva[1], BMedSci; Marlos Aureliano Dias de Sousa[1], MHS, MD; Luciano Alves Matias Silveira[1], MHS, MD; Thiago César Pardi[1], MHS, MD; Fernanda Castro Gazotto[1], MD; Marcos Vinícius Silva[2], BMedBiol, MHS, PhD; Virmondes Rodrigues Jr[2], MHS, MD, PhD; Wellington Francisco Rodrigues[2], BMedBiol, MHS, PhD; Carlo Jose Freire Oliveira[2], BMedBiol, MHS, PhD

[1]Institute of Health Sciences, Federal University of Triangulo Mineiro, Uberaba, Brazil
[2]Laboratory of Immunology and Bioinformatics, Institute of Biological and Natural Sciences, Federal University of Triangulo Mineiro, Uberaba, Brazil
[3]Institute of Language Studies, University of Campinas, Campinas, Brazil
[*]these authors contributed equally

Corresponding Author:
Carlo Jose Freire Oliveira, BMedBiol, MHS, PhD
Laboratory of Immunology and Bioinformatics
Institute of Biological and Natural Sciences
Federal University of Triangulo Mineiro
Uberaba
Brazil
Phone: 55 34988567251
Email: carlo.oliveira@uftm.edu.br

Abstract

Background: COVID-19 is a highly transmissible illness caused by SARS-CoV-2. The disease has affected more than 200 countries, and the measures that have been implemented to combat its spread, as there is still no vaccine or definitive medication, have been based on supportive interventions and drug repositioning. Brazil, the largest country in South America, has had more than 140,000 recorded deaths and is one of the most affected countries. Despite the extensive quantity of scientifically recognized information, there are still conflicting discussions on how best to face the disease and the virus, especially with regard to social distancing, preventive methods, and the use of medications.

Objective: The main purpose of this study is to evaluate the Brazilian population's basic knowledge about COVID-19 to demonstrate how Brazilians are managing to identify scientifically proven information.

Methods: A cross-sectional study design was used. An original online questionnaire survey was administered from June 16 to August 21, 2020, across all five different geopolitical regions of the country (ie, the North, Northeast, Center-West, Southeast, and South). The questionnaire was comprised of questions about basic aspects of COVID-19, such as the related symptoms, conduct that should be followed when suspected of infection, risk groups, prevention, transmission, and social distancing. The wrong questionnaire response alternatives were taken from the fake news combat website of the Brazilian Ministry of Health. Participants (aged ≥18 years) were recruited through social networking platforms, including Facebook, WhatsApp, and Twitter. The mean distributions, frequencies, and similarities or dissimilarities between the responses for the different variables of the study were evaluated. The significance level for all statistical tests was less than .05.

Results: A total of 4180 valid responses representative of all the states and regions of Brazil were recorded. Most respondents had good knowledge about COVID-19, getting an average of 86.59% of the total score with regard to the basic aspects of the disease. The region, education level, age, sex, and social condition had a significant association (P<.001) with knowledge about the disease, which meant that women, the young, those with higher education levels, nonrecipients of social assistance, and more economically and socially developed regions had more correct answers.

Conclusions: Overall, Brazilians with social media access have a good level of basic knowledge about COVID-19 but with differences depending on the analyzed subgroup. Due to the limitation of the platform used in carrying out the study, care should be taken when generalizing the study findings to populations with less education or who are not used to accessing social networking platforms.

(JMIR Public Health Surveill 2021;7(1):e24756) doi:10.2196/24756

KEYWORDS

COVID-19; coronavirus; perception; knowledge; Brazil; cross-sectional; online survey; health information

Introduction

COVID-19 is a highly transmissible multi-organ viral disease caused by SARS-CoV-2, a new coronavirus [1]. The most severe cases can be fatal and are present in risk groups that include males, older adults, people who are obese, and patients with other comorbidities [2]. The disease is currently the largest public health issue worldwide, having reached, since March 11, 2020, the status of a global pandemic [3]. The virus can be transmitted from person to person through droplets, aerosols, airborne routes, and contaminated surfaces. The most common symptoms of infection are fever, dry cough, fatigue, headache, loss of smell, and shortness of breath [1]. The disease does not yet have a vaccine or definitive treatment. For this reason, measures such as social distancing, proper hygiene, and the use of individual and collective protective equipment have been instituted by different health authorities, which have been shown to be central to preventing the transmission of the virus and controlling the spread of the disease [4,5]. In addition, knowledge about the infection and its signs and symptoms, whether by the general population or by health professionals, has also been shown to be effective in aiding early diagnosis, better monitoring, and more effective treatment [4].

Brazil is a country of continental dimensions and, in addition to its geographical and cultural differences found within its borders, presents significant economic and educational vulnerabilities. Since the appearance of the country's first case of COVID-19, much discussion has ensued on how best to face the disease, especially with regard to social distancing (eg, whether "vertical isolation" or "horizontal isolation" should be practiced), the use of medications without World Health Organization approval (eg, azithromycin, ivermectin, and hydroxychloroquine), and the monitoring of the disease from the onset of symptoms to the admission of the patient to a hospital specialized in treating the infection [6]. In addition to these discussions and related challenges, miraculous "cures," inconsistency between policies and scientific evidence, conspiracy theories, and increases in fake news have been widely disseminated on social networks, which has caused confusion among the general population and hindered the fight against the disease.

Many countries have sought to understand all there is to know about the pandemic to better fight this dangerous disease. For this reason, several researchers have conducted studies to track the public's knowledge and misperceptions regarding COVID-19. Studies with this focus have already been conducted among general and specific populations in mainland China [7], Colombia [8], Hong Kong [9], India [10], Iran [11], Israel [12], the United Kingdom [13], and the United States [14,15], among others. These studies, despite the differences in their findings, clearly demonstrate that populations present a certain level of knowledge about COVID-19. On the other hand, these studies have also revealed how much the disease has had economic, psychosocial, and behavioral impacts that also need to be mitigated. At the time of writing (September 2020), Brazil is the third country with the highest number of confirmed cases and has had more than 140,000 deaths from the disease [16]; however, no studies have evaluated the population's basic knowledge about COVID-19. Thus, this study seeks to evaluate the public's knowledge and misperceptions about COVID-19 and the preventive measures adopted to date in the country.

Methods

Participants

A cross-sectional anonymous online survey (see Multimedia Appendix 1) was carried out using Google Forms, a service for form and questionnaire creation that is free for everyone who has a Google account. This tool allows for the creation of different types of questions, the collection and organization of the responses received, and generation of spreadsheets and graphs of the final data in real time.

In an attempt to make possible the implementation of the research and aiming at easy access to the online survey, respondents were recruited via the divulgation of information regarding the research on the university's and researcher's social medias. With the expectation of reaching the largest possible number and diversity of people, the disclosure was made in four of the main social medias used in Brazil: WhatsApp, Instagram, Facebook, and Twitter. For a better representation of the overall Brazilian population, the researchers also used promotional tools on these social networks—paying for advertisements to enable the survey form to reach different audiences from all regions of the country. According to the Brazilian Institute of Geography and Statistics, the population of Brazil in 2020 reached 211.8 million, of whom around 134 million have access to the internet. Thus, according to statistical analysis, a sample number of 2500 participants would be representative of the population using internet in the country, with a 2% margin of error and 95% confidence level. Still, in addition to the stages of confusion and risk of bias control, the data was evaluated to 64% above the estimated sample number, making a total of 4100 participants distributed in the 5 macroregions of Brazil.

The online form was available for about 2 months between June 16 and August 21, 2020, and can be found in its full version in Multimedia Appendix 1. This specific time period was selected because it was the peak of the pandemic's "first wave" in the country (ie, the period when, for the first time, the pandemic reached a peak in cases and deaths) [17]. The 2-month availability was due to the geographical extent of Brazil and the need for representativity of the population from each region. As a country of continental dimensions, the disease has not spread homogeneously throughout the country. In addition, cities far from the research centers that the researchers belonged were more difficult to access.

This study was approved by the ethics and research committee of the Federal University of Triângulo Mineiro in Minas Gerais State, Brazil. Upon access to the web-based survey form, respondents were provided with an explanation as to the purpose of the research as well as the prerequisites for participation. Potential participants could decide freely whether to participate in the study. Those aged ≥18 years, the target study participants, were then asked to select the option of electronically signing the free and informed consent form.

In Brazil, the legislation defines 18 years as the age of majority, making the individual fully capable to respond by himself. Considering that the research was online and there was an urgent need for it to be carried out at the height of the COVID-19 spread in the country, there would be a greater bureaucratic obstacle if the research had to include younger individuals because more documents would need to be filled out and analyzed by the ethics committee. This group is also considered vulnerable, meaning that it would be necessary to have authorization from a guardian older than 18 years. Therefore, we chose to recruit only participants aged ≥18 years. If the participants consented, they were directed to answer the questions on the form. There was no financial compensation for participants who responded to the survey; thus, participation was voluntary and anonymous.

The survey was elaborated on considering data from the official website of the Brazilian Ministry of Health, the country's highest health authority responsible for organizing and preparing public plans and policies aimed at health across the country. Thus, the basis for outlining the questions and the correct alternatives was developed according to the information available on the COVID-19 webpage of the Brazilian Ministry of Health [18]. To elaborate on the incorrect alternatives and seeking to address some difficulties faced by Brazilians in terms of understanding the disease, the website of the Brazilian Ministry of Health was used. This website compiles some fake news about the disease to discuss its flaws from a scientific point of view and to clarify it for the population [19].

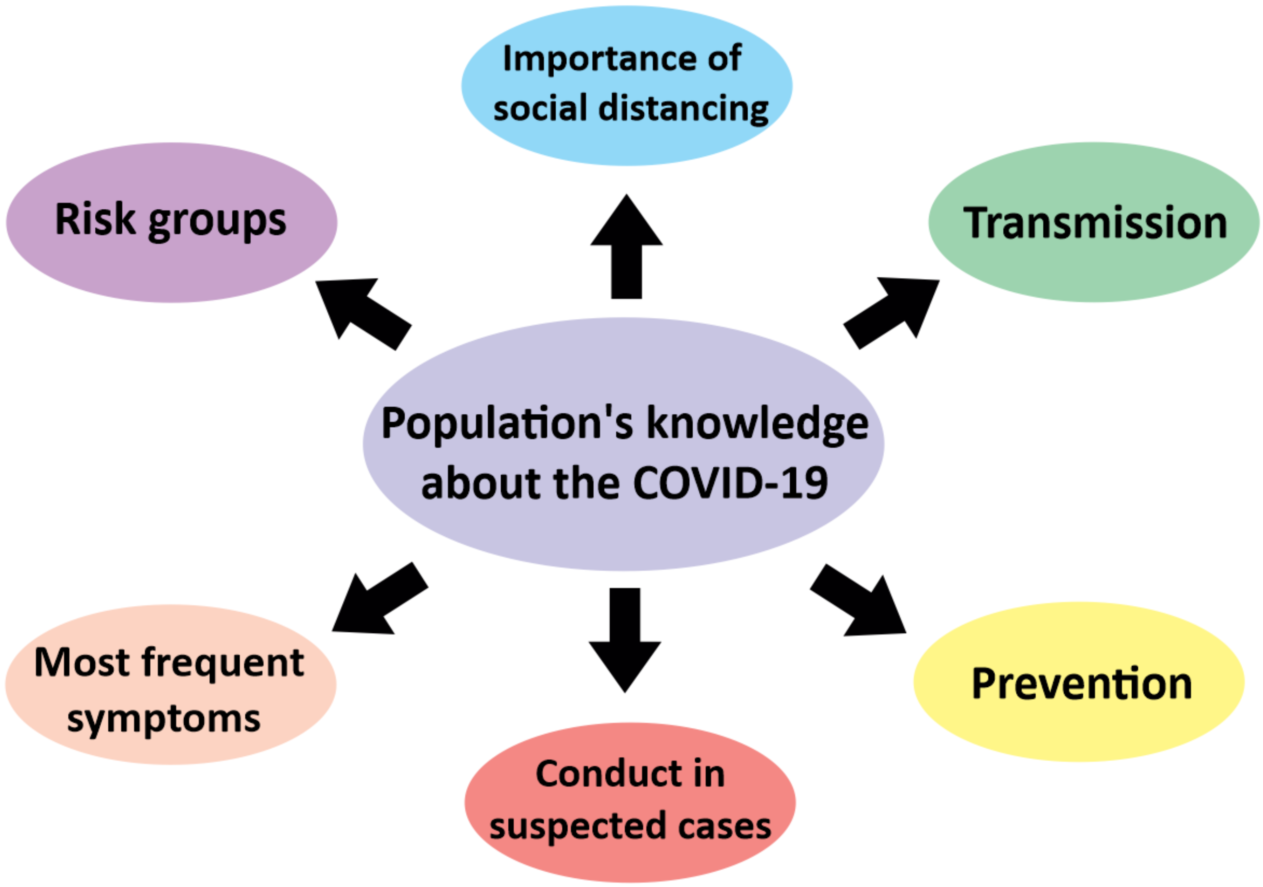
As the objective was to understand the Brazilian residents' knowledge about COVID-19, the questionnaire that participants had access to was in the official language spoken and written in all regions of the country, the Portuguese language. After the completion of data collection, an English translation of the questions and alternatives was provided for this publication (see Multimedia Appendix 1).

Data Collection and Quality Assurance

At the end of the data disclosure period, we stopped collecting questionnaire responses on Google Forms. We then obtained a spreadsheet with all the survey data, with each row representing the responses of a participant and each column representing the answers to a question. Using a filter tool available in Excel (Microsoft Corporation), it was possible to identify and exclude both the responses of participants who declared they were younger than 18 years and the responses of participants who declared they were not residents of Brazil.

In the survey form, participants were asked 26 questions that were divided into five main blocks of themes: (1) general information about COVID-19 (ie, questions about transmission, most common symptoms, conduct in case of infection, risk groups, and social distancing), (2) sharing information about the disease (ie, questions about how participants obtain information about COVID-19, what information they receive and share, and how they check and analyze the news they find on social media), (3) identification of misinformation (analyzing whether the participants recognize fake news about the disease), (4) economic and social impact of the pandemic (ie, questions about the main family fears and challenges, and what most hinders the search for information about the disease), and (5) sociodemographic information about the participants (ie, age, sex, education level, region of residence, profession, whether they were recipients of government benefits, and number of people living in the same household). In this study, only the questions that directly assessed the participants' basic knowledge in health education about COVID-19 (see Figure 1) were referred to in the analysis; specifically, this included questions on the forms of transmission (question 1), main symptoms of the disease (question 2), conduct in a suspected case (question 3), identification of risk groups (question 4), understanding of social distancing (question 5), and disease prevention (question 12). The other questions did not directly evaluate knowledge about health and disease, and were not referred to in the analysis. The number of alternatives for each question took into account the proportion of fake news available on the website of the Ministry of Health. Thus, questions that addressed more fake news verified by the website were considered, proportionally, as issues of greater confusion among the population, receiving more alternatives and presenting greater participation in the final score. Before the evaluation of the participants' responses, a pilot test with 200 participants (randomized by automation) was carried out to ensure reliability. In our study, the Cronbach alpha was .66 for general information about COVID-19, .94 for sharing information about the disease, .88 for identification of misinformation, .70 for economic and social impact, and .68 for sociodemographic information about the participants. Quality assurance was accomplished by checking, by at least two independent evaluators, the data collection, extraction and entry to the software, and data analysis.

Figure 1. Basic health knowledge investigated in the study.



To count the final score of each participant in the six questions chosen for analysis, we classified the questions based on their format (checkboxes or multiple choice). Questions 1, 2, 4, and 12 were arranged in a checkbox format, and the participant could select more than one alternative. Questions 3 and 5 were in the multiple choice format, with the participant being able to choose only one alternative.

In the questions arranged in checkboxes, all alternatives, correct or incorrect, were evaluated. There were four possibilities for selecting alternatives: (1) the alternative was correct and was selected, (2) the alternative was correct and was not selected, (3) The alternative was incorrect and was not selected, and (4) the alternative was incorrect and was selected. In options 1 and 3, the participant got the choices right and scored 1 point. In options 2 and 4, the participant made a mistake in the choices and scored 0 points. That is, the participants score 1 point both when choosing which alternative is correct and when choosing which is incorrect. Thus, the maximum score for these questions corresponded to the total number of alternatives in the question and indicated that the participant marked the correct alternatives and did not mark the incorrect ones, scoring 1 point for all alternatives. Question 1 was composed of 7 alternatives, with a maximum value of 7 points; question 2 had 9 alternatives, with a maximum value of 9 points; question 4 had 10 alternatives, with a maximum value of 10 points; and question 12 had 11 alternatives, with a maximum value of 11 points.

In multiple choice questions, there is a limitation in the number of checked alternatives and only one could be selected. There were only two possibilities for selecting alternatives: (1) the alternative was correct and was selected and (2) the alternative was incorrect and was selected. In option 1, the participant got it right and scored 1 point. In option 2, the participant made a mistake and scored 0 points. Therefore, the participants only scored points when they selected the correct alternative. Thus, the maximum score for these questions was always 1 and indicated that the participant chose the correct alternative. Therefore, although questions 3 and 5 have 3 alternatives each, the maximum value of these questions was 1, indicating the selection of the correct alternative.

Taking into account the maximum value that questions in checkboxes and multiple choice could receive, the participant could score between 0 and 39 points in total. Simply put, Higher scores had more correct answers, meaning more knowledge about COVID-19 was evident. Scores between 0-9 were regarded as *poor knowledge* about the disease, scores between 10-19 were regarded as *regular knowledge*, scores between 20-29 were regarded as *good knowledge*, and scores between 30-39 were regarded as *optimal knowledge*. The required knowledge investigated in this study can be considered basic, as they mainly concerned practical aspects of people's day-to-day life, not entering into the theoretical or scientific merits of a complex multi-organ disease such as COVID-19.

At the end of the questionnaire, participants had access to some links that directed them to sources of scientifically safe information about the disease, such as the official website of the Brazilian Ministry of Health [18,19] and the website of a main research institution in infectious diseases in the country, the Oswaldo Cruz Foundation (FIOCRUZ) [20]. Participants also had the possibility to check their own answers, the right and wrong answers, and the explanation for each alternative.

Statistical Analysis

The data were tabulated using Excel and analyzed using SPSS 21 (IBM Corp) and GraphPad Prism 7.0 (GraphPad Software, Inc). The data were evaluated for their distribution (using D'Agostino-Pearson and Shapiro-Wilk normality tests), and the variances were compared (using the *F* test and Bartlett test). Unpaired tests to compare the distributions of the different variables were used (Kruskal-Wallis with Dunn multiple comparisons test and Mann-Whitney *U* test). The hypotheses were tested using chi-square, Fisher exact, and chi-square with Yates correction tests. To assess the association measures, odds ratios (Baptista-Pike) with their respective confidence intervals were used in the definitive analysis.

To assess the effect of associations between the tested variables in the third table (ie, transmission, symptoms, conduct in suspected infection, risk groups, social distancing, and prevention), the lowest scores (the poor outcomes) were compared with the other scores (the best outcomes) between the descriptions for each variable. For the grouped variables, the scores were normalized in relative frequencies and were compared with the scores up to 50%, with the others (above 50%) between the descriptions for each variable. Multivariate analysis was performed to determine the hierarchical groupings of the different variables. After adjusting the proximity matrix using the squared Euclidean distance, the results were plotted on a dendrogram. Spearman test was used to investigate correlations. The significance levels in all statistical tests were less than .05 (5%) [21].

Results

A total of 4436 responses were received; however, 17 were excluded from the analysis due to the respondents having been from other countries, and 239 were excluded for having been filled out by people younger than 18 years, thus leaving a total of 4180 valid responses. Of these valid responses, 2051 (49.07%) came from the Southeast, 871 (20.84%) from the Northeast, 697 (16.67%) from the South, 285 (6.82%) from the North, and 276 (6.60%) from the Central-West geopolitical regions of Brazil (Figure 2). The average age of respondents was 34.57 years; 2937 (70.26%) were women, 2040 (48.80%) held a bachelor's degree or above, and 3504 (83.83%) lived with a maximum of four people in the same house. Among the respondents, 3252 (77.80%) stated that they did not receive any kind of government assistance. Most (n=3641, 87.11%) had not traveled to other countries in the past year. These and other demographic information are shown in Table 1.

Figure 2. Distribution of the Brazilian population by region and its relationship with the distribution of the study population.

REGION	ABSOLUTE BRAZILIAN POPULATION (% OF TOTAL POPULATION)	ABSOLUTE SAMPLE POPULATION (% OF TOTAL SAMPLE)
North	18 672 591 (**8.82**)	285 (**6.82**)
Northeast	57 374 243 (**27.09**)	871 (**20.84**)
Center-west	16 504 303 (**7.79**)	276 (**6.60**)
Southeast	89 012 240 (**42.04**)	2051 (**49.07**)
South	30 192 315 (**14.26**)	697 (**16.67**)

Table 1. Demographic characteristics of participants.

Characteristic	Participants (N=4180), n (%)
Sex	
Male	1243 (29.74)
Female	2937 (70.26)
Age (years)	
18-19	315 (7.54)
20-29	1718 (41.10)
30-39	761 (18.21)
40-49	583 (13.95)
50-59	558 (13.35)
≥60	245 (5.86)
Region	
North	285 (6.82)
Northeast	871 (20.84)
Central-West	276 (6.60)
Southeast	2051 (49.07)
South	697 (16.67)
Education	
Middle and high school	2140 (51.20)
Higher and postgraduate education	2040 (48.80)
Household size	
1 person	354 (8.47)
2 people	841 (20.12)
3 people	1177 (28.16)
4 people	1132 (27.08)
5 people	442 (10.57)
6 people or more	234 (5.60)
Receives government social assistance	
Yes	928 (22.20)
No	3252 (77.80)

Figure 2 illustrates the distribution of the participants by region. For the total score, measured between 0 and 39 possible points, the average score of the participants was 33.77 points, varying between 20 and 39 points in total, depending on the respondent. This means that, on average, the participants reached 86.59% of the total possible score. Table 2 shows the distribution of responses (true or false) for each response alternative to the questions presented. In the table, for each alternative, we can observe if the item was considered false or true (in parenthesis) and the number of people who appropriately marked it.

Table 2. Questionnaire of knowledge about COVID-19.

Questions	Participants (N=4180), n (%)
What are the main forms of transmission of COVID-19?	
Through sneezing, coughing, or talking to infected people (true)	4082 (97.66)
Direct contact with domestic animals (false)	132 (3.16)
Bringing hand to face after touching contaminated surfaces (true)	3765 (90.07)
Bites from contaminated insects (false)	26 (0.62)
Taking filtered water in cities with many cases of infection (false)	232 (5.55)
Using products that came from China, where the coronavirus appeared (false)	85 (2.03)
Contact with contaminated people (eg, kiss, hug, or handshake; true)	3954 (94.59)
What are the three most common symptoms of COVID-19?	
Diarrhea and vomiting (false)	418 (10.00)
Skin wounds (false)	19 (0.45)
Persistent fatigue (true)	1250 (29.90)
Stuffy nose (false)	270 (6.46)
Fever (true)	3775 (90.31)
Shortness of breath (false)	3410 (81.58)
Cough (true)	3192 (76.36)
Headache (false)	1144 (27.37)
Sneezing (false)	542 (12.97)
What is the possible conduct after infection?	
The virus is not that dangerous, so you can continue your life normally (false)	6 (0.14)
You should be isolated at home and seek help if you feel short of breath or get worse (true)	3338 (79.86)
You must immediately go to the hospital to seek medical attention (false)	836 (20.00)
Which risk groups are most likely to get infected?	
People with heart or kidney problems (true)	2806 (67.13)
People with vision problems (eg, blindness or myopia; false)	19 (0.45)
Wheelchair users (false)	85 (2.03)
People with respiratory diseases and smokers (true)	4070 (97.37)
Older adults (true)	3995 (95.57)
People with cancer (true)	2668 (63.83)
Adolescents and young adults (false)	59 (1.41)
People with diabetes or high blood pressure (true)	3778 (90.38)
Pregnant women (true)	1531 (36.63)
There are no risk groups (false)	24 (0.57)
Importance of social distancing	
Necessary (true)	4101 (98.11)
Makes no difference (false)	66 (1.58)
Harmful (false)	13 (0.31)
Which alternatives are true about COVID-19?	
There is already a vaccine against COVID-19 (false)	226 (5.41)
Wearing gloves and masks for everyday activities decreases the chance of becoming infected with the virus (true)	3590 (85.89)
Gargling with warm water, salt, and vinegar prevents coronavirus (false)	72 (1.72)

Questions	Participants (N=4180), n (%)
Hot water or tea kills the coronavirus (false)	46 (1.10)
70% gel alcohol kills the coronavirus (true)	3678 (87.99)
Chloroquine protects people from becoming infected with the coronavirus (false)	114 (2.73)
There are already drugs that cure COVID-19 (false)	143 (3.42)
Soap, sanitary water, liquid alcohol, and common detergents kill the coronavirus (true)	3404 (81.44)
Drinking alcohol kills the virus (false)	8 (0.19)
Social distancing has no scientific proof (false)	140 (3.35)
Once the person has had the coronavirus infection, they cannot have it again because they are immune (false)	567 (13.56)

After determining the sociodemographic profile of the participants (Table 1) and the survey questions (Table 2), the percentage of the population's knowledge about COVID-19 with regard to the different research variables (Figure 1) was evaluated, including transmission, symptoms, conduct for suspected infection, risk groups, perception of social distancing, and prevention. Overall, the participants had a good perception of the COVID-19 outbreak since the percentage of correct answers was above 90% and never below 70% for some of the variables evaluated. There was no participant that had a poor or regular knowledge score. On the other hand, of the 4180 responses, 252 (6.03%) had good knowledge scores and 3928 (93.97%) had optimal knowledge scores. When the level of perception between each variable was assessed, a statistically significant difference (P<.001) was found between them, with knowledge about symptoms being the parameter with the lowest understanding by the respondents (73.08% of correct responses). This limited understanding of the symptoms of the disease was 14 times lower than the knowledge about the importance of social distancing. There was an important lack of understanding concerning the conduct to be taken in cases of suspected SARS-CoV-2 infection (20.14%), risk groups (15.36%), disease prevention (6.92%), and disease transmission (4.15%; Figure 3). It is important to note that all the variables studied were linked in the range of the squared Euclidean distance, and an intimate relationship was observed between social distancing, transmission, and prevention. On the other hand, in general, choosing the right or wrong answer did not respect these relationships, as the similarities of the average connections between the groups were not consistent (quadratic R^2=0.22; Figure 4).

Figure 3. Distribution and association among COVID-19 health education indicators. The differences or similarities between the participants' levels of correct answers on questions regarding the symptoms of COVID-19, the conduct of those suspected of SARS-CoV-2 infection, risk groups, disease prevention, disease transmission, and perception of social distancing. The relative distribution, in percentages, of the levels of right and wrong answers for each variable and comparisons between them is demonstrated. *Statistically significant differences between groups (Kruskal-Wallis test with Dunn multiple comparison tests).

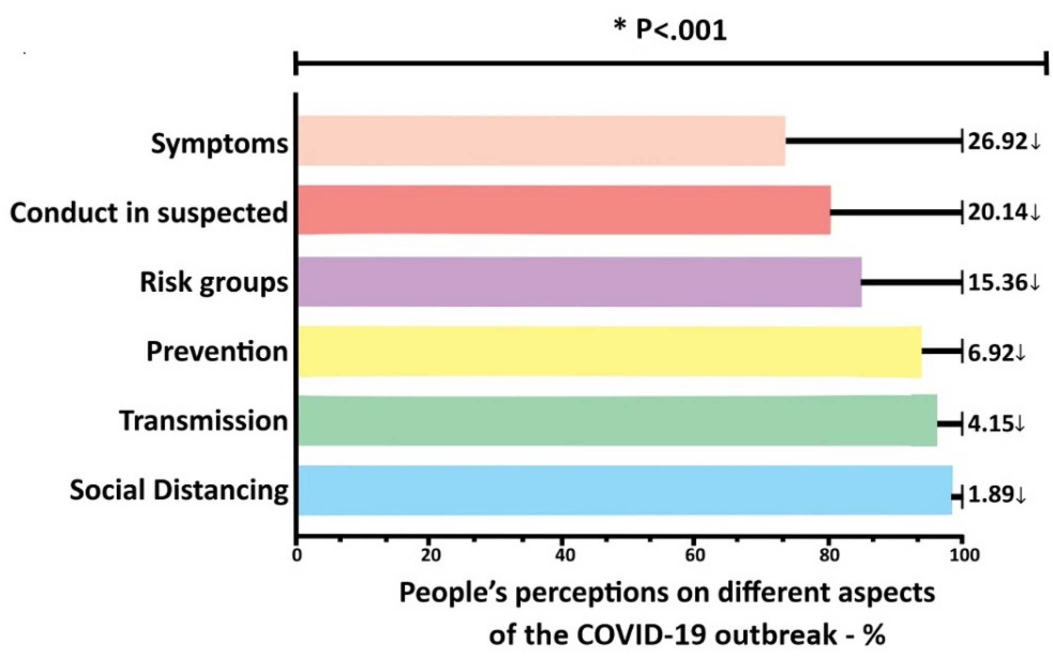
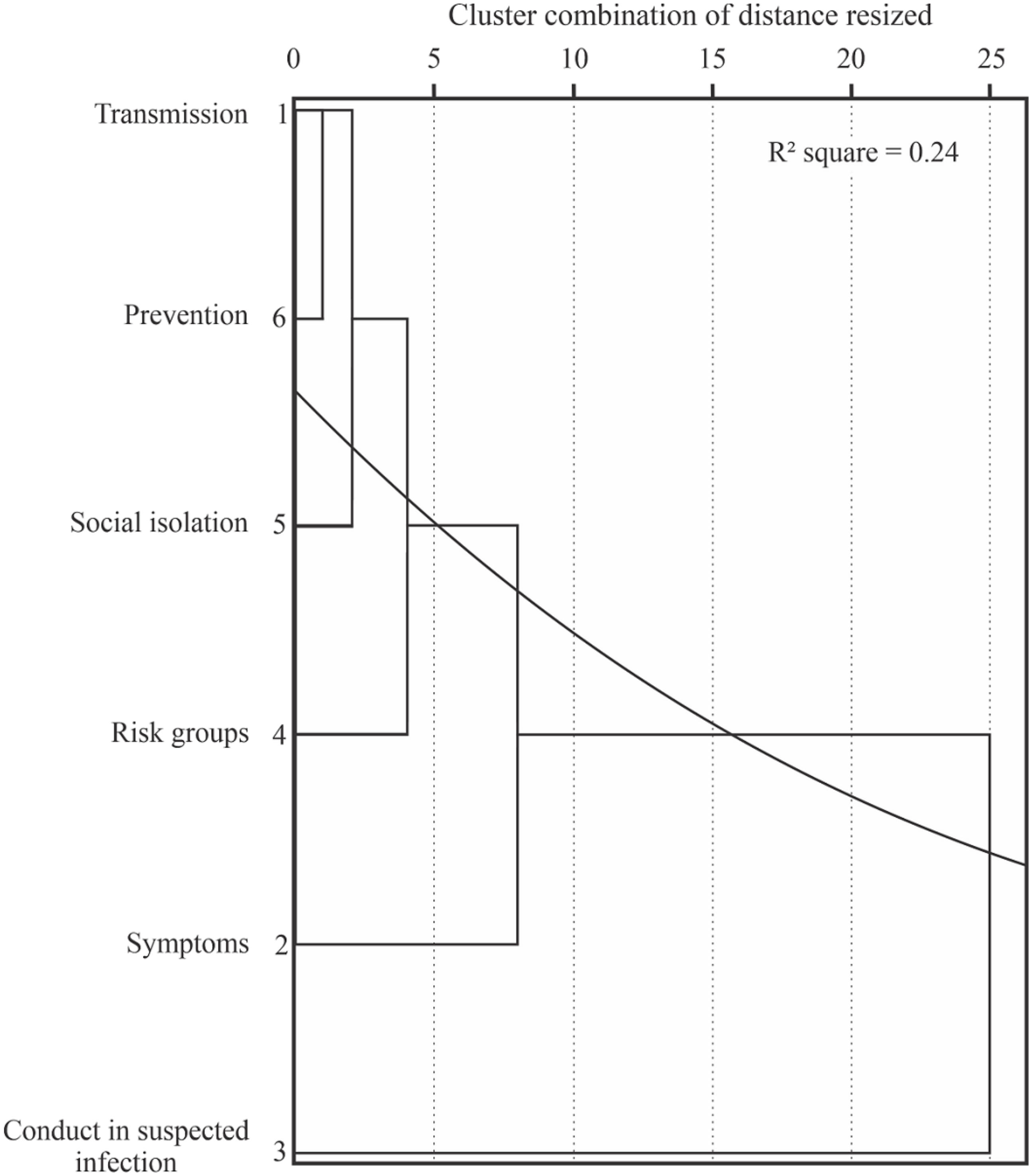


Figure 4. Research variables were combined in clusters and, after obtaining the square Euclidean distance and plotting on a dendrogram, the keys indicated the similarity between the variables evaluated in the survey. The central line indicates an adjustment by the square R2.



The step immediately after evaluating all the participants together was to investigate their perception of COVID-19 by considering their sociodemographic profiles and each research variable. Thus, each of the respondent's health education indicators was evaluated considering, for example, sex (male and female), education level (middle/high school and higher/postgraduate education), age (18-83 years), socioeconomic vulnerability (receiving or not receiving government support), and the number of people per residence (1-6). These indicators were evaluated for their respective distributions for each profile (unpaired), and the odds ratios between the worst outcomes (lowest score) were compared with those between the best outcomes. We also grouped the media scores of all parameters (ie, transmission, symptoms, conduct in the case of an infection, risk groups, and social distancing) and evaluated the odds ratios of them in each population profile (ie, sex, education level, age, socioeconomic vulnerability, people by residence, and Human Development Index [HDI]) to have a percentage of correct answers of up to 50%, as described in the Methods section and Table 3.

Table 3. Sample statistical analysis based in sociodemographic profiles.

Variables	Mean (SD)	M-W[a] test	K-W[b] test	Spearman *r* (95% CI)	OR[c] (95% CI)	*P* value
Transmission						
Sex		1,765,703	N/A[d]	N/A	0 (0 to 2.73)	*.01*[e]
Male	6.68 (0.63)					
Female	6.72 (0.64)					
Education level		2,041,706	N/A	N/A	1.91 (0.22 to 27.66)	*<.001*
Middle and high school	6.66 (0.69)					
Higher/postgraduate	6.76 (0.572)					
Age (years)		N/A	68.82		N/A	*<.001*
18-19	6.71 (0.76)[f,g]			N/A		N/A
20-29	6.80 (0.50)[g]			N/A		N/A
30-39	6.66 (0.69)[f,h]			N/A		N/A
40-49	6.60 (0.71)[h]			N/A		N/A
50-59	6.63 (0.70)[f,h]			N/A		N/A
≥60	6.60 (0.74)[f,h]			N/A		N/A
18-83	34.57 (14.01)			–0.08 (–0.12 to 0.05)		*<.001*
Socioeconomic vulnerability		1,412,808	N/A	N/A	1.75 (0.12 to 15.09)	*<.001*
RGS[i]	6.62 (0.75)					
Not RGS	6.73 (0.60)					
People in residence (n)		N/A	N/A	–0.02 (–0.05 to 0.01)	N/A	.26
1-40	3.35 (1.61)					
HDI[j]		N/A	N/A	0.10 (0.07 to 0.13)	N/A	*<.001*
0.683-0.850	0.779 (0.04)					
Symptoms						
Sex		1,825,002	N/A	N/A	1.03 (0.50 to 2.10)	.99
Male	6.57 (1.08)					
Female	6.58 (1.06)					
Education level		2,171,656	N/A	N/A	1.92 (0.95 to 3.99)	.73
Middle and high school	6.57 (1.09)					
Higher/Postgraduate	6.58 (1.04)					
Age (years)		N/A	18.66		N/A	*.002*
18-19	6.52 (1.15)[f,h]			N/A		N/A
20 to 29	6.65 (1.02)[f]			N/A		N/A
30 to 39	6.54 (1.03)[f,h]			N/A		N/A
40 to 49	6.57 (1.13)[f,h]			N/A		N/A
50 to 59	6.51 (1.12)[h]			N/A		N/A
≥60	6.44 (1.08)[h]			N/A		N/A
18-83	34.57 (14.01)			–0.10 (–0.12 to 0.05)		*<.001*
Socioeconomic vulnerability		1,398,824	N/A	N/A	2.61 (1.30 to 5.16)	*<.001*
RGS	6.45 (1.11)					

Variables	Mean (SD)	M-W[a] test	K-W[b] test	Spearman *r* (95% CI)	OR[c] (95% CI)	*P* value
Not RGS	6.61 (1.05)					
People in residence (n)		N/A	N/A	–0.01 (–0.04 to 0.02)	N/A	.66
1-40	3.35 (1.61)					
HDI		N/A	N/A	0.07 (0.04 to 0.10)	N/A	*<.001*
0.683-0.850	0.779 (0.04)					
Conduct in suspected infection						
Sex		1,809,913	N/A	N/A	0.95 (0.80 to 1.12)	.53
Male	0.80 (0.40)					
Female	0.80 (0.40)					
Education level		2,149,210	N/A	N/A	0.91 (0.78 to 1.06)	.21
Middle and high school	0.81 (0.39)					
Higher/postgraduate	0.79 (0.41)					
Age (years)		N/A	295.40		N/A	*<.001*
18-19	0.87 (0.33)[f]			N/A		N/A
20-29	0.90 (0.30)[f]			N/A		N/A
30-39	0.78 (0.42)[g]			N/A		N/A
40-49	0.70 (0.46)[h]			N/A		N/A
50-59	0.68 (0.47)[h]			N/A		N/A
≥60	0.55 (0.50)[k]			N/A		N/A
18-83	34.57 (14.01)			–0.25 (–0.28 to –0.22)		*<.001*
Socioeconomic vulnerability		1,473,256	N/A	N/A	1.15 (0.97 to 1.38)	.11
RGS	0.78 (0.41)					
Not RGS	0.80 (0.40)					
People in residence (n)		N/A	N/A	0.01 (–0.02 to 0.04)	N/A	.73
1-40	3.35 (1.61)					
HDI		N/A	N/A	0.12 (0.09 to 0.15)	N/A	*<.001*
0.683-0.850	0.779 (0.04)					
Risk groups						
Sex		1,641,327	N/A	N/A	2.37 (0.89 to 6.32)	*<.001*
Male	8.31 (1.29)					
Female	8.53 (1.27)					
Education level		2,010,364	N/A	N/A	0.71 (0.25 to 2.06)	*<.001*
Middle and high school	8.37 (1.31)					
Higher/postgraduate	8.56 (1.24)					
Age (years)		N/A	33.51		N/A	*<.001*
18-19	8.14 (1.38)[f,g]			N/A		N/A
20-29	8.52 (1.20)[f,g,h]			N/A		N/A
30-39	8.56 (1.29)[f,h]			N/A		N/A
40-49	8.52 (1.29)[f,g,h]			N/A		N/A
50-59	8.37 (1.38)[g,h]			N/A		N/A

Variables	Mean (SD)	M-W[a] test	K-W[b] test	Spearman *r* (95% CI)	OR[c] (95% CI)	*P* value
≥60	8.29 (1.39)[g]			N/A		N/A
18-83	34.57 (14.01)			0.02 (–0.01 to 0.05)		.13
Socioeconomic vulnerability		1,449,382	N/A	N/A	0.58 (0.13 to 2.26)	.06
RGS	8.39 (1.31)					
Not RGS	8.49 (1.27)					
People in residence (n)		N/A	N/A	–0.02 (–0.06 to 0.01)	N/A	.10
1-40	3.35 (1.61)					
HDI		N/A	N/A	0.01 (–0.02 to 0.04)	N/A	.50
0.683-0.850	0.779 (0.04)					
Social distancing						
Sex		1,780,394	N/A	N/A	3.21 (2.06 to 5.00)	*<.001*
Male	0.96 (0.19)					
Female	0.99 (0.11)					
Education level		2,181,870	N/A	N/A	0.98 (0.63 to 1.53)	.92
Middle and high school	0.98 (0.13)					
Higher/postgraduate	0.98 (0.14)					
Age (years)		N/A	18.93		N/A	*.002*
18-19	0.97 (0.16)[f,h]			N/A		N/A
20-29	0.99 (0.11)[f]			N/A		N/A
30-39	0.97 (0.16)[f,h]			N/A		N/A
40-49	0.98 (0.14)[f,h]			N/A		N/A
50-59	0.99 (0.12)[f]			N/A		N/A
≥60	0.95 (0.22)[h]			N/A		N/A
18-83	34.57 (14.01)			–0.03 (–0.06 to 0.00)		*.03*
Socioeconomic vulnerability		1,507,802	N/A	N/A	0.96 (0.55 to 1.63)	.88
RGS	0.98 (0.13)					
Not RGS	0.98 (0.14)					
People in residence (n)		N/A	N/A	–0.02 (–0.05 to 0.01)	N/A	.16
1-40	3.35 (1.61)					
HDI		N/A	N/A	–0.01 (–0.04 to 0.02)	N/A	.42
0.683-0.850	0.779 (0.04)					
Prevention						
Sex		1,803,511	N/A	N/A	7.1 (1.06 to 92.31)	.50
Male	10.21 (1.03)					
Female	10.25 (0.94)					
Education level		2,010,310	N/A	N/A	0.95 (0.15 to 6.09)	*<.001*
Middle and high school	10.16 (1.01)					
Higher/postgraduate	10.32 (0.91)					
Age (years)		N/A	37.61		N/A	*<.001*
18-19	10.05 (1.01)			N/A		N/A
20-29	10.35 (0.85)			N/A		N/A

Variables	Mean (SD)	M-W[a] test	K-W[b] test	Spearman *r* (95% CI)	OR[c] (95% CI)	*P* value
30-39	10.17 (1.03)			N/A		N/A
40-49	10.25 (0.90)			N/A		N/A
50-59	10.15 (1.06)			N/A		N/A
≥60	10.07 (1.29)			N/A		N/A
18-83	34.57 (14.01)			–0.03 (–0.06 to 0.00)		.07
Socioeconomic vulnerability		1,407,623	N/A	N/A	0.00 (0 to 3.51)	*<.001*
RGS	10.17 (0.95)					
Not RGS	10.26 (0.97)					
People in residence (n)		N/A	N/A	0.01 (–0.02 to 0.04)	N/A	.54
1-40	3.35 (1.61)					
HDI		N/A	N/A	0.10 (0.03 to 0.09)	N/A	*<.001*
0.683-0.850	0.779 (0.04)					
Grouped variables						
Sex		1,680,078	N/A	N/A	1.25 (1.01 to 1.56)	*<.001*
Male	33.55 (2.45)					
Female	33.86 (2.42)					
Education level		1,981,133	N/A	N/A	1.69 (1.37 to 2.08)	*<.001*
Middle and high school	33.56 (2.55)					
Higher/postgraduate	33.99 (2.28)					
Age (years)		N/A	110.10		N/A	*<.001*
18-19	33.27 (2.41)[f,k]			N/A		N/A
20-29	34.21 (2.11)[h]			N/A		N/A
30-39	33.69 (2.57)[g]			N/A		N/A
40-49	33.63 (2.57)[f,g]			N/A		N/A
50-59	33.32 (2.58)[f,k]			N/A		N/A
≥60	32.92 (2.84)[k]			N/A		N/A
18-83	34.57 (14.01)			–0.10 (–0.12 to –0.05)		*<.001*
Socioeconomic vulnerability		1,334,021	N/A	N/A	1.50 (1.19 to 1.87)	*<.001*
RGS	33.39 (2.53)					
Not RGS	33.88 (2.39)					
People in residence (n)		N/A	N/A	–0.02 (–0.05 to 0.01)	N/A	.26
1-40	3.35 (1.61)					
HDI		N/A	N/A	0.10 (0.07 to 0.13)	N/A	*<.001*

Variables	Mean (SD)	M-W[a] test	K-W[b] test	Spearman r (95% CI)	OR[c] (95% CI)	P value
0.683-0.850	0.779 (0.04)					

[a]M-W: Mann-Whitney.

[b]K-W: Kruskal-Wallis.

[c]OR: odds ratio.

[d]N/A: not applicable.

[e]Italics indicate statistically significant difference.

[f]Statistically significant difference between these groups.

[g]Statistically significant difference between these groups.

[h]Statistically significant difference between these groups.

[i]RGS: receiving government support.

[j]HDI: Human Development Index.

[k]Statistically significant difference between these groups.

With regard to sex, women had a better understanding and greater knowledge about transmission (0.60% more; P=.01) and risk groups (2.65% more; P<.001), with no difference in odds ratios. Regarding the understanding of the importance of social distancing, in addition to a better average of female performance (3.13% more; P<.001), there was also a higher probability of low performance for men (OR 3.21, 95% CI 2.06-5.00). A similar result was observed when the average score of all parameters was grouped, with a better average accuracy by women (P<.001) and a greater probability of incorrect responses by men (OR 1.69, 95% CI 1.01-1.56; Table 3).

When the population's perception about COVID-19 was assessed taking into account the education level, no significant differences were found between the groups concerning knowledge about symptoms, actions in case of suspected SARS-CoV-2 infection, and social distancing (all P>.05). On the other hand, the higher education and postgraduate group obtained a better average for knowledge about transmission (1.50% more; P<.001), risk groups (2.27% more; P<.001), and prevention (1.57% more; P<.001). When the average scores of all parameters were grouped, the higher/postgraduate group had a better average performance (P<.001), and the middle and high school group had a higher probability of lower performance (OR 1.69, 95% CI 1.37-2.08; Table 3).

The age of the participants influenced the understanding and comprehension of all COVID-19 health education indicators (all P<.05). The younger participants had a better understanding. In addition, a negative and significant correlation (all P<.05) was observed in relation to knowledge about transmission, symptoms, conduct in suspected infection, risk groups, and social distancing. On the other hand, concerning the indicator of prevention, no significant correlation was observed (all P>.05). After grouping the average scores for all parameters, the negative and significant correlation between age and percentage of correct answers was maintained (P<.001), with the greatest average difference observed in the group 60 years and older, and in the group aged 20-29 years (percentage of correct answers 3.92% higher; P<.001; Table 3).

The influence of socioeconomic vulnerability on the population's perception of COVID-19 was estimated through the investigation of participants who received or did not receive any type of financial support or resource from the government during the pandemic. Regardless of whether or not they received any financial support, no statistically significant differences were found with regard to the understanding of conduct in case of suspected SARS-CoV-2 infection (P=.11), risk groups (P=.06), and the importance of social distancing (P=.88). On the other hand, a higher percentage of correct answers was observed in the population with better socioeconomic conditions for knowledge about transmission (percentage of correct answers 1.66% higher; P<.001), symptoms (percentage of correct answers 2.48% higher; P<.001), prevention (percentage of correct answers 0.88% higher; P<.001), or even when the average scores of all parameters were grouped (percentage of correct answers 1.47% higher; P<.001). In addition, a greater probability for a lower percentage of correct answers was observed among people with socioeconomic vulnerability regarding symptoms (OR 2.61, 95% CI 1.30-5.16) and when the average scores of all parameters were grouped (OR 1.50, 95% CI 1.19-1.87; Table 3).

The relationship between the number of people per household and correct answers was also assessed; however, no statistically significant association was found for any of the examined questions (all P>.05). Correlations were also investigated regarding the HDI and the percentage of correct answers. Positive and significant correlations (all P<.001) were found in all evaluated variables, except for knowledge about risk groups (P=.50; Table 3).

Taking into account the vast territorial extent of Brazil and its cultural, climatic, and political influences, among others, we also evaluated the possible differences in correct answers of questions related to health education on COVID-19 by region (Table 4). Regardless of region, a significant percentage of the respondents answered questions incorrectly about COVID-19. In the specific cases of knowledge about the main risk groups and social distancing, the percentage of incorrect responses was similar in all regions (average incorrect responses 15.62% and 2.40%, respectively). Knowledge of the population in each region, on the other hand, about transmission, symptoms, conduct in cases of suspected infection, prevention, and the average score of all variables together showed significant differences between regions (Table 4).

Table 4. Sample statistical analysis based in geopolitical regions.

Variables	North	Northeast	Central-West	Southeast	South	*P* value (K-W[a] test)
Transmission						*<.001* [b]
Mean (SD)	6.57 (0.74)[c]	6.62 (0.71)[d]	6.68 (0.72)[e]	6.79 (0.53)[f]	6.64 (0.70)[g]	
CV[h] (%)	11.34	10.72	10.77	7.88	10.52	
Reduction (%)	–6.14	–5.43	–4.57	–3.00	–5.14	
Symptoms						*<.001*
Mean (SD)	6.38 (1.12)[c]	6.53 (1.06)[c,e]	6.45 (1.10)[c,e]	6.64 (1.03)[d,e]	6.60 (1.13)[e]	
CV (%)	17.57	16.24	17.00	15.55	17.13	
Reduction (%)	–29.11	–27.44	–28.33	–26.22	–26.67	
Conduct in cases of suspected infection						*<.001*
Mean (SD)	0.67 (0.47)[c]	0.74 (0.44)[d]	0.72 (0.45)[c,d]	0.86 (0.39)[e]	0.77 (0.42)[d]	
CV (%)	70.84	58.70	61.76	40.51	54.40	
Reduction (%)	–33.00	–26.00	–28.00	–14.00	–23.00	
Risk groups						.52
Mean (SD)	8.32 (1.40)	8.42 (1.35)	8.50 (1.20)	8.50 (1.21)	8.45 (1.36)	
CV (%)	16.79	16.06	14.11	14.23	16.15	
Reduction (%)	–16.80	–15.80	–15.00	–15.00	–15.50	
Social distancing						.28
Mean (SD)	0.98 (0.14)	0.98 (0.13)	0.96 (0.19)	0.98 (0.13)	0.98 (0.13)	
CV (%)	14.69	13.25	19.42	13.18	13.80	
Reduction (%)	–2.00	–2.00	–4.00	–2.00	–2.00	
Prevention						*<.001*
Mean (SD)	10.03 (1.12)[c]	10.19 (0.98)[c]	10.06 (1.08)[c]	10.33 (0.89)[d]	10.19 (1.02)[c]	
CV (%)	11.19	9.60	10.72	8.65	9.99	
Reduction (%)	–8.82	–7.36	–8.55	–6.09	–7.36	
Grouped variables						*<.001*
Mean (SD)	32.94 (2.78)[c]	33.48 (2.56)[d,e]	33.37 (2.42)[c,e]	34.10 (2.20)[f]	33.64 (2.62)[d]	
CV (%)	8.44	7.64	7.27	6.44	7.79	
Reduction (%)	–15.54	–14.15	–14.44	–12.56	–13.74	

[a]K-W: Kruskal-Wallis.

[b]Italics indicate statistically significant difference.

[c]Statistically significant difference between these groups.

[d]Statistically significant difference between these groups.

[e]Statistically significant difference between these groups.

[f]Statistically significant difference between these groups.

[g]Statistically significant difference between these groups.

[h]CV: coefficient of variation.

Respondents from the North region of Brazil had the highest percentage of wrong answers on transmission (6.14%), followed by those from the Northeast (5.43%), South (5.14%), Central-West (4.57%), and Southeast (3%). Respondents from the North region also presented a higher percentage of wrong answers on questions about the symptoms (29.11%), conduct to be taken in case of suspected infection (33%), prevention (8.82%), and when all the variables were grouped (15.54%). The Northeast was the second region in the number of wrong answers about COVID-19, with percentage of incorrect responses of 27.44%, 26%, 7.36%, and 14.15% for symptoms, conduct in cases of suspected infection, prevention, and when all variables were grouped, respectively. The Central-West was third in percentage of wrong answers with 28.33%, 28%, 8.55%, and 14.44% for symptoms, conduct in case of suspected infection, prevention, and when all variables were grouped, respectively. Meanwhile, the South had percentage of incorrect responses of 26.67%, 23%, 7.36%, and 13.74% for symptoms, conduct in case of suspected infection, prevention, and grouping of all variables, respectively. On the other hand, the Southeast had the lowest percentage incorrect responses, with percentages of 26.22%, 14.00%, 6.09%, and 12.56% for symptoms, conduct in case of suspected infection, prevention, and grouping the average score of all parameters, respectively (Table 4).

Discussion

Sample Data

We assessed the Brazilian population's basic knowledge about COVID-19. To this end, an online survey was made available that allowed people 18 years or older and who use social media for communication and information to test their knowledge about the disease and infection. The findings were elucidating, as this method of gathering information allowed for the evaluation of people from different regions of Brazil, different social groups, of different ages, and with different education levels (ranging from people with only basic education to people with graduate degrees). The survey respondents were predominantly female, young (younger than 40 years), from the Southeast region, and composed of people who did not receive government assistance. Women's greater concern with health [22] and the massive use of social media by young individuals with better social and economic conditions are aspects that may explain the predominant final configuration of the findings [23]. In general, the respondents presented satisfactory basic knowledge about COVID-19, scoring an average of 86.59% of the maximum possible survey score but with statistically significant differences depending on the question, group, or region analyzed. That is, there were differences in the groups analyzed that showed that knowledge about the disease, although reasonable, differed depending on the respondent. Similar studies in other countries have been carried out and have presented participants with characteristics close to those obtained in our study, that is, satisfactory basic knowledge of the disease but with inequalities depending on the analyzed group [7,11,15,24]. This was the case in a study carried out in China [7], in which participants scored an average of 90% of the total possible score with a predominantly young female sample and with just over half of the interviewees having completed undergraduate and graduate courses.

With respect to social media disseminating metrics, a total of R $803.76 (US $155.77) were invested so that 5908 clicks on the form link could be reached. This means, on average, that R $0.14 (US $0.03) per click was spent. A total of 239,414 people were reached, generating 349,320 impressions. That is, 2.47% of the people who were reached with the dissemination accessed the link. These metrics show that despite a relatively low investment per click, searches for paid ads like this arouse the interest of a minority of the people reached. Considering that the research producers also collaborated on disseminating the form on social media, we observed that it was filled out by 4180 participants, meaning that not all of those who clicked on the link followed the form until its completion. This low adherence was expected due to, among other reasons, the fact that the metrics are proportional to the perception of the institution's credibility [25]. Thus, as science in Brazil still has a lack of credibility among the population, especially due to problems of communication between science and the practical appropriation of the scientific knowledge, it was already expected that there would be a minority of responses in relation to the total number of individuals impacted by the dissemination [26].

Main Findings

As illustrated in Figure 3, there was a statistical difference among the number of correct answers according to the investigated question. More correct answers were recorded on questions related to the importance of social distancing, treatment, and prevention. We observed that, similarly, these same items were answered correctly in the United States, China, Pakistan, the United Kingdom, and India [6,7,11,24], showing that these are issues with global levels of comprehension. The lowest scores occurred on questions concerned with knowledge about the symptoms of the disease, risk groups, and conduct for patients suspected of infection. Participants overestimated shortness of breath, did not recognize some risk groups such as pregnant women, and demonstrated more misconceptions regarding the conduct in suspected cases. This difference, depending on the questioned item, was also found by other studies in other countries. However, the wrong questions vary; in Brazil, as described, the question with the fewest correct answers was that related to symptoms, while in a survey carried out in Pakistan [24], the question related to transmission received the fewest correct answers. This indicates that these different scores may depend on the specific aspect about COVID-19 being investigated, and these aspects may vary according to the country studied.

In our research, as shown in Table 3, the three symptoms that the respondents most believed to be connected with COVID-19 were fever, dry cough, and shortness of breath. The first two symptoms are correct; however, the third is not: shortness of breath has occurred in a minority of COVID-19 cases, and the correct response alternative would be the symptom of fatigue [27]. Although symptoms such as diarrhea, skin wounds, vomiting, stuffy nose, shortness of breath, headache, and loss of taste or smell may be present, they are less frequent [27]. Because the question was limited to the three most common symptoms (persistent tiredness, fever, and cough), the alternatives that encompassed these less common symptoms were false. It is important to note that when checking the answers participants were informed that, although these other symptoms were not the most frequent, they could show up. Research conducted in India, the United Kingdom, and the United States [10,13] showed a similar situation (ie, the replacement of fatigue by shortness of breath as one of the three main symptoms of the disease). The media dissemination of more serious cases and of deaths related to shortness of breath, hospitalization, and the use of pulmonary ventilators may have led the general population to believe that dyspnea is a common manifestation of the disease. It is important to note that this perception can negatively influence people's behavioral conduct and create a dualistic view of COVID-19 (ie, it can lead individuals to believe that they are infected only when shortness of breath is present; otherwise, they are healthy, which is not entirely true). In other words, this perception can be worrying in relation to milder and asymptomatic cases, which normally do not exhibit shortness of breath. In the absence of the manifestation of this symptom, these groups may not conduct themselves appropriately because they believe they are not infected, thereby becoming potential transmitters of the disease.

The second question that received the most incorrect answers was related to a patient's conduct in case of suspected SARS-COV-2 infection. In Brazil, the Ministry of Health has relayed that in case of suspected COVID-19, individuals need to stay at home in isolation and only seek health services in certain situations such as when symptoms are more severe. This course of action is in line with the recommendations of the World Health Organization. However, at the beginning of the pandemic, the increased concern of the population caused hospitals to be filled with people with mild symptoms who sought medical care even though their chances of complications were low, for they neither were from risk groups nor had worrying symptoms. This overload ended up bringing crowds to health services, including people without COVID-19 who were concerned with any sign common to the disease, contributing to the spread of the virus in the population. Surveys conducted in China and India [7,10], places where the same course of action has also been endorsed by their governments, have had participants show high levels of correctness in answers. This means that, regarding one's conduct in case of suspected COVID-19 infection, efforts to disseminate the relevant correct information have resulted in a relatively adequate awareness among the population; however, in Brazil, it seems that it is necessary to reinforce this awareness.

The third question that received the most incorrect answers was related to identifying risk groups among the general population who were more likely to deal with the most severe forms of the disease. Some risk groups were adequately recognized by the 4180 participants; however, other risk groups were less recognized, as was the case for people with heart or kidney problems (n=2806, 67.13%), people with cancer (n=2668, 63.83%), and pregnant women (n=1531, 36.63%). At the beginning of the pandemic, there was still no certainty about the inclusion of pregnant women in the risk group. Months later, the Brazilian Ministry of Health officially included this part of the population in this group. It is possible that this initial confusion may have influenced the correct recognition of risk groups. This factor may have a negative impact by decreasing the precaution around these groups, which require more attention. In China and Pakistan [7,24], older adults, people who are obese, and patients with chronic diseases have also been recognized by several studies as being the most common risk groups. In the United States and the United Kingdom [13], older adults were also recognized as a risk group, followed by adults with health problems. However, 53.8% of Americans and 39.1% of British people also recognized children as a risk group, which, from a scientific point of view, is incorrect [28]. People therefore more easily recognize important risk groups such as older adults and people with chronic diseases. However, others are being forgotten, such as pregnant women in Brazil, or are being incorrectly assigned, such as children in the United States and the United Kingdom. With regard to the questions that were answered correctly the most in our research (ie, those regarding prevention, transmission, and social distancing), we observed similar items in surveys conducted in China, India, Pakistan, the United Kingdom, and the United States, and the findings of these surveys showed that their participants' comprehension levels had already reached that of many nations in the world [7,10,13,24].

Regarding sex, women obtained more correct answers than men in the questions on general knowledge about COVID-19—a finding that conflicts with the literature: in the United States and the United Kingdom [14], no difference in the correctness of survey answers was observed between men and women. In Pakistan [24], men were more correct than women; in China [7], as well as in Brazil, women were more correct than men. The lack of similarity in findings among countries and the differences between the sexes seem to indicate that our findings with respect to sex may be underdetermined by other characteristics including economic, social, and cultural development. We also separated the study participants as undergraduates, graduates, and those who were neither. In this sense, participants who had higher educational qualifications were observed to be better informed about the disease and infection. In China and Pakistan [7,24], similar findings were found, indicating that a higher level of education indicates a higher level of knowledge about COVID-19. These findings underscore that investments in education (in addition to contributing to scientific development) create a more informed population, as is the case with regard to COVID-19 and, moreover, any disease that may affect the general population, whether as a pandemic or not.

Differences were identified when the participants were grouped by age, with the younger participants tending to have more correct answers than those who were older, especially when comparing older adults with individuals aged 20-29 years. In Pakistan, this correlation was also found [24], but in China [7], this was not the case; older adults were the second most successful group. Although our data do not correspond worldwide, it is possible that in some countries younger individuals, who tend to have had earlier experience with the internet and social networks, have a more critical perception with the information conveyed on social networks.

Concerning the aspects of socioeconomic vulnerability, HDI, and regional differences, no research was found in the literature that has investigated these aspects. Thus, we consider our study findings to be unique and important in more accurately understanding people's knowledge of the pandemic. The information we have obtained evidences the reasoning that more socioeconomically advanced groups have greater knowledge about COVID-19. That is, a correlation was found between the lack of government assistance and the highest HDI in the region with the largest number of correct answers regarding COVID-19. In practical terms, development and income may be predictors for better levels of access to and interpretation of information, including those related to COVID-19, thus enabling a better understanding of the disease.

The different regions of the country had different levels of success, which reflects the important social and economic differences between the regions [29]. This highlights the need to develop specific public policies for each location, with greater emphasis on conduct awareness in case of suspected infection in the North region and identification of the main symptoms in other regions of the country.

Practical Applications

It is important to consider that one of the key points that began this research was the significant spread of fake news, conspiracy theories, and contradictory orientations among the population. Fake news is not a phenomenon exclusive to the COVID-19 pandemic; it has been verified in other contexts, especially in political elections [30]. However, in the pandemic, their impact can be dramatic; when spread in a sustained manner, they have a disruptive effect on the preventive measures necessary to combat COVID-19. With less prevention, more people become contaminated, greater overload occurs in the health system, and consequently, more deaths are accounted. This is especially problematic in Brazil, a transitional country whose public health system has a burden of diseases and lack of resources [31].

In Brazil, studies have indicated that 9 out of 10 people have read or heard at least one source of false information about COVID-19, while 7 out of 10 believe in at least one uninformative source about the disease [32]. This significant proportion of misinformation is not a mere disinterested product without scientific knowledge. Fake news has several purposes in validating points of view that are incompatible with science but serve political, economic, and even criminal interests. As fake news spreads six times faster than true information, producers can create this content to generate network traffic for financial return with advertising, or there may even be scams asking for money for respected scientific institutions to fight COVID-19 [32,33].

However, when analyzing the study population, it was possible to verify that there is a satisfactory knowledge about COVID-19 when true information and fake news are mixed. Participants demonstrated that they were able to differentiate the two types of information. Thus, although more studies are needed, it is possible to suggest that the impact of fake news on the knowledge of COVID-19 in the population of our study was limited. This does not mean that fake news has a limited impact on the Brazilian population in general, as this study did not fully analyze it nor did it select all the fake news that exists among the population; only a few of the main ones selected by the Brazilian Ministry of Health were used, and their verification had already been made public in advance. This result shows a certain effectiveness in campaigns against the Brazilian government's lack of information at the beginning of the pandemic and underscores the importance of continuing this action. In spite of this, it is necessary to consider that the knowledge assessed was considered basic, that is, excessively technical aspects of a complex disease such as COVID-19 were not addressed. The alternatives were based on practical aspects disclosed by the Ministry of Health that the population could use in their daily life.

Nevertheless, even with the good theoretical knowledge demonstrated by the population of this study, the practice is still not represented in the population's behavior. In Brazil, the practice of social distancing is unsatisfactory; agglomeration cases are recurrent; and, although efficient, preventive measures still do not show significant adherence by the population [34]. Thus, there is a gap between theorical knowledge and satisfactory practice. In a way, this shows that the problem of a lack of adherence to preventive measures cannot be attributed exclusively to fake news. In other words, the lack of knowledge is not the only factor that impacts the generation of an effective practice against the pandemic in Brazilians who use social networks. To consolidate the practice of fighting COVID-19, in addition to producing knowledge, it is necessary to provide more conditions for its practical implementation. The need to investigate and to correct other social, political, economic, and cultural conditions that are preventing a disciplined coping with the pandemic is evident, not exclusively attributing the responsibility for low public engagement to the fake news.

In addition, our findings are useful to political authorities, journalistic or media groups, and even to social media. This is because the findings unfolded here diagnose some weak spots in the population's knowledge about COVID-19. Despite how satisfactory the general knowledge of the disease may be, failures were observed in certain groups such as men, older adults, and undereducated people; in locations such as those with the lowest HDI; and in aspects of the disease, such as the most common symptoms, conduct in suspected cases, and identification of risk groups. Some of these findings have even been confirmed in studies from other countries, showing a similarity that goes beyond continents. Ultimately, public and private institutions responsible for informing the population need to focus their efforts on these shortcomings.

We also demonstrated that investments in education and socioeconomic improvements can have a positive impact on the knowledge and actions of the population, which can be useful not only in coping with COVID-19 but also in other diseases or possible future pandemics. These two pillars, in addition to allowing investigations that improve the effort to fight and treat the disease, are themselves capable of educating citizens more immune to fake news.

Limitations

Online surveys have some limitations. Participants, for example, could search for answers on the internet or choose random alternatives to quickly complete the questionnaire that would impair some of the data.

As this study had no deadline and was voluntary and anonymous, participants were free in their decision to engage. They were also warned that their knowledge would not be exposed, leaving them more comfortable to answer the questions, avoiding any related bias. Despite this and taking into account the state of the pandemic and social isolation in Brazil, dissemination through social media through a form by Google Forms proved to be a viable solution for assessing knowledge about COVID-19.

The findings of this study only apply to people who use social platforms, can read and write, present some level of knowledge about the pandemic, and have compatible electronic equipment to answer the survey. In other words, Brazil is a country that still has a high rate of illiteracy [35], and a large portion of the population does not have the internet and equipment necessary to access the online survey. Thus, although the survey findings represent an important portion of the population, it cannot be generalized as being applicable to the entire Brazilian population. In addition, the survey was optional, which may indicate that a large part of the responses came from participants with a greater interest in information concerning the disease. This could have an influence in the participants' good performance. To reduce this limitation, we sought to evaluate a high number of participants, which eventually brought greater representativeness to the sample.

Finally, for this publication, the questionnaire data were translated from Portuguese (the official language of Brazil) into English. Some translation problems could change certain interpretations of sentences. To avoid that, we submitted the revised version of the manuscript to a professional academic English editing service.

Conclusions

The Brazilian population with access to social networks demonstrated satisfactory basic knowledge about COVID-19. Despite this, there were differences among the issues, groups analyzed, and regions of the country. In general, participants had better knowledge about prevention, transmission, and social distancing but made more mistakes in identifying the main symptoms, risk groups, and correct conduct in cases of infection. Better performances were also observed among women, young people between 20 and 29 years of age, undergraduates and graduates, and those who did not receive any type of government assistance. In addition, a positive correlation was identified between the best HDI and the level of knowledge about the disease.

Acknowledgments

The authors would like to thank all the participants who contributed to the study and permitted an investigation representative of a large part of the Brazilian population. This study was supported by the Fundação de Amparo à Pesquisa do Estado Minas Gerais, National Council for Scientific and Technological Development, and Coordination for the Improvement of Higher Education Personnel (finance code 001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The complete version of the form given to participants.
[DOCX File , 47 KB - publichealth_v7i1e24756_app1.docx]

References

1. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
2. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436. [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
3. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-09-13]
4. Fadaka AO, Sibuyi NRS, Adewale OB, Bakare OO, Akanbi MO, Klein A, et al. Understanding the epidemiology, pathophysiology, diagnosis and management of SARS-CoV-2. J Int Med Res 2020 Aug;48(8):300060520949077 [FREE Full text] [doi: 10.1177/0300060520949077] [Medline: 32842818]
5. Coronavirus disease (COVID-19) advice for the public: Mythbusters. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-09-24]
6. Werneck G, Carvalho M. The COVID-19 pandemic in Brazil: chronicle of a health crisis foretold. Cad Saude Publica 2020;36(5):e00068820 [FREE Full text] [doi: 10.1590/0102-311x00068820] [Medline: 32402007]
7. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
8. Pedrozo-Pupo JC, Pedrozo-Cortés MJ, Campo-Arias A. Perceived stress associated with COVID-19 epidemic in Colombia: an online survey. Cad Saude Publica 2020;36(5):e00090520 [FREE Full text] [doi: 10.1590/0102-311x00090520] [Medline: 32490918]
9. Chan EYY, Huang Z, Lo ESK, Hung KKC, Wong ELY, Wong SYS. Sociodemographic predictors of health risk perception, attitude and behavior practices associated with health-emergency disaster risk management for biological hazards: the case

of COVID-19 pandemic in Hong Kong, SAR China. Int J Environ Res Public Health 2020 May 29;17(11) [FREE Full text] [doi: 10.3390/ijerph17113869] [Medline: 32485979]
10. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Asian J Psychiatr 2020 Jun;51:102083 [FREE Full text] [doi: 10.1016/j.ajp.2020.102083] [Medline: 32283510]
11. Taghrir MH, Borazjani R, Shiraly R. COVID-19 and Iranian medical students; a survey on their related-knowledge, preventive behaviors and risk perception. Arch Iran Med 2020 Apr 01;23(4):249-254. [doi: 10.34172/aim.2020.06] [Medline: 32271598]
12. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of public perception of the Israeli government's early emergency instructions regarding COVID-19: online survey study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
13. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
14. McFadden SM, Malik AA, Aguolu OG, Willebrand KS, Omer SB. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
15. Jones J, Sullivan PS, Sanchez TH, Guest JL, Hall EW, Luisi N, et al. Similarities and differences in COVID-19 awareness, concern, and symptoms by race and ethnicity in the United States: cross-sectional survey. J Med Internet Res 2020 Jul 10;22(7):e20001 [FREE Full text] [doi: 10.2196/20001] [Medline: 32614778]
16. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-02-10]
17. Coronavirus panel. Coronavírus Brazil. URL: https://covid.saude.gov.br/ [accessed 2020-11-23]
18. Coronavírus (Covid-19). Ministério da Saúde. 2020. URL: https://coronavirus.saude.gov.br/ [accessed 2020-12-03]
19. Health without fake news. Ministério da Saúde. URL: https://www.saude.gov.br/fakenews [accessed 2020-05-09]
20. Fiocruz. Covid-19 - new coronavirus. Fiocruz. 2020. URL: https://portal.fiocruz.br/coronavirus [accessed 2020-12-03]
21. Arango HG. Bioestatística: Teórica e Computacional. Rio de Janeiro: Guanabara Koogam; 2001.
22. PNS 2019: sete em cada dez pessoas que procuram o mesmo serviço de saúde vão à rede pública. Agencia de Noticias. 2020. URL: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28793-pns-2019-sete-em-cada-dez-pessoas-que-procuram-o-mesmo-servico-de-saude-vao-a-rede-publica [accessed 2020-11-28]
23. PNAD Contínua TIC 2017: Internet chega a três em cada quatro domicílios do país. Agencia de Noticias. 2018. URL: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/23445-pnad-continua-tic-2017-internet-chega-a-tres-em-cada-quatro-domicilios-do-pais [accessed 2020-11-28]
24. Hayat K, Rosenthal M, Xu S, Arshed M, Li P, Zhai P, et al. View of Pakistani residents toward coronavirus disease (COVID-19) during a rapid outbreak: a rapid online survey. Int J Environ Res Public Health 2020 May 12;17(10) [FREE Full text] [doi: 10.3390/ijerph17103347] [Medline: 32408528]
25. Patino-Hernandez D, Fernández-Ávila DG, Celis-Preciado CA, Munoz-Velandia OM. Social networks and traditional metrics of impact in pulmonary medicine journals: a correlation study. Adv Respir Med 2019;87(6):209-213 [FREE Full text] [doi: 10.5603/ARM.2019.0058] [Medline: 31970722]
26. Massarani L, Moreira IDC. Science communication in Brazil: a historical review and considerations about the current situation. An Acad Bras Cienc 2016 Sep;88(3):1577-1595 [FREE Full text] [doi: 10.1590/0001-3765201620150338] [Medline: 27556221]
27. Folha informativa COVID-19 - Escritório da OPAS e da OMS no Brasil. Pan American Health Organization. 2020. URL: https://www.paho.org/pt/covid19 [accessed 2020-11-26]
28. Johansen TB, Astrup E, Jore S, Nilssen H, Dahlberg BB, Klingenberg C, et al. Infection prevention guidelines and considerations for paediatric risk groups when reopening primary schools during COVID-19 pandemic, Norway, April 2020. Euro Surveill 2020 Jun;25(22) [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.22.2000921] [Medline: 32524956]
29. Marson F, Ortega M. COVID-19 in Brazil. Pulmonology 2020;26(4):241-244 [FREE Full text] [doi: 10.1016/j.pulmoe.2020.04.008] [Medline: 32371054]
30. Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 2019 Jan 02;10(1):7. [doi: 10.1038/s41467-018-07761-2] [Medline: 30602729]
31. Paim J, Travassos C, Almeida C, Bahia L, Macinko J. The Brazilian health system: history, advances, and challenges. Lancet 2011 May 21;377(9779):1778-1797. [doi: 10.1016/S0140-6736(11)60054-8] [Medline: 21561655]
32. Galhardi C, Freire N, Minayo M, Fagundes M. Fact or fake? An analysis of disinformation regarding the Covid-19 pandemic in Brazil. Cien Saude Colet 2020 Oct;25(suppl 2):4201-4210 [FREE Full text] [doi: 10.1590/1413-812320202510.2.28922020] [Medline: 33027357]
33. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]

34. Cai para 33,8 milhões número de pessoas rigorosamente isoladas na pandemia. Agencia de Noticias. 2020. URL: https://tinyurl.com/y2qof6l6 [accessed 2020-11-28]
35. Mapa Do Analfabetismo Do Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Aníso Teixeira. 2016. URL: http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/6978610 [accessed 2020-02-10]

Abbreviations

HDI: Human Development Index

Edited by G Eysenbach; submitted 04.10.20; peer-reviewed by G Kernohan, B Chakalov, M Antunes, A Bhatia, FS Lodhi; comments to author 18.11.20; revised version received 11.12.20; accepted 24.12.20; published 21.01.21.

Please cite as:
Guimarães VHA, de Oliveira-Leandro M, Cassiano C, Marques ALP, Motta C, Freitas-Silva AL, de Sousa MAD, Silveira LAM, Pardi TC, Gazotto FC, Silva MV, Rodrigues Jr V, Rodrigues WF, Oliveira CJF
Knowledge About COVID-19 in Brazil: Cross-Sectional Web-Based Study
JMIR Public Health Surveill 2021;7(1):e24756
URL: http://publichealth.jmir.org/2021/1/e24756/
doi:10.2196/24756
PMID:33400684

Original Paper

COVID-19–Related Hospitalization Rates and Severe Outcomes Among Veterans From 5 Veterans Affairs Medical Centers: Hospital-Based Surveillance Study

Cristina V Cardemil[1,2], MD, MPH; Rebecca Dahl[1], MPH; Mila M Prill[1], MSPH; Jordan Cates[1], PhD; Sheldon Brown[3,4], MD; Adrienne Perea[3], BS; Vincent Marconi[5,6,7], MD; LaSara Bell[5], MS; Maria C Rodriguez-Barradas[8,9], MD; Gilberto Rivera-Dominguez[8,9], MD; David Beenhouwer[10,11], MD; Aleksandra Poteshkina[10], BS; Mark Holodniy[12,13,14], MD; Cynthia Lucero-Obusan[12,13], MD; Neha Balachandran[1], MBBS, MPH; Aron J Hall[1], DVM, MSPH; Lindsay Kim[1,2], MD, MPH; Gayle Langley[1], MD, MPH

[1]Centers for Disease Control and Prevention, Atlanta, GA, United States
[2]United States Public Health Service, Rockville, MD, United States
[3]James J. Peters VA Medical Center, New York, NY, United States
[4]Icahn School of Medicine at Mt. Sinai, New York, NY, United States
[5]Atlanta VA Medical Center, Atlanta, GA, United States
[6]Emory University School of Medicine, Atlanta, GA, United States
[7]Rollins School of Public Health, Emory University, Atlanta, GA, United States
[8]Infectious Diseases Section, Department of Medicine, Baylor College of Medicine, Houston, TX, United States
[9]Michael E. DeBakey VA Medical Center, Houston, TX, United States
[10]VA Greater Los Angeles Healthcare System, Los Angeles, CA, United States
[11]David Geffen School of Medicine at UCLA, Los Angeles, CA, United States
[12]Public Health Surveillance and Research, Department of Veterans Affairs, Washington, DC, United States
[13]VA Palo Alto Health Care System, Palo Alto, CA, United States
[14]Division of Infectious Diseases and Geographic Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Cristina V Cardemil, MD, MPH
Centers for Disease Control and Prevention
1600 Clifton Road
Atlanta, GA, 30329
United States
Phone: 1 404 639 8241
Email: cristina.cardemil@nih.gov

Abstract

Background: COVID-19 has disproportionately affected older adults and certain racial and ethnic groups in the United States. Data quantifying the disease burden, as well as describing clinical outcomes during hospitalization among these groups, are needed.

Objective: We aimed to describe interim COVID-19 hospitalization rates and severe clinical outcomes by age group and race and ethnicity among US veterans by using a multisite surveillance network.

Methods: We implemented a multisite COVID-19 surveillance platform in 5 Veterans Affairs Medical Centers located in Atlanta, Bronx, Houston, Palo Alto, and Los Angeles, collectively serving more than 396,000 patients annually. From February 27 to July 17, 2020, we actively identified inpatient cases with COVID-19 by screening admitted patients and reviewing their laboratory test results. We then manually abstracted the patients' medical charts for demographics, underlying medical conditions, and clinical outcomes. Furthermore, we calculated hospitalization incidence and incidence rate ratios, as well as relative risk for invasive mechanical ventilation, intensive care unit admission, and case fatality rate after adjusting for age, race and ethnicity, and underlying medical conditions.

Results: We identified 621 laboratory-confirmed, hospitalized COVID-19 cases. The median age of the patients was 70 years, with 65.7% (408/621) aged ≥65 years and 94% (584/621) male. Most COVID-19 diagnoses were among non-Hispanic Black (325/621, 52.3%) veterans, followed by non-Hispanic White (153/621, 24.6%) and Hispanic or Latino (112/621, 18%) veterans. Hospitalization rates were the highest among veterans who were ≥85 years old, Hispanic or Latino, and non-Hispanic Black (430, 317, and 298 per 100,000, respectively). Veterans aged ≥85 years had a 14-fold increased rate of hospitalization compared with those aged 18-29 years (95% CI: 5.7-34.6), whereas Hispanic or Latino and Black veterans had a 4.6- and 4.2-fold increased rate of hospitalization, respectively, compared with non-Hispanic White veterans (95% CI: 3.6-5.9). Overall, 11.6% (72/621) of the patients required invasive mechanical ventilation, 26.6% (165/621) were admitted to the intensive care unit, and 16.9% (105/621) died in the hospital. The adjusted relative risk for invasive mechanical ventilation and admission to the intensive care unit did not differ by age group or race and ethnicity, but veterans aged ≥65 years had a 4.5-fold increased risk of death while hospitalized with COVID-19 compared with those aged <65 years (95% CI: 2.4-8.6).

Conclusions: COVID-19 surveillance at the 5 Veterans Affairs Medical Centers across the United States demonstrated higher hospitalization rates and severe outcomes among older veterans, as well as higher hospitalization rates among Hispanic or Latino and non-Hispanic Black veterans than among non-Hispanic White veterans. These findings highlight the need for targeted prevention and timely treatment for veterans, with special attention to older aged, Hispanic or Latino, and non-Hispanic Black veterans.

(JMIR Public Health Surveill 2021;7(1):e24502) doi:10.2196/24502

KEYWORDS

veteran; COVID-19; hospitalization; outcome; fatality; mortality; prevention; at-risk

Introduction

COVID-19 has been found to be more prevalent in older adults, men, and those with certain underlying comorbidities, thereby resulting in higher rates of hospitalization and deaths among these groups [1,2]. The COVID-19 burden has also disproportionately affected certain racial and ethnic groups in the United States, including non-Hispanic Black, Hispanic or Latino, and American Indian or Alaska Native persons [3].

Over 9 million veterans (defined here as former members of the armed forces) are enrolled in the Veterans Affairs (VA)–integrated health care program [4]. These US veterans constitute a population that is older, is predominantly male, has a high rate of chronic medical conditions, and has a higher representation from certain racial and ethnic groups compared with the general population. We built a COVID-19 surveillance system on the infrastructure of SUPERNOVA (Surveillance Platform for Enteric and Respiratory Infectious Organisms in the Veterans Affairs population), a population-based platform that captures information on cases of veterans with acute gastroenteritis and acute respiratory illness [5] from 5 VA Medical Centers (VAMCs) located in Atlanta, Georgia; Bronx, New York; Houston, Texas; Los Angeles, California; and Palo Alto, California. This surveillance system helps understand the epidemiology, hospitalization rates, and underlying medical conditions among these US veterans. There are limited studies describing the impact of COVID-19 on this specific population, but recently published articles indicate that although higher rates of hospitalization and severe outcomes are not observed among Black and Hispanic veterans, they are more likely to test positive for COVID-19 and have higher risks for sepsis and respiratory, neurologic, and renal in-hospital complications than White veterans [6-9]. The objective of this study was to describe the interim COVID-19 hospitalization rates, severe outcomes, and prevalence of underlying medical conditions by age group and race and ethnicity among veterans in this network.

Methods

Study Population

SUPERNOVA sites served 396,280 unique veterans in outpatient and inpatient settings in the fiscal year 2019 (Atlanta: n=118,258; Bronx: n=24,116; Houston: n=109,890; Los Angeles: n=82,574; and Palo Alto: n=61,442). The veteran catchment population at these sites was 32% (126,188/396,280) non-Hispanic Black, 47% (186,174/396,280) non-Hispanic White, and 11% (42,098/396,280) Hispanic or Latino. The overall median age of the veterans in the catchment population was 64 years, and 87% (342,944/396,280) of all veterans were male.

Screening and Data Abstraction

To identify eligible patients for COVID-19 surveillance, trained research personnel screened admission logs and diagnoses [5], lists of patients being tested for COVID-19, and the results from respiratory specimens submitted for testing. Patients were eligible for inclusion in this study if they were admitted to the VAMC and had a positive molecular test for SARS-CoV-2 within 14 days prior to admission or during hospitalization. COVID-19 testing was ordered by clinicians either prior to or during hospitalization, and tests were conducted at the hospital or at public or commercial laboratories serving each VAMC. Data were manually summarized from patients' electronic medical charts. Research personnel reviewed the patient's entire hospitalization course such as admission and discharge notes; International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) discharge codes; and problem lists to identify patients with any underlying medical conditions on a prespecified list. The complete list of variables for abstraction included over 140 underlying medical conditions grouped into 10 categories. Therefore, for rapid abstraction and analyses, the data abstraction for this report was intentionally limited to the underlying medical conditions (ie, chronic kidney disease, chronic obstructive pulmonary disease (COPD), coronary artery disease, heart failure, hypertension, diabetes, and obesity) that had been identified from surveillance early during the pandemic as being more prevalent among hospitalized COVID-19 patients [10]. This report presents data analyzed for patients admitted from February 27 to July 17, 2020, and their clinical outcomes available through July 31, 2020.

Statistical Analysis

Hospitalization incidence per 100,000 person-years among veterans was calculated based on previously published methods by using the catchment population of unique individuals served at each facility [11]. Severe clinical outcomes were defined as patients requiring invasive mechanical ventilation, admission to an intensive care unit, or deaths among hospitalized patients. Binomial regression was used to generate incidence rate ratios and 95% CIs for hospitalization rates, and relative risks and 95% CI for severe outcomes. All models were adjusted for age and race and ethnicity; outcome models were also adjusted for the number of underlying medical conditions.

Case fatality rate (CFR) was calculated as a 14-day moving average. We plotted the distribution of cases and compared the CFR over 2 time periods: one early during the pandemic and one later in the summer months. Patients who were still hospitalized were excluded from the analysis. Differences between groups and time periods were compared by chi-square tests and considered significant if $P<.05$.

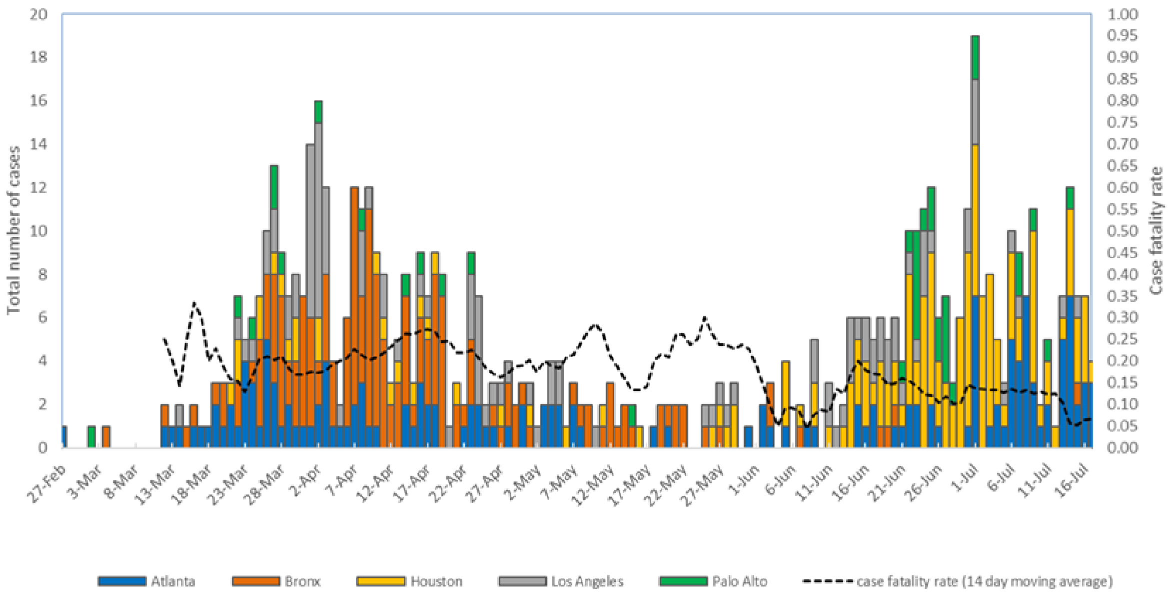
Ethical Review

This study was reviewed, informed consent was waived, and a final approval was obtained by 5 VA sites, research and development committees, and the Centers for Disease Control and Prevention institutional review board(s).

Results

From February 27 through July 17, 2020, a total of 621 laboratory-confirmed, hospitalized COVID-19 cases were identified at the 5 VAMCs. As of July 31, 2020, 27 of the 621 (4.3%) patients continued to remain hospitalized. These cases initially peaked in late March and April 2020, before declining and then plateauing in May and June, and then increased again in the latter half of June and early July, with the highest peak in new hospital admissions recorded on July 1, 2020 (Figure 1). The CFR among patients whose hospitalization was completed was higher from February 27 to May 31 (75/343, 21.9%) than from June 1 to July 17 (30/251, 12%; $P<.01$).

Figure 1. Epidemic curve of hospitalized COVID-19 cases and case fatality rate for US veterans at 5 Veterans Affairs Medical Centers by hospital admission date, from February 27 to July 17, 2020. Case fatality rate represents a 14-day moving average. Rate suppressed prior to March 12, 2020, due to low case counts (n≤3).



The Bronx VAMC had the highest number of hospitalized veterans diagnosed with COVID-19 (166/621, 27%), followed by Atlanta (157/621, 25%), Houston (156/621, 25%), Los Angeles (105/621, 17%), and Palo Alto (37/621, 6%). Across all VAMCs, the patients' median age was 70 years, with 66% (408/621) of the patients aged ≥65 years and 94% (584/621), male. The highest percentage of hospitalizations due to confirmed COVID-19 diagnoses were among non-Hispanic Black veterans (325/621, 52.3%), followed by non-Hispanic White (153/621, 24.6%), Hispanic or Latino (112/621, 18%), non-Hispanic Asian (9/621, 2%), non-Hispanic American Indian or Alaskan Native (5/621, 1%), and non-Hispanic Native Hawaiian or Pacific Islander (2/621, 0.3%) veterans. About 2.4% (15/621) of all veterans had an unknown race or ethnicity.

The overall hospitalization rate was 205.7 per 100,000 population (Table 1). Hospitalization rates were the highest among veterans who were aged ≥85 years, Hispanic or Latino, and non-Hispanic Black (430, 317, and 298 per 100,000, respectively). Veterans aged ≥85 years had a 14-fold increased rate of hospitalization relative to those aged 18-29 years (95% CI: 5.7-34.6), after adjusting for race and ethnicity. Compared with non-Hispanic White veterans, Hispanic or Latino veterans had a 4.6-fold increased rate of hospitalization (95% CI: 3.6-5.9), and non-Hispanic Black veterans had a 4.2-fold increased rate of hospitalization (95% CI: 3.4-5.1), after adjusting for age.

Table 1. Distribution characteristics, incidence rates, and adjusted incidence rate ratios of US veterans hospitalized with COVID-19—overall, by age, and by race and ethnicity, at 5 Veterans Affairs Medical Centers from February 27 to July 17, 2020.

Variable	Overall distribution of characteristics, n (%)	Hospitalization incidence rates, per 100,000	Adjusted incidence rate ratios of hospitalization (95% CI)[a]
Overall	621 (100)	205.7	N/A[b]
Age range (years)			
18-29	5 (0.8)	48.4	ref[c]
30-39	21 (3.4)	61.7	1.31 (0.50-3.48)
40-49	37 (6)	121.2	2.25 (0.88-5.74)
50-64	150 (24.2)	190.9	3.68 (1.51-9.00)
65-74	218 (35.1)	238.6	5.89 (2.42-14.32)
75-84	107 (17.2)	282.7	8.16 (3.32-20.06)
≥85	83 (13.4)	429.7	14.00 (5.66-34.63)
Age group (years)			
<65	213 (34.3)	138.8	ref
≥65	408 (65.7)	274.7	2.65 (2.23-3.15)
Race and ethnicity			
Non-Hispanic White	153 (24.6)	96.1	ref
Hispanic or Latino	112 (18)	317.3	4.57 (3.58-5.85)
Non-Hispanic Black	325 (52.3)	297.9	4.18 (3.43-5.10)

[a]Incidence rate ratio by age group is adjusted for race and ethnicity; incidence rate ratio by race and ethnicity is adjusted for age.

[b]N/A: not applicable.

[c]ref: reference group.

Overall, 11.6% (72/621) of the patients required invasive mechanical ventilation, 26.6% (165/621) were admitted to the intensive care unit (ICU), and 16.9% (105/621) died while hospitalized (Table 2). Although the adjusted relative risk for invasive mechanical ventilation and ICU admission was not significantly different by age group or by race and ethnicity, an increased risk of death was observed among older adults. Notably, veterans who were aged ≥65 years had a 4.5-fold increased risk of death while hospitalized with COVID-19 compared with those aged <65 years (95% CI: 2.4-8.6), after adjusting for race and ethnicity and the number of underlying medical conditions. Additionally, most deaths were among those who were aged ≥65 years (94/105, 89.5%), and veterans who were aged ≥85 years had the highest CFR (33/105, 39.8%).

Table 2. Prevalence and adjusted relative risk of invasive mechanical ventilation, intensive care unit (ICU) admission, and deaths among US veterans hospitalized with COVID-19—overall, by age, and by race and ethnicity, at 5 Veterans Affairs Medical Centers from February 27 to July 17, 2020.

Variable	Prevalence of invasive mechanical ventilation, n (%)	Adjusted relative risk for invasive mechanical ventilation (95% CI)[a]	Prevalence of ICU admission, n (%)	Adjusted relative risk for ICU admission (95% CI)[a]	Case fatality, n (%)	Adjusted relative risk for case fatality (95% CI)[a]
Overall	72 (11.6)	N/A[b]	165 (26.6)	N/A	105 (16.9)	N/A
Age range (years)						
18-29	0 (0)	—[c]	1 (20)	—	0 (0)	—
30-39	1 (4.8)	—	4 (19)	—	0 (0)	—
40-49	2 (5.4)	—	3 (8.1)	—	1 (2.7)	—
50-64	14 (9.3)	ref[d]	37 (24.7)	ref[d]	10 (6.7)	ref[d]
65-74	31 (14.2)	1.51 (0.81-2.83)	62 (28.4)	1.12 (0.77-1.60)	36 (16.5)	2.65 (1.31-5.37)
75-84	14 (13.1)	1.10 (0.50-2.42)	39 (36.4)	1.34 (0.90-2.00)	25 (23.4)	3.20 (1.52-6.74)
≥85	10 (12)	1.12 (0.48-2.59)	19 (22.9)	0.88 (0.53-1.47)	33 (39.8)	6.28 (3.15-12.54)
Age group (years)						
<65	17 (8)	ref[e]	45 (21.1)	ref[e]	11 (5.2)	ref[e]
≥65	55 (13.5)	1.48 (0.85-2.59)	120 (29.4)	1.24 (0.90-1.71)	94 (23)	4.50 (2.37-8.56)
Race and ethnicity						
Non-Hispanic White	16 (10.5)	ref	47 (30.7)	ref	33 (21.6)	ref
Hispanic or Latino	14 (12.5)	1.16 (0.66-2.03)	25 (22.3)	0.90 (0.66-1.21)	19 (17)	0.86 (0.59-1.26)
non-Hispanic Black	35 (10.8)	1.38 (0.70-2.69)	84 (25.8)	0.79 (0.52-1.21)	46 (14.2)	0.90 (0.56-1.45)

[a]Relative risk by age group is adjusted for race and ethnicity and the number of underlying medical conditions (chronic kidney disease, chronic obstructive pulmonary disease, coronary artery disease, heart failure, hypertension, diabetes, and obesity); relative risk by race and ethnicity is adjusted for age and underlying medical conditions.

[b]N/A: not applicable.

[c]Due to the small number of cases in the younger age groups, only veterans aged ≥50 years were included in these models.

[d]Veterans aged 50-64 years were the reference group for veterans aged 65-74 years, 75-84 years, and ≥85 years.

[e]For the models comparing veterans aged <65 years with ≥65 years, all cases and age groups were included (≥18 years) and combined into the 2 age groups (ie, 18-64 and ≥65 years old). Veterans aged 18-64 years old served as the reference group.

Most patients had at least one of the 7 underlying medical conditions (547/621, 88.1%) (Table 3). By age group, veterans who were aged ≥65 years had a higher prevalence of the following underlying medical conditions than those aged <65 years: chronic kidney disease (104/408, 25.5% vs 30/213, 14.1%; *P*=.001), COPD (76/408, 18.6% vs 14/213, 6.6%; *P*<.001), coronary artery disease (103/408, 25.2% vs 17/213, 8%; *P*<.001), heart failure (78/408, 19.1% vs 11/213, 5.2%; *P*<.001), hypertension (322/408, 78.9% vs 120/213, 56.3%; *P*<.001), and diabetes (218/408, 53.4% vs 88/213, 41.3%; *P*=.004). Veterans aged <65 years had a higher prevalence of obesity than those aged ≥65 years (93/213, 43.7% vs 84/408, 20.6%; *P*<.001). By race and ethnicity, non-Hispanic Black veterans (78/325, 24%) had a higher prevalence of chronic kidney disease than Hispanic or Latino (20/112, 17.9%) and non-Hispanic White veterans (29/153, 19%; *P*=.26), whereas non-Hispanic White veterans had a higher prevalence of COPD (non-Hispanic White: 40/153, 26.1%; non-Hispanic Black: 32/325, 9.9%; Hispanic or Latino: 15/112, 13.4%; *P*<.001).

Table 3. Prevalence of underlying medical conditions among US veterans hospitalized with COVID-19—overall, by age, and by race and ethnicity, at 5 Veterans Affairs Medical Centers from February 27 to July 17, 2020.

Underlying medical conditions	Overall (n=621), n (%)	Age group			Race and ethnicity			
		<65 years (n=213), n (%)	≥65 years (n=408), n (%)	*P* value	Hispanic or Latino (n=112), n (%)	Non-Hispanic Black (n=325), n (%)	Non-Hispanic White (n=153), n (%)	*P* value[a]
Any underlying medical condition[b]	547 (88.1)	166 (77.9)	381 (93.4)	<.001	91 (81.3)	287 (88.3)	140 (91.5)	.04
Chronic kidney disease[c]	134 (21.6)	30 (14.1)	104 (25.5)	.001	20 (17.9)	78 (24)	29 (19)	.26
COPD[d]/emphysema	90 (14.5)	14 (6.6)	76 (18.6)	<.001	15 (13.4)	32 (9.8)	40 (26.1)	<.001
Coronary artery disease	120 (19.3)	17 (8)	103 (25.2)	<.001	27 (24.1)	50 (15.4)	34 (22.2)	.06
Heart failure	89 (14.3)	11 (5.2)	78 (19.1)	<.001	17 (15.2)	44 (13.5)	25 (16.3)	.71
Hypertension[e]	442 (71.2)	120 (56.3)	322 (78.9)	<.001	74 (66.1)	233 (71.7)	110 (71.9)	.49
Diabetes	306 (49.3)	88 (41.3)	218 (53.4)	.004	56 (50)	166 (51.1)	71 (46.4)	.63
Obesity[e]	177 (28.5)	93 (43.7)	84 (20.6)	<.001	32 (28.6)	97 (29.8)	40 (26.1)	.70

[a]Due to the small number of cases in the other race and ethnicity groups (non-Hispanic Asian, non-Hispanic American Indian or Alaskan Native, and non-Hispanic Native Hawaiian or Pacific Islander), comparisons were limited to the 3 race and ethnicity groups listed in this table.

[b]Any of the 7 underlying medical conditions listed in this table.

[c]Chronic kidney disease or chronic renal insufficiency.

[d]COPD: chronic obstructive pulmonary disease.

[e]Presence of hypertension and obesity were not determined based on blood pressure or height and weight measurements during the current hospitalization but were noted as part of the patient's underlying medical history after review of clinical admission and discharge notes as well as discharge codes of the International Classification of Diseases, Tenth Revision, Clinical Modification.

Discussion

The burden of COVID-19 hospitalizations at these 5 VAMCs was the highest among older, Hispanic or Latino, and non-Hispanic Black veterans. Veterans aged ≥50 years had a higher rate of hospitalization than those aged 18-29 years, and Hispanic or Latino and non-Hispanic Black veterans had a 4.6- and 4.2-fold higher rate of hospitalization, respectively, than non-Hispanic White veterans. More severe outcomes were noted among older adults, with 9 out of every 10 deaths reported among those who were aged ≥65 years. Moreover, these older veterans also had a 4.5-fold increased risk of death following hospitalization with COVID-19 as compared to those aged <65 years. These findings highlight the need for targeted prevention, control and treatment efforts for veterans, with special attention to increasing age and race and ethnicity (ie, Hispanic or Latino and non-Hispanic Black).

Our COVID-19 surveillance data demonstrated that Hispanic or Latino and non-Hispanic Black veterans were hospitalized for COVID-19 at higher rates than were non-Hispanic White veterans, which is consistent with the findings of studies in the United States and other countries [1,3,12]. Despite the higher hospitalization rates in these racial and ethnic groups, our findings did not show differences in severe outcomes by race and ethnicity, as measured by the need for invasive mechanical ventilation, ICU admission, or in-hospital deaths due to COVID-19. However, recently published data indicates a disproportionate burden of certain in-hospital complications among US veterans of racial and ethnic minority groups [9]. Therefore, there is an urgent need to understand sociodemographic and economic factors, access to care, living conditions, prior medical history, and occupational groups and exposures [13] that may be contributors to disparities in COVID-19 hospitalization rates as well as the short- and long-term clinical course after infection among US veteran minority populations.

Similar to nationwide death certificate trends, we observed a decline in the CFR since the start of the pandemic [10]. Continued surveillance will help track whether these trends persist and understand possible explanations for such trends, which may include earlier and wider access to testing and treatment, increased knowledge and availability of interventions, and possible differences in the demographics of those infected. The higher death rates observed among older individuals in this study was also noted in previous studies from China, Italy, and the United States [2,14-17]. Another surveillance system in the United States, Coronavirus Disease 2019-associated Hospitalization Surveillance Network (COVID-NET), has demonstrated a similar increase in hospitalization incidence and death with increasing age [17], indicating the need for continued prevention efforts for older adults.

Nearly all patients (88.1%) in this cohort had an underlying medical condition, and veterans aged ≥65 years had higher prevalence of several key comorbidities. Many studies have indicated the heightened risk of both older age and specific underlying medical conditions (eg, chronic kidney disease, COPD, serious heart conditions, and diabetes) for severe COVID-19 outcomes, including hospitalization, ICU admission, mechanical ventilation, and death [17,18]. Veterans who use VA health care are older and more likely to be diagnosed with multiple health conditions than the general population [19], making this an important group to monitor for severe outcomes resulting from COVID-19.

This study has limitations. First, early during the pandemic, limited COVID-19 testing was conducted nationwide, and this may have led to an under-ascertainment of patients with COVID-19. Second, although prior studies have indicated that 91% of veterans enrolled in VA healthcare in these 5 site catchment areas use their local VAMCs for care [5], veterans hospitalized elsewhere would not have been captured in our surveillance, thereby resulting in underestimation of hospitalization rates. Third, as the number of cases in younger age groups was much smaller than in older adults, we were not able to calculate relative statistics for severe clinical outcomes in veterans aged 18-49 years. Finally, veterans are predominantly male; therefore, these results may not be representative of all US adults.

In conclusion, multisite COVID-19 surveillance conducted at 5 VAMCs across the United States from February 27 to July 17, 2020, demonstrated higher hospitalization rates and more deaths among older veterans as well as higher hospitalization rates among Hispanic or Latino and non-Hispanic Black veterans than among non-Hispanic White veterans. To prevent contracting and spreading COVID-19, veterans and persons living with them should follow recommendations such as practicing social and physical distancing, regularly washing hands, and wearing masks, including inside the house. From a systems-level perspective, policies and procedures that focus on veterans, including timely and high-quality care as well as access to therapeutics and future vaccines, will help protect this population and reduce morbidity and mortality therein. Ongoing data collection at these sites will aid in further characterizing the clinical characteristics of COVID-19, better understanding the risk factors and management of hospitalized patients, characterization of the immune response to SARS-CoV-2 infection through collection of patients' serum samples, and description of long-term outcomes following hospitalization.

Acknowledgments

We would like to thank the participants at the various study sites for their time and contributions. We would also like to thank the following staff for their contributions to this surveillance platform, without whom this research would not have been possible: Awilda Mero and Ilda Graham (Bronx VAMC); Alexis Whitmire and Amy Hartley (Atlanta VAMC); Rosalba Gomez Morones, Bashir Lengi, and Julio Rojas-Quintero (Houston VAMC); Evan Goldin (Los Angeles VAMC); Theresa Peters and Madhuri Agrawal (Palo Alto VA).

This work was supported by the Centers for Disease Control and Prevention.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the Agency for Toxic Substances and Disease Registry, the Department of Veterans Affairs, or the United States government.

Conflicts of Interest

None declared.

References

1. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 - COVID-NET, 14 states, March 1-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):458-464 [FREE Full text] [doi: 10.15585/mmwr.mm6915e3] [Medline: 32298251]
2. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
3. Gold JAW, Wong KK, Szablewski CM, Patel PR, Rossow J, da Silva J, et al. Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 - Georgia, March 2020. MMWR Morb Mortal Wkly Rep 2020 May 08;69(18):545-550 [FREE Full text] [doi: 10.15585/mmwr.mm6918e1] [Medline: 32379729]
4. Veterans Health Administration. U.S. Department of Veterans Affairs. URL: https://www.va.gov/health/aboutvha.asp [accessed 2021-01-12]
5. Cardemil CV, Balachandran N, Kambhampati A, Grytdal S, Dahl R, Rodriguez-Barradas M, et al. Incidence, etiology, and severity of acute gastroenteritis among prospectively enrolled patients in 4 Veterans Affairs hospitals and outpatient centers, 2016-18. Clin Infect Dis 2020 Jun 25. [doi: 10.1093/cid/ciaa806] [Medline: 32584956]

6. Rentsch C, Kidwai-Khan F, Tate J, Park L, King J, Skanderson M, et al. Covid-19 by race and ethnicity: a national cohort study of 6 million United States veterans. medRxiv Preprint posted online on May 18, 2020. [FREE Full text] [doi: 10.1101/2020.05.12.20099135] [Medline: 32511524]
7. Rentsch CT, Kidwai-Khan F, Tate JP, Park LS, King JT, Skanderson M, et al. Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: a nationwide cohort study. PLoS Med 2020 Sep;17(9):e1003379 [FREE Full text] [doi: 10.1371/journal.pmed.1003379] [Medline: 32960880]
8. Ioannou GN, Locke E, Green P, Berry K, O'Hare AM, Shah JA, et al. Risk factors for hospitalization, mechanical ventilation, or death among 10 131 US veterans with SARS-CoV-2 infection. JAMA Netw Open 2020 Sep 01;3(9):e2022310 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22310] [Medline: 32965502]
9. Cates J, Lucero-Obusan C, Dahl RM, Schirmer P, Garg S, Oda G, et al. Risk for in-hospital complications associated with COVID-19 and influenza - Veterans Health Administration, United States, October 1, 2018-May 31, 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 23;69(42):1528-1534 [FREE Full text] [doi: 10.15585/mmwr.mm6942e3] [Medline: 33090987]
10. COVIDView: A Weekly Surveillance Summary of U.S. COVID-19 Activity. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html [accessed 2020-07-28]
11. Grytdal S, Browne H, Collins N, Vargas B, Rodriguez-Barradas M, Rimland D, et al. Trends in incidence of norovirus-associated acute gastroenteritis in 4 Veterans Affairs Medical Center populations in the United States, 2011-2015. Clin Infect Dis 2020 Jan 01;70(1):40-48 [FREE Full text] [doi: 10.1093/cid/ciz165] [Medline: 30901024]
12. Lassale C, Gaye B, Hamer M, Gale CR, Batty GD. Ethnic disparities in hospitalisation for COVID-19 in England: The role of socioeconomic factors, mental health, and inflammatory and pro-inflammatory factors in a community-based cohort study. Brain Behav Immun 2020 Aug;88:44-49 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.074] [Medline: 32497776]
13. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? BMJ 2020 Apr 20;369:m1548. [doi: 10.1136/bmj.m1548] [Medline: 32312785]
14. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/s0140-6736(20)30566-3]
15. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ 2020 Mar 26;368:m1091 [FREE Full text] [doi: 10.1136/bmj.m1091] [Medline: 32217556]
16. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
17. Kim L, Garg S, O'Halloran A, Whitaker M, Pham H, Anderson E, et al. Risk factors for intensive care unit admission and in-hospital mortality among hospitalized adults identified through the U.S. coronavirus disease 2019 (COVID-19)-associated hospitalization surveillance network (COVID-NET). Clin Infect Dis 2020 Jul 16 [FREE Full text] [doi: 10.1093/cid/ciaa1012] [Medline: 32674114]
18. People with Certain Medical Conditions. Centers for Disease Control and Prevention - COVID-19. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [accessed 2020-07-23]
19. Eibner C, Krull H, Brown KM, Cefalu M, Mulcahy AW, Pollard M, et al. Current and projected characteristics and unique health care needs of the patient population served by the Department of Veterans Affairs. Rand Health Q 2016 May 09;5(4):13 [FREE Full text] [Medline: 28083423]

Abbreviations

CFR: case fatality rate
COPD: chronic obstructive pulmonary disease
ICU: intensive care unit
SUPERNOVA: Surveillance Platform for Enteric and Respiratory Infectious Organisms in the Veterans Affairs population
VA: Veterans Affairs
VAMC: Veterans Affairs Medical Centers

Edited by T Sanchez; submitted 22.09.20; peer-reviewed by K Kapinos, B Smith, J Ferguson; comments to author 16.11.20; revised version received 05.12.20; accepted 14.12.20; published 22.01.21.

Please cite as:
Cardemil CV, Dahl R, Prill MM, Cates J, Brown S, Perea A, Marconi V, Bell L, Rodriguez-Barradas MC, Rivera-Dominguez G, Beenhouwer D, Poteshkina A, Holodniy M, Lucero-Obusan C, Balachandran N, Hall AJ, Kim L, Langley G
COVID-19–Related Hospitalization Rates and Severe Outcomes Among Veterans From 5 Veterans Affairs Medical Centers: Hospital-Based Surveillance Study
JMIR Public Health Surveill 2021;7(1):e24502
URL: http://publichealth.jmir.org/2021/1/e24502/
doi:10.2196/24502
PMID:33338028

Original Paper

The Influence of Average Temperature and Relative Humidity on New Cases of COVID-19: Time-Series Analysis

Zonglin He[1,2*], MBBS; Yiqiao Chin[2*], MBBS; Shinning Yu[2], MBBS; Jian Huang[3], MPH, PhD; Casper J P Zhang[4], MPH, PhD; Ke Zhu[1], MBBS; Nima Azarakhsh[5], PhD; Jie Sheng[6], BA, MSc; Yi He[7], BA, MSc; Pallavi Jayavanth[5], PhD; Qian Liu[8], PhD; Babatunde O Akinwunmi[9], MD, MPH, MMSc, FWACS; Wai-Kit Ming[1], MD, MPH, PhD, MMSc, EMBA

[1]School of Medicine, Jinan University, Guangzhou, China

[2]Faculty of Medicine, International School, Jinan University, Guangzhou, China

[3]MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

[4]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong

[5]International School, Jinan University, Guangzhou, China

[6]College of Economics, Jinan University, Guangzhou, China

[7]Department of Statistics, University of Oxford, Oxford, United Kingdom

[8]National Media Experimental Teaching Demonstration Center, School of Journalism and Communication, Jinan University, Guangzhou, China

[9]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, MPH, PhD, MMSc, EMBA
School of Medicine
Jinan University
601 West Huangpu Avenue
Guangzhou
China
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Abstract

Background: The influence of meteorological factors on the transmission and spread of COVID-19 is of interest and has not been investigated.

Objective: This study aimed to investigate the associations between meteorological factors and the daily number of new cases of COVID-19 in 9 Asian cities.

Methods: Pearson correlation and generalized additive modeling (GAM) were performed to assess the relationships between daily new COVID-19 cases and meteorological factors (daily average temperature and relative humidity) with the most updated data currently available.

Results: The Pearson correlation showed that daily new confirmed cases of COVID-19 were more correlated with the average temperature than with relative humidity. Daily new confirmed cases were negatively correlated with the average temperature in Beijing (r=–0.565, P<.001), Shanghai (r=–0.47, P<.001), and Guangzhou (r=–0.53, P<.001). In Japan, however, a positive correlation was observed (r=0.416, P<.001). In most of the cities (Shanghai, Guangzhou, Hong Kong, Seoul, Tokyo, and Kuala Lumpur), GAM analysis showed the number of daily new confirmed cases to be positively associated with both average temperature and relative humidity, especially using lagged 3D modeling where the positive influence of temperature on daily new confirmed cases was discerned in 5 cities (exceptions: Beijing, Wuhan, Korea, and Malaysia). Moreover, the sensitivity analysis showed, by incorporating the city grade and public health measures into the model, that higher temperatures can increase daily new case numbers (beta=0.073, Z=11.594, P<.001) in the lagged 3-day model.

Conclusions: The findings suggest that increased temperature yield increases in daily new cases of COVID-19. Hence, large-scale public health measures and expanded regional research are still required until a vaccine becomes widely available and herd immunity is established.

(JMIR Public Health Surveill 2021;7(1):e20495) doi:10.2196/20495

KEYWORDS

COVID-19; coronavirus; meteorological factors; temperature; humidity; weather; transmission; virus; Asia; time-series; analysis; public health

Introduction

In December 2019, several cases of pneumonia of unknown etiology were reported in Wuhan, Hubei Province, China [1]. A novel strain of coronavirus was identified from the nasopharyngeal swab specimens of infected patients, which was later named SARS-CoV-2, which results in the disease COVID-19. As the number of infectees increased, the World Health Organization declared the outbreak as a public health emergency of international concern on January 31, 2020 [2].

SARS-CoV-2, of the *Coronaviridae* family, is an enveloped, single-stranded, positive-sense RNA (ribonucleic acid) virus, which is closely related to the SARS (severe acute respiratory syndrome)-like coronaviruses, and based on the phylogenetic analysis, these coronaviruses have a common ancestor that resembles the bat coronavirus HKU9-1 [3,4]. Evidence has shown that SARS-CoV-2 can transmit from person to person via respiratory droplets, fecal-oral route, direct contact, and aerosols [5-7]. Moreover, the long incubation period (1-14 days) increases the difficulty of controlling the COVID-19 outbreak. Studies have shown the mean incubation period to be 5.1 days (range: 2.2-11.5 days, 95% CI 4.5-5.8) [8]; one study estimated the mean incubation period to be 6.4 days (range: 2-14 days) [9]. By early July 2020, 215 countries and regions had reported high infection rates, with over 7,000,000 confirmed cases, 400,000 deaths, and a fatality rate of over 5.84% worldwide [10].

Human respiratory pathogens (bacterial pathogens like pneumococcus and viruses like rubella and influenza) usually exhibit an annual, seasonal pattern, with an increase in incidence during winter and a decrease during summer. Although there is much data on the influenza virus, respiratory syncytial virus, and the SARS outbreak in 2003 following this pattern, it is difficult to predict whether COVID-19 will follow the trend and be eliminated during warmer seasons, since our understanding of the forces driving the seasonality of infectious diseases remains limited. As influenza is a common viral disease, a proportion of the population already possesses some levels of immunity, and when more patients recover, herd immunity constrains the transmission of the virus. The low or absent prevalence of SARS-CoV or the Middle East respiratory syndrome coronavirus (MERS-CoV) in the summer was also observed to have strongly relied on the use of effective therapeutic treatment and strict public health measures [11,12]. However, SARS-CoV-2 is a novel virus among humans, immunity to this ongoing viral pandemic is limited, and an effective pharmaceutical therapy or vaccine is yet to be available.

Seasonality in the outbreak of respiratory infectious diseases may be due to seasonal variations in host physiology (susceptibility, individual immunity, and herd immunity) [13], genetics [14], viral stability [15,16] and infectivity [17-20], presence of latent infectors to spread the virus [21,22], and atmospheric dispersion and transocean intercontinental migration [23-26], which are mainly driven by meteorological factors, including temperature and humidity [17]. Geological features and latitude play a major role in forming a meteorological pattern. Lowen et al [19] assumed the seasonality of influenza and respiratory syncytial virus epidemics in temperate climates were mainly attributable to the low absolute humidity, and specific factors associated with it, namely low temperature, increased population, and low micronutrient levels (such as low vitamin D levels) [27,28]. During the winter in temperate zones, temperature and humidity are low, there is dryness and coldness, and viruses are more easily transmitted via aerosols than direct or indirect contact, where the low temperature renders the virus viable and stable in aerosols and on surfaces. However, in the rainy seasons of tropical regions, where it is hot and wet, aerosol transmission decreases but transmission by direct contact increases [29]. Although high temperatures can decrease the stability of the virus and reduce the level of aerosolization of viral droplets, the amount of virus deposited on surfaces increases as the temperature increases [30].

Because of the relatively stable structure of the coronaviruses, the infectivity of the coronavirus would not be affected by a relative humidity >95% and a temperature of 28 °C to 33 °C [30]. For example, the transmission of SARS is more efficient in temperatures between 16 °C and 28 °C, and the wide use of air-conditioning also provides shelter for the breeding and transmission of SARS, where the virus is stable for 3 weeks at room temperature [30]. SARS-CoV-2 has many similarities with SARS-CoV-1, but whether meteorological factors influence viral transmission has not been established. Therefore, in this study, we investigated whether and how meteorological factors affect the spread of COVID-19, with the specific aim of examining the relationship between meteorological factors and the number of COVID-19 daily cases in Asian cities at different latitudes. The objective was to provide scientific evidence concerning the future progression of COVID-19 based on climate factors.

Methods

Meteorological Data

We obtained daily meteorological data from the National Meteorological Information Center [31]. The meteorological data of cities in China were collected from the National Meteorological Bureau and Data Center of the Ministry of Ecology [32,33] from January 20 to March 18, 2020. The meteorological data in Seoul, Tokyo, Kuala Lumpur, and Singapore were obtained from timeanddate.com [34] from January 20 to March 18, 2020. We retrieved the highest and lowest temperatures and relative humidity for four different quarters of the day and computed the average values of temperature and relative humidity for that day.

COVID-19 Surveillance Data

The daily number of domestic COVID-19 cases in 5 cities in China (Wuhan, Beijing, Shanghai, Guangzhou, and Hong Kong) and Singapore were obtained from the Ministry of Health in China [35] and Singapore [36], respectively, from the inception of cases to March 18, 2020. Given that the daily number of domestic cases at the city level in Japan, Korea, and Malaysia were not available from their corresponding Ministries of Health, as an alternative, we used the number of domestic cases at the national level in these 3 countries for analysis, which was obtained from the Ministry of Health of Japan, Korea, and Malaysia, respectively. The clinical criterion for the diagnosis of COVID-19 was based on the high-throughput sequencing or reverse transcription polymerase chain reaction assay via nasopharyngeal swab.

Statistical Analysis

First, the normality of the daily new cases and meteorological data was evaluated by examining their skewness and kurtosis. Subsequently, descriptive analysis was performed to analyze the city-specific characteristics of confirmed COVID-19 cases. Second, to analyze the overall correlation between the two meteorological factors and daily new confirmed cases, Pearson correlation and covariances between the daily COVID-19 new cases and daily meteorological factors were tested, and the linearity between variables was tested using linear regression (with city level and public health measures as controlling factors), as shown in Multimedia Appendix 1. All statistical analyses were performed using STATA 14.0 (Stata Corp).

Subsequently, city-specific generalized additive models (GAMs) with a Poisson family and logarithm link function were used to estimate the associations of daily COVID-19 new cases with average temperature and relative humidity. GAMs are useful for identifying exposure-response relationships from various types of data, particularly in exploring nonparametric relationships [37]. The model was built as follows:

$$\textit{Daily new cases}_i \sim \textit{Poisson}\ (\mu_i)$$

$$\log\ \mu_i = \beta_0 + f_1(\textit{Date}_i) + f_2(\textit{Temperature}_i) + f_3(\textit{Relative humidity}_i)$$

where the terms f_1 to f_2 are the smoothing functions, and β_0 is the intercept. The GAM analysis was performed in R software, version 3.6.0 (The R Project for Statistical Computing), using the package “mgcv.” We first established a basic temporal model for COVID-19 cases without including meteorological variables. To adjust for long-term trends and seasonality, we included penalized spline functions of time in the model. The degrees of freedom (df) for time was optimized by minimizing the absolute values of the partial autocorrelation function (PACF) of residuals for lags up to 30 days [38-41]. Additionally, the selection of an optimal model was based on the lowest Akaike information criterion (AIC). Next, we built meteorological models based on the temporal models to account for the lagged effect of meteorological variables and the incubation period of COVID-19 [7,42]. Specifically, we examined the effect of meteorological variables with different time lags, including 1-day lag (lag 1d), 3-day lag (lag 3d), 5-day lag (lag 5d), single-week lag (lag 7d), and 2-week lag (lag 14d), to capture immediate effects and lagged effects, respectively. Automated penalized splines were used to fit the association between the daily new cases and each of the meteorological variables. The date when the accumulated cases exceeded 30 in each city or country was selected as the inception of the date incorporated in the model to equalize the starting speed of outbreak and thus avoid misinterpretation and overfitting.

To control for the autocorrelation, the model’s residuals were examined for serial correlation using PACF. Default plots showing the smooth components of a fitted GAM were produced. The percentage of deviance explained by each variable was calculated, which represents the scale of a linear predictor that contributes to the component smooth functions. At last, the graph of smooth components was grouped based on similar latitudes to mitigate the influence by latitude.

Sensitivity Analysis

To verify model results, a sensitivity analysis was performed. Owing to the different public health measures put forward in different cities or countries, the city scale was taken into consideration in the model. We postulated that the during the following 15 days after the accumulated cases exceeded 40, public health measures and precautions were not readily available; thus, the transmission of the disease resembles natural transmission. The model was built as follows:

$$\textit{Daily new cases}_i \sim \textit{Poisson}\ (\mu_i)$$

$$\log\ \mu_i = \beta_0 + f_1(\textit{Date}_i) + f_2(\textit{Temperature}_i) + f_3(\textit{Relative humidity}_i) + f_4(\textit{City level}_i)$$

where the terms f_1 is the smoothing functions, $f_{2\text{-}4}$ are the linear functions, and β_0 is the intercept. GAM with linear components facilitate the analysis of the effects of explanatory variables in a way that closely resembles the analysis of covariates in a standard linear model, but with less confining assumptions. This is achieved by specifying a link function, which links the systematic component of the linear model with a wider class of outcome variables and residual forms.

Results

Descriptive Statistical Results

The daily new cases of COVID-19 and meteorological data from January 20 to March 18, 2020 (61 days), in 9 different cities are shown in Tables 1 and 2. During the study period, the temperature in the 5 cities in China ranged from –9 °C to 26.25 °C (Beijing: –9 °C to 26 °C; Shanghai: 0 °C to 22 °C; Guangzhou 8.75 °C to 26.25 °C; Wuhan: 1 °C to 21.75 °C; Hong Kong: 11.25 °C to 26.25 °C). The temperature in Singapore ranged from 25.5 °C to 32.75 °C; in Seoul, from –9.5 °C to 12.25 °C; in Tokyo, from 0.75 °C to 18.5 °C; and in Kuala Lumpur, from 24.5 °C to 33.5 °C. Multimedia Appendices 2 and 3 show the distribution of daily COVID-19 new cases associated with average temperature and average relative humidity over time, respectively.

Table 1. Descriptive statistics for different Asian cities.

Variables (latitude)	Mean (SD)	Minimum	Q1	Median	Q3	Maximum
Beijing, China (39°56′N)						
Domestic case (persons)	8.00 (9.60)	0.00	1.00	5.00	12.00	39.00
Highest temperature (°C)	8.29 (5.34)	–4.00	5.00	8.00	11.00	26.00
Lowest temperature (°C)	–3.10 (3.44)	–9.00	–6.00	–3.00	–1.00	4.00
Relative humidity (%)	52.49 (19.48)	23.50	36.25	54.75	67.00	92.75
Shanghai, China (31°14′N)						
Domestic case (persons)	7.00 (8.20)	0.00	0.00	3.00	13.00	27.00
Highest temperature (°C)	11.34 (3.61)	4.00	8.75	10.75	13.50	21.50
Lowest temperature (°C)	8.28 (3.18)	1.50	6.25	8.25	10.25	15.00
Relative humidity (%)	71.11 (14.70)	42.00	60.00	71.75	82.50	94.75
Wuhan, China (30°37′N)						
Domestic case (persons)	1035.00 (1968.80)	0.00	120.00	502.50	1524.50	13,436.00
Highest temperature (°C)	11.45 (4.40)	3.50	8.00	11.00	14.00	20.75
Lowest temperature (°C)	8.22 (3.76)	1.50	5.25	8.25	11.00	17.50
Relative humidity (%)	74.03 (11.70)	51.75	66.25	74.75	84.75	93.25
Guangzhou, China (23°10′N)						
Domestic case (persons)	7.00 (9.80)	0.00	0.00	3.00	12.00	38.00
Highest temperature (°C)	19.32 (3.95)	11.25	16.25	20.00	22.00	26.25
Lowest temperature (°C)	16.55 (3.82)	8.75	13.75	17.00	19.75	22.50
Relative humidity (%)	67.95 (13.19)	34.75	60.00	68.75	78.50	87.50
Hong Kong, China (22°15′N)						
Domestic case (persons)	2.00 (2.50)	0.00	0.00	1.50	3.00	13.00
Highest temperature (°C)	20.66 (3.11)	13.50	18.50	21.00	23.25	26.25
Lowest temperature (°C)	18.39 (3.12)	11.25	16.50	19.25	20.50	24.25
Relative humidity (%)	70.31 (12.67)	31.00	65.00	74.00	79.00	88.50

Table 2. Descriptive statistics for different Asian cities, continued.

Variables (latitude)	Mean (SD)	Minimum	Q1	Median	Q3	Maximum
Singapore (1°18′N)						
Domestic case (persons)	3.00 (3.10)	0.00	1.00	3.00	4.00	14.00
Highest temperature (°C)	30.46 (0.97)	27.75	29.75	30.50	31.25	32.75
Lowest temperature (°C)	27.64 (0.77)	26.00	27.25	27.75	28.25	29.00
Relative humidity (%)	76.09 (4.95)	63.75	73.25	75.50	79.25	89.50
Seoul, Korea (37°33′N)						
Domestic case (persons)	142.00 (216.00)	0.00	0.00	3.00	210.00	909.00
Highest temperature (°C)	5.45 (3.93)	–6.00	3.25	6.25	8.00	12.25
Lowest temperature (°C)	0.95 (3.56)	–9.50	–1.25	1.50	3.75	6.25
Relative humidity (%)	68.50 (10.78)	48.25	59.50	66.75	76.75	91.50
Tokyo, Japan (35°69′N)						
Domestic case (persons)	15.00 (17.90)	0.00	0.00	8.00	25.00	60.00
Highest temperature (°C)	10.30 (3.11)	3.50	8.75	10.25	12.50	18.50
Lowest temperature (°C)	6.92 (2.65)	0.75	5.25	6.75	8.75	14.00
Relative humidity (%)	61.41 (17.52)	32.75	47.25	59.50	76.25	97.75
Kuala Lumpur, Malaysia (3°8′N)						
Domestic case (persons)	14.00 (37.20)	0.00	0.00	0.00	7.00	190.00
Highest temperature (°C)	30.98 (1.26)	26.50	30.25	31.00	31.75	33.50
Lowest temperature (°C)	27.19 (1.03)	24.50	26.50	27.25	28.00	29.00
Relative humidity (%)	77.19 (6.74)	62.50	74.00	77.50	82.25	92.50

To test the potential collinearity between the meteorological parameters, a series of Pearson correlations and covariances, as well as linear regressions were conducted (Table 3 and Multimedia Appendix 1). Daily confirmed new cases were negatively correlated with average temperature in Beijing (r=–0.565, P<.001), Shanghai (r=–0.471, P<.001), and Guangzhou (r=–0.530, P<.001). In contrast, Japan exhibited a positive correlation (r=0.416, P<.001). The correlation between average temperature and relative humidity was found to be positive in Shanghai, Guangzhou, Hong Kong, Korea, and Japan, and negative in Beijing, Wuhan, Singapore, and Malaysia, according to the pairwise Pearson correlation test (Table 3).

Table 3. Pearson correlation coefficient (*r*) between daily new COVID-19 cases and meteorological factors.

Variables	Daily new cases		Average temperature	
	r	*P* value	*r*	*P* value
Beijing, China				
Daily new cases	1.00	—[a]	—	—
Average temperature	–0.57	<.001	1.00	—
Humidity	0.15	.30	–0.26	.05
Shanghai, China				
Daily new cases	1.00	—	—	—
Average temperature	–0.47	<.001	1.00	—
Humidity	–0.11	.45	0.09	.50
Wuhan, China				
Daily new cases	1.00	—	—	—
Average temperature	0.04	.81	1.00	—
Humidity	0.10	.48	–0.42	<.001
Guangzhou, China				
Daily new cases	1.00	—	—	—
Average temperature	–0.53	<.001	1.00	—
Humidity	–0.29	.05	0.34	.01
Hong Kong, China				
Daily new cases	1.00	—	—	—
Average temperature	0.08	.56	1.00	—
Humidity	0.07	.60	0.58	<.001
Singapore				
Daily new cases	1.00	—	—	—
Average temperature	0.27	.04	1.00	—
Humidity	0.04	.74	–0.58	<.001
Seoul, Korea				
Daily new cases	1.00	—	—	—
Average temperature	0.29	.02	1.00	—
Humidity	0.14	.30	0.32	.01
Tokyo, Japan				
Daily new cases	1.00	—	—	—
Average temperature	0.42	<.001	1.00	—
Humidity	0.20	.14	0.21	.11
Kuala Lumpur, Malaysia				
Daily new cases	1.00	—	—	—
Average temperature	0.21	.11	1.00	—
Humidity	–0.17	.20	–0.74	<.001

[a]Not applicable.

City-Specific GAM Analysis of Daily New COVID-19 Cases With Meteorological Factors

The final GAM model of daily new COVID-19 cases incorporated date (time-series), average temperature, and mean relative humidity. All estimates and significance levels were listed in Multimedia Appendix 1. The models with the best performance (lowest AIC) for each city were as follows: no-lag model for Shanghai and Singapore; lag 1d model for Beijing and Wuhan; lag 5d model for Guangzhou, Korea, and Kuala Lumpur; and lag 14d model for Hong Kong and Japan (Table 4).

GAM results were reported using the smoothing components plot for temperature and relative humidity in Multimedia Appendices 4 and 5, respectively. The smoothing components plots demonstrated the estimated smoothing spline functions with the linear effect subtracted out, and each panel represented the weighted sum of basis functions for each time-varying covariate and corresponds to the hypothesized model.

The significant smoothers indicated that the correlations between new cases of COVID-19 and explanatory variables were nonlinear. As shown in Multimedia Appendix 4, the no-lag model suggested that holding all linear and the other nonlinear terms fixed, daily new cases of COVID-19 were not influenced by the temperature, but the case number decreased when the temperature reached 5 °C, 18 °C, and 29 °C in Beijing, Hong Kong, and Singapore, respectively ($P<.01$ for all; Multimedia Appendix 4). While the magnitude of the results may look small relative to the base rate of case accrual, these plots were on the log-case scale, so the effect on the case number is multiplicative. The distributions of COVID-19 cases displayed greater uncertainty at a lower temperature in Beijing and Wuhan. Beijing and Wuhan have fluctuating patterns throughout the 6 models both regarding temperature and relative humidity except in the lag 5d and lag 4d models for Wuhan, which may be due to the change in diagnostic method on February 12, when a total of 13,436 cases were added. Nevertheless, the relationship between relative humidity and new cases were less evident in the no-lag model, and the distributions of COVID-19 cases displayed greater uncertainty in Wuhan.

Table 4. The selection of generalized additive modeling by Akaike information criterion (AIC).

City and model	*df*	AIC
Beijing, China		
No lag	24.01	265.47
Lag 1d[a]	21.71	226.47
Lag 3d	17.72	273.59
Lag 5d	21.46	263.25
Lag 7d	23.38	243.26
Lag 14	19.79	230.50
Shanghai, China		
No lag[a]	7.91	178.01
Lag 1d	7.50	181.60
Lag 3d	7.04	180.93
Lag 5d	7.08	180.60
Lag 7d	6.83	182.73
Lag 14	10.51	178.06
Guangzhou, China		
No lag	14.17	199.85
Lag 1d	14.40	192.44
Lag 3d	18.32	194.49
Lag 5d[a]	20.97	192.35
Lag 7d	10.63	200.94
Lag 14	12.89	195.34
Wuhan, China		
No lag	27.95	4812.57
Lag 1d[a]	27.87	3637.25
Lag 3d	27.89	6778.17
Lag 5d	27.98	5344.86
Lag 7d	27.86	10,351.23
Lag 14	26.45	6735.71
Hong Kong, China		
No lag	12.46	193.98
Lag 1d	16.49	192.14
Lag 3d	11.92	195.65
Lag 5d	14.12	193.45
Lag 7d	13.17	199.01
Lag 14[a]	6.53	189.09
Singapore		
No lag[a]	10.94	183.22
Lag 1d	10.88	185.19
Lag 3d	9.57	202.13
Lag 5d	10.83	207.79
Lag 7d	16.18	230.09

City and model	*df*	AIC
Lag 14	14.83	277.87
Seoul, Korea		
No lag	27.54	461.16
Lag 1d	25.86	381.98
Lag 3d	27.19	439.14
Lag 5d[a]	26.67	363.42
Lag 7d	26.24	420.27
Lag 14	27.29	452.58
Tokyo, Japan		
No lag	24.58	264.80
Lag 1d	19.77	266.44
Lag 3d	24.37	276.50
Lag 5d	18.28	272.25
Lag 7d	17.61	288.62
Lag 14[a]	18.33	264.45
Kuala Lumpur, Malaysia		
No lag	16.37	142.50
Lag 1d	16.74	138.72
Lag 3d	21.70	132.95
Lag 5d[a]	17.93	125.05
Lag 7d	19.80	132.67
Lag 14	21.58	133.89

[a]Denotes the model with the lowest AIC value.

Sensitivity Analysis of Daily New COVID-19 Cases With Meteorological Factors

After taking the city level into consideration in the model, a GAM model of daily new COVID-19 cases was built, incorporating date (time-series), city level, average temperature, and mean relative humidity, where the date was analyzed as spline functions.

Overall, we found a significant association between meteorological factors and daily new cases of COVID-19 in 9 Asian cities. Under GAM, high temperature tended to increase the number of daily new cases of COVID-19 whereas high relative humidity decreased the count (Table 5). Relative humidity does not influence the daily new cases significantly, but higher temperatures can exert an increase as high as 7% on daily new cases in lag 3d (beta=0.073, SE 0.006, *P*<.001; Table 4). The time-series analysis revealed that notwithstanding the lagged time effects, the overall influence of temperature was steadily positive. Nevertheless, the relative humidity exerted a negative influence on the transmission of daily new cases of COVID-19 (Table 5).

Table 5. Generalized linear modeling of the effects of temperature and relative humidity on daily new cases of COVID-19.

Model	Estimate	Standard error	Z value	*P* value
No-lag model				
Temperature	0.062	0.006	10.861	<.001
Relative humidity	–0.018	0.001	–15.105	<.001
Lag 1d model				
Temperature	0.015	0.007	2.284	.02
Relative humidity	–0.007	0.001	–6.513	<.001
Lag 3d model				
Temperature	0.073	0.006	11.594	<.001
Relative humidity	–0.006	0.001	–4.501	<.001
Lag 5d model				
Temperature	–0.026	0.006	–4.152	<.001
Relative humidity	0.003	0.001	2.471	.01
Lag 7d model				
Temperature	0.063	0.005	13.887	<.001
Relative humidity	–0.008	0.001	–6.620	<.001

Discussion

Principal Findings

In this study, we investigated the associations between meteorological factors and patterns of daily new cases of COVID-19 across 9 Asian cities. The city-specific GAM analysis revealed a positive relationship between temperature and daily new cases of COVID-19 in Guangzhou, Singapore (except in the lagged 14-day model), Hong Kong (except in the lagged 7-day and 14-day model), and Beijing (high curvilinearity). Relative humidity positively associated with the number of daily new cases in Singapore (except in the lagged 14-day model), Hong Kong (except in the lagged 3-day model). Moreover, the sensitivity test using GAM with linear components revealed that high temperature significantly increases the daily new cases of COVID-19, while high relative humidity significantly reduced, to a lower extent, the daily new cases of COVID-19. Therefore, our analysis suggests, unlike influenza, seasonality of COVID-19 may not be expected, and the pandemic is unlikely to diminish during warmer seasons (ie, summer).

Researchers have long been investigating how meteorological factors affect the viral infectivity, where GAM has been frequently used, as it allows smooth components to be estimated for time, meteorological factors, and other covariates, together with a nonsmoothed period effect. Experiments from the mid-20th century reported that the influenza virus is more stable in cool and dry air [15,16]. With increasing temperature, the viability of the influenza virus in aerosol or droplets [43] and the aerosol transmission diminishes [17]. Lowen et al [19] reported that aerosol transmission of influenza between guinea pigs was completely blocked at temperature higher than 30 °C despite evidence of continuous viral shedding from infectious individuals; nevertheless, direct contact transmission was not affected, which was equally efficient at 30 °C and 20 °C. Chan et al [30] reported that the viability of the SARS virus was rapidly lost (>3 log 10) at high temperatures (38 °C) and high relative humidity (>95%). The better stability of the SARS coronavirus at low temperatures and low humidity environment might facilitate its transmission in the community in subtropical areas (such as Hong Kong) during the spring and in air-conditioned environments. This might also explain why some Asian countries in tropical areas (such as Malaysia, Indonesia, or Thailand) with high temperatures and high relative humidity environment did not have major community outbreaks of SARS. However, such an explanation is not currently convincing in light of the results of the present study, where higher temperatures were associated with increases in the daily new cases of COVID-19 in some of the investigated regions. The high temperature and high relative humidity in tropical Asian countries like Singapore and Malaysia also seem to have little influence on the growing number of daily new cases (Multimedia Appendix 4). However, this could have been confounded by multiple factors.

There are several reasons underlying the continuous growth in COVID-19 cases in Singapore and Malaysia, such as the high population density, mass gatherings, use of air-conditioning, and shortage of medical resources [30]. SARS-CoV-2 can persist at room temperature for up to 9 days, and its heat sensitivity renders it susceptible to increased temperature, affecting its persistence in the outdoor environment. However, this coronavirus was still found to be infective up to 2 weeks in an air-conditioned environment [30]. As air-conditioners may increase the probability of viral spread, it may be advisable to reduce the use of air-conditioners and keep areas well ventilated [44]. Moreover, during low temperature and high humidity, it is advisable to avoid mass gatherings, since there is evidence supporting transmission by direct contact or close contact in tropical areas.

Humidity can influence aerosol transmission by altering the proportion of respiratory droplets undergoing aerosolization and influencing the stability and viability of the virus within these aerosols. Respiratory droplets are generated in the high humidity setting of the respiratory tract. Upon entering an environment with low humidity, respiratory droplets reduce in size within seconds due to evaporation. At higher environmental humidity, respiratory droplets evaporate more slowly, and hence are larger and settle faster, and less aerosol nuclei are produced [45,46]. Previous studies have shown that influenza transmission in mice decreased as relative humidity increased from 47% to 70% [13,47]. Moreover, humidity can also influence indirect transmission by changing the mass of respiratory droplets accumulating on surfaces and affecting the survival of the virus on surfaces. While increased humidity reduces the number of droplet nuclei formed, the same mechanisms (reduced droplet evaporation and faster droplet settling) result in a greater mass of respiratory droplets on surfaces [45,46]. Areas with relatively low temperature and humidity have a higher infection rate compared to tropical areas since cold and dry weather is suitable for viral survival and transmission [48]. The viability of the influenza virus appears greater at lower humidity, and exhibits progressively reduced survival with increasing relative humidity over the 27%-84% range, with an increase in survival at 99% relative humidity. The mechanisms underlying this may be that the reduced evaporation of droplets at high relative humidity maintain the solute concentration, thus protecting the virus [17,49]. In addition, the 30° N to 50° N latitudes have become a zone for COVID-19 transmission with a similar average temperature between 5 °C and 11 °C and 47%-79% humidity, which may be influenced by the transoceanic migration of the virus, but the underlying mechanism is still not understood [50].

Moreover, under laboratory conditions with constant humidity and temperature, circadian and circannual rhythms in variation of susceptibility of hosts (mice) have been observed [13,51]. Mice were substantially more susceptible to invasive pneumococcal disease in early morning hours than any other time of day [51], and more susceptible to influenza in winter than in summer [13], which was thought to be attributable to the daily and seasonal variation of melatonin [52]. In addition, even in areas where many spend summers in air-conditioned spaces, marked annual variations in incidence constantly exist, and similar strains of the virus appear almost simultaneously across vast stretches of ocean in areas of similar latitude around the globe [23,24].

Therefore, although a higher temperature is associated with lower effectiveness of virus transmission, yet it does not necessarily suggest a reduced chance for virus survival. The natural fading out of the virus in the summer is unlikely given the widespread use of air-conditioners in developed areas and dense populations in cities. Until an effective vaccine becomes widely available for the establishment of herd immunity, alongside efficient pharmaceutical therapies, strict public health measures should be implemented, including social distancing, quarantine, contact tracing, face mask wearing, and hand washing.

Strengths and Limitations

This is the first study to statistically analyze the relationship between meteorological factors and the daily new cases of COVID-19. Because of the nonlinear nature of the data, we also performed GAMs to quantitate and visualize the relationship using spline functions. However, there are several limitations. First, the number of cases in Malaysia, Korea, and Japan were obtained from the epidemiologic reports released by the Department of Health in the corresponding countries, instead of a daily update of case numbers, which was not available for these countries. Hence, some cases may be missing. Second, the duration of the study period was short and the number of cases in Malaysia, Singapore, and Japan were small. Third, we only considered two meteorological factors (temperature and humidity) in this study. Other covariates such as wind speed, pollutant concentration, population density, air-conditioning use, and rainfall, which could also influence the spread of COVID-19, were not included. Moreover, while we collected the meteorological data from the capital cities of Malaysia, Japan, and Korea, the number of domestic cases at the national level in these three countries was used for analysis due to lack of available data at their city level. Most importantly, we did not incorporate public health measures into the modeling, which may greatly confound the results, but we chose the date when accumulated confirmed cases exceeded 30, based on the postulation that a certain level of public health measures had been carried out at that time; thus, the confounding effects can be mitigated at some point.

Conclusions

In this study, we found high temperature to be associated with daily new cases of COVID-19. Therefore, unlike influenza, seasonality in COVID-19 prevalence may not be expected, and the pandemic is unlikely to decrease in numbers during the warmer seasons. Strict public health measures such as social distancing, quarantine, contact tracing, face mask wearing, and hand washing are needed until a vaccine becomes widely available to induce herd immunity.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The effects of temperature and relative humidity on daily new cases of COVID-19.

[DOCX File , 29 KB - publichealth_v7i1e20495_app1.docx]

Multimedia Appendix 2

Daily temperature and distribution of daily new cases of COVID-19 over time.
[PNG File , 919 KB - publichealth_v7i1e20495_app2.png]

Multimedia Appendix 3
Daily relative humidity and distribution of daily new COVID-19 cases over time.
[PNG File , 861 KB - publichealth_v7i1e20495_app3.png]

Multimedia Appendix 4
Smoothing components plots for daily new COVID-19 cases associated with average temperature.
[PNG File , 1329 KB - publichealth_v7i1e20495_app4.png]

Multimedia Appendix 5
Smoothing components plots for daily new COVID-19 cases associated with average relative humidity.
[PNG File , 1278 KB - publichealth_v7i1e20495_app5.png]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7]
2. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization. 2020 Jan 30. URL: https://tinyurl.com/rjdtx2k [accessed 2020-02-24]
3. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci 2020 Mar;63(3):457-460 [FREE Full text] [doi: 10.1007/s11427-020-1637-5] [Medline: 32009228]
4. Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. Viruses 2020 Jan 24;12(2):135 [FREE Full text] [doi: 10.3390/v12020135] [Medline: 31991541]
5. Cowling B, Leung G. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. Euro Surveill 2020 Feb;25(6):2020 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.6.2000110] [Medline: 32046814]
6. Tang J. The emergence and spread of the 2019 novel coronavirus (2019-nCoV). Infectious Diseases Hub. 2020. URL: https://www.id-hub.com/2020/02/10/the-emergence-and-spread-of-the-2019-novel-coronavirus-2019-ncov/ [accessed 2020-02-24]
7. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill 2020 Jan;25(4) [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.4.2000058] [Medline: 32019669]
8. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine 2020 May 05;172(9):577-582. [doi: 10.7326/m20-0504]
9. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
10. WHO Director-General's opening remarks at the Mission briefing on COVID-19 - 12 March 2020. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19---12-march-2020 [accessed 2020-03-15]
11. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016 Aug 27;14(8):523-534 [FREE Full text] [doi: 10.1038/nrmicro.2016.81] [Medline: 27344959]
12. Cai Q, Lu J, Xu Q, Guo Q, Xu D, Sun Q, et al. Influence of meteorological factors and air pollution on the outbreak of severe acute respiratory syndrome. Public Health 2007 Apr;121(4):258-265 [FREE Full text] [doi: 10.1016/j.puhe.2006.09.023] [Medline: 17307207]
13. Schulman J, Kilbourne E. Experimental Transmission of Influenza Virus Infection in Mice. II. Some Factors Affecting the Incidence of Transmitted Infection. J Exp Med 1963;118(2):267-275. [doi: 10.1084/jem.118.2.267]
14. Horby P, Nguyen N, Dunstan S, Baillie JK. The role of host genetics in susceptibility to influenza: a systematic review. PLoS One 2012;7(3):e33180 [FREE Full text] [doi: 10.1371/journal.pone.0033180] [Medline: 22438897]
15. Hemmes JH, Winkler KC, Kool SM. Virus Survival as a Seasonal Factor in Influenza and Poliomyelitis. Nature 1960 Oct;188(4748):430-431. [doi: 10.1038/188430a0]

16. Loosli CG, Lemon HM, Robertson OH, Appel E. Experimental Air-Borne Influenza Infection. I. Influence of Humidity on Survival of Virus in Air. Experimental Biology and Medicine 1943 Jun 01;53(2):205-206. [doi: 10.3181/00379727-53-14251p]
17. Paynter S. Humidity and respiratory virus transmission in tropical and temperate settings. Epidemiol Infect 2015 Apr;143(6):1110-1118. [doi: 10.1017/S0950268814002702] [Medline: 25307020]
18. Finkelman BS, Viboud C, Koelle K, Ferrari MJ, Bharti N, Grenfell BT. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. PLoS One 2007 Dec 12;2(12):e1296 [FREE Full text] [doi: 10.1371/journal.pone.0001296] [Medline: 18074020]
19. Lowen AC, Steel J, Mubareka S, Palese P. High Temperature (30°C) Blocks Aerosol but Not Contact Transmission of Influenza Virus. JVI 2008 Jun 01;82(11):5650-5652. [doi: 10.1128/jvi.00325-08]
20. Coronavirus disease 2019 (COVID-19) Situation Report – 28. World Health Organization. 2020 Feb 17. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200217-sitrep-28-covid-19.pdf?sfvrsn=a19cf2ad_2 [accessed 2020-02-24]
21. Addie D, Schaap I, Nicolson L, Jarrett O. Persistence and transmission of natural type I feline coronavirus infection. J Gen Virol 2003 Oct;84(Pt 10):2735-2744. [doi: 10.1099/vir.0.19129-0] [Medline: 13679608]
22. Arbour N, Côté G, Lachance C, Tardieu M, Cashman NR, Talbot PJ. Acute and Persistent Infection of Human Neural Cell Lines by Human Coronavirus OC43. J. Virol 1999 Apr 01;73(4):3338-3350. [doi: 10.1128/jvi.73.4.3338-3350.1999]
23. Langmuir AD, Schoenbaum SC. The epidemiology of influenza. Hosp Pract 1976 Oct 06;11(10):49-56. [doi: 10.1080/21548331.1976.11707011] [Medline: 67988]
24. Thacker SB. The persistence of influenza A in human populations. Epidemiol Rev 1986;8(1):129-142. [doi: 10.1093/oxfordjournals.epirev.a036291] [Medline: 3533582]
25. Hope-Simpson RE, Golubev DB. A new concept of the epidemic process of influenza A virus. Epidemiol Infect 1987 Aug 19;99(1):5-54 [FREE Full text] [doi: 10.1017/s0950268800066851] [Medline: 3301379]
26. Hammond GW, Raddatz RL, Gelskey DE. Impact of atmospheric dispersion and transport of viral aerosols on the epidemiology of influenza. Rev Infect Dis 1989 May 01;11(3):494-497. [doi: 10.1093/clinids/11.3.494] [Medline: 2665004]
27. Tamerius J, Nelson MI, Zhou SZ, Viboud C, Miller MA, Alonso WJ. Global influenza seasonality: reconciling patterns across temperate and tropical regions. Environ Health Perspect 2011 Apr;119(4):439-445 [FREE Full text] [doi: 10.1289/ehp.1002383] [Medline: 21097384]
28. Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN. Influenza Seasonality: Underlying Causes and Modeling Theories. JVI 2007 Jun 01;81(11):5429-5436. [doi: 10.1128/jvi.01680-06]
29. Tamerius JD, Shaman J, Alonso WJ, Alonso WJ, Bloom-Feshbach K, Uejio CK, et al. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. PLoS Pathog 2013 Mar 7;9(3):e1003194 [FREE Full text] [doi: 10.1371/journal.ppat.1003194] [Medline: 23505366]
30. Chan KH, Peiris JSM, Lam SY, Poon LLM, Yuen KY, Seto WH. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. Adv Virol 2011;2011:734690-734697 [FREE Full text] [doi: 10.1155/2011/734690] [Medline: 22312351]
31. National Meteorological Information Center. URL: https://data.cma.cn/en [accessed 2020-05-15]
32. Ministry of Ecology. URL: http://english.mee.gov.cn/ [accessed 2020-05-15]
33. China Meteorological Administration. URL: http://www.cma.gov.cn/en2014/ [accessed 2020-05-15]
34. Timeanddate.com. URL: https://www.timeanddate.com/ [accessed 2020-05-15]
35. Ministry of Health in China. URL: http://www.nhc.gov.cn/ [accessed 2020-12-08]
36. Ministry of Health Singapore. URL: https://www.moh.gov.sg/ [accessed 2020-12-08]
37. Hastie TJ. Generalized additive models for medical research. In: Hastie TJ, Tibshirani RJ, editors. Generalized Additive Models. Roca Baton, FL: Chapman & Hall/CRC; 1995:187-196.
38. Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, Monopolis Y, et al. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology 2001 Sep;12(5):521-531. [doi: 10.1097/00001648-200109000-00011] [Medline: 11505171]
39. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. J Royal Statistical Soc A 2006 Mar;169(2):179-203. [doi: 10.1111/j.1467-985x.2006.00410.x]
40. Kan H, London SJ, Chen G, Zhang Y, Song G, Zhao N, et al. Differentiating the effects of fine and coarse particles on daily mortality in Shanghai, China. Environ Int 2007 Apr;33(3):376-384 [FREE Full text] [doi: 10.1016/j.envint.2006.12.001] [Medline: 17229464]
41. Huang Y, Deng T, Yu S, Gu J, Huang C, Xiao G, et al. Effect of meteorological variables on the incidence of hand, foot, and mouth disease in children: a time-series analysis in Guangzhou, China. BMC Infect Dis 2013 Mar 13;13(1):134 [FREE Full text] [doi: 10.1186/1471-2334-13-134] [Medline: 23497074]
42. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5]

43. Dublineau A, Batéjat C, Pinon A, Burguière AM, Leclercq I, Manuguerra J. Persistence of the 2009 pandemic influenza A (H1N1) virus in water and on non-porous surface. PLoS One 2011 Nov 23;6(11):e28043 [FREE Full text] [doi: 10.1371/journal.pone.0028043] [Medline: 22132205]
44. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
45. Xie X, Li Y, Chwang ATY, Ho PL, Seto WH. How far droplets can move in indoor environments--revisiting the Wells evaporation-falling curve. Indoor Air 2007 Jun;17(3):211-225. [doi: 10.1111/j.1600-0668.2007.00469.x] [Medline: 17542834]
46. Yang W, Marr LC. Dynamics of airborne influenza A viruses indoors and dependence on humidity. PLoS One 2011 Jun 24;6(6):e21481 [FREE Full text] [doi: 10.1371/journal.pone.0021481] [Medline: 21731764]
47. Schulman JL, Kilbourne ED. Airborne Transmission of Influenza Virus Infection in Mice. Nature 1962 Sep;195(4846):1129-1130. [doi: 10.1038/1951129a0]
48. To K, Lo AW. Exploring the pathogenesis of severe acute respiratory syndrome (SARS): the tissue distribution of the coronavirus (SARS-CoV) and its putative receptor, angiotensin-converting enzyme 2 (ACE2). J Pathol 2004 Jul 21;203(3):740-743 [FREE Full text] [doi: 10.1002/path.1597] [Medline: 15221932]
49. Yang W, Elankumaran S, Marr LC. Relationship between humidity and influenza A viability in droplets and implications for influenza's seasonality. PLoS One 2012 Oct 3;7(10):e46789 [FREE Full text] [doi: 10.1371/journal.pone.0046789] [Medline: 23056454]
50. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. SSRN 2020 Mar 09:3550308. [doi: 10.2139/ssrn.3550308] [Medline: 32714105]
51. Feigin RD, San Joaquin VH, Haymond MW, Wyatt RG. Daily periodicity of susceptibility of mice to pneumococcal infection. Nature 1969 Oct 25;224(5217):379-380. [doi: 10.1038/224379a0] [Medline: 5343888]
52. Nelson RJ, Drazen DL. Melatonin mediates seasonal adjustments in immune function. Reprod. Nutr. Dev 1999;39(3):383-398. [doi: 10.1051/rnd:19990310]

Abbreviations

AIC: Akaike information criterion
df: degrees of freedom
GAM: generalized additive modeling
MERS-CoV: Middle East respiratory syndrome coronavirus
PACF: partial autocorrelation function
RNA: ribonucleic acid
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by R Ho, R Poluru; comments to author 29.06.20; revised version received 12.07.20; accepted 24.10.20; published 25.01.21.

Please cite as:
He Z, Chin Y, Yu S, Huang J, Zhang CJP, Zhu K, Azarakhsh N, Sheng J, He Y, Jayavanth P, Liu Q, Akinwunmi BO, Ming WK
The Influence of Average Temperature and Relative Humidity on New Cases of COVID-19: Time-Series Analysis
JMIR Public Health Surveill 2021;7(1):e20495
URL: https://publichealth.jmir.org/2021/1/e20495
doi:10.2196/20495
PMID:33232262

Original Paper

Predictors of COVID-19 Information Sources and Their Perceived Accuracy in Nigeria: Online Cross-sectional Study

Olufemi Erinoso[1], MPH, BDS; Kikelomo Ololade Wright[2], MPH, MD; Samuel Anya[3], MD, PhD; Yetunde Kuyinu[2], MPH, MD; Hussein Abdur-Razzaq[3], MPH, MD; Abiodun Adewuya[4], MSc, MD

[1]Department of Oral and Maxillofacial Surgery, Lagos State University Teaching Hospital, Lagos, Nigeria

[2]Department of Community Health & Primary Health Care, Lagos State University College of Medicine, Lagos State University Teaching Hospital, Lagos, Nigeria

[3]Research Unit, Lagos State Ministry of Health, Lagos, Nigeria

[4]Department of Behavioural Medicine, Lagos State University College of Medicine, Lagos State University Teaching Hospital, Lagos, Nigeria

Corresponding Author:
Olufemi Erinoso, MPH, BDS
Department of Oral and Maxillofacial Surgery
Lagos State University Teaching Hospital
1-5b Oba Akinjobi Road
GRA Ikeja
Lagos, 23401
Nigeria
Phone: 234 1 8950825
Email: olufemierinoso@gmail.com

Abstract

Background: Effective communication is critical for mitigating the public health risks associated with the COVID-19 pandemic.

Objective: This study assesses the source(s) of COVID-19 information among people in Nigeria, as well as the predictors and the perceived accuracy of information from these sources.

Methods: We conducted an online survey of consenting adults residing in Nigeria between April and May 2020 during the lockdown and first wave of COVID-19. The major sources of information about COVID-19 were distilled from 7 potential sources (family and friends, places of worship, health care providers, internet, workplace, traditional media, and public posters/banners). An open-ended question was asked to explore how respondents determined accuracy of information. Statistical analysis was conducted using STATA 15.0 software (StataCorp Texas) with significance placed at $P<.05$. Approval to conduct this study was obtained from the Lagos State University Teaching Hospital Health Research Ethics Committee.

Results: A total of 719 respondents completed the survey. Most respondents (n=642, 89.3%) obtained COVID-19–related information from the internet. The majority (n=617, 85.8%) considered their source(s) of information to be accurate, and 32.6% (n=234) depended on only 1 out of the 7 potential sources of COVID-19 information. Respondents earning a monthly income between NGN 70,000-120,000 had lower odds of obtaining COVID-19 information from the internet compared to respondents earning less than NGN 20,000 (odds ratio [OR] 0.49, 95% CI 0.24-0.98). In addition, a significant proportion of respondents sought accurate information from recognized health organizations, such as the Nigeria Centre for Disease Control and the World Health Organization.

Conclusions: The internet was the most common source of COVID-19 information, and the population sampled had a relatively high level of perceived accuracy for the COVID-19 information received. Effective communication requires dissemination of information via credible communication channels, as identified from this study. This can be potentially beneficial for risk communication to control the pandemic.

(JMIR Public Health Surveill 2021;7(1):e22273) doi:10.2196/22273

KEYWORDS

COVID-19; communication; health information; public health; infodemiology; infodemic; accuracy; cross-sectional; risk; information source; predictor; Nigeria

Introduction

COVID-19, caused by the virus SARS-CoV-2, was first identified in Wuhan, Hubei Province, China, in December 2019 [1], and was thereafter recognized as a pandemic by the World Health Organization (WHO) on March 11, 2020 [2]. As of July 6, 2020, more than 11 million cases and 537,419 deaths related to COVID-19 have been reported in 213 countries [3]. In Africa, over 470,000 cases have been recorded with more than 29,000 cases from Nigeria [3]. The virus is typically spread from one person to another via respiratory droplets and contact with contaminated surfaces [4].

COVID-19 has led to unprecedented local and global public health measures, such as obligatory movement restrictions, social and physical distancing, and prolonged closures of schools and leisure centers. Guidelines on risk factors and preventive measures have emerged in Nigeria from state health agencies, the Nigeria Centre for Disease Control (NCDC), and the Federal Ministry of Health [5]. As our understanding of SARS-CoV-2 increases, local COVID-19 interventions are beginning to focus on the sources and perceived accuracy of disseminated preventive information, which portends an important process in the prevention and control of the disease [6].

Potential sources of information during disease outbreaks include the internet and social media platforms (Facebook, WhatsApp, Instagram, and Twitter), traditional media (television, radio and newspapers), places of worship, health care providers, family members and friends, and workplaces. Therefore, due to the varied sources of information on COVID-19, identifying factors related to such sources can support public educational interventions [7].

In a survey conducted in the Nigerian population during the Ebola virus disease (EVD) outbreak in 2014 [8], the majority of respondents depended on traditional media for information on the disease, while less than one-third depended on the internet for their information. This may suggest that perception and trust of information source are intertwined and contribute to use [8].

As the Nigerian government has begun easing COVID-19 lockdown measures, effective communication is an essential component for mitigating the risks associated with the inevitable clustering of people in public places and other practices capable of not only fueling the spread of the disease but spiking the number of cases and mortality from COVID-19 [9]. Therefore, effectively communicating the efficacy of practical interventions such as personal hygiene, hand washing, use of face masks, and social distancing among other strategies may help curb transmission. Consequently, identifying common sources of COVID-19 information among the population, and the perceived accuracy of these information sources, will possibly guide risk communication processes and the dissemination of evidence-based COVID-19 public health information. The dissemination of this information can be channeled to the most commonly used sources with the highest levels of perceived accuracy.

This study aims to identify the sources of COVID-19 information among Nigerians, as well as the predictors and the perceived accuracy of these sources. Findings from this study will support policymakers in disseminating targeted evidence-based anti–COVID-19 information to populations at risk and those affected. This can invariably empower the public with the capacity to make informed health decisions and improve health outcomes.

Methods

This study is part of an online survey descriptive cross-sectional survey on COVID-19 conducted by the Lagos State University College of Medicine between April 22 and May 20, 2020. The survey was conducted to assess psychological distress, adherence, and sources of information during the COVID-19 outbreak.

Study Design and Population

Study participants were consenting individuals aged 18 years or above, residing in Nigeria at the time of the study. A multistage sampling technique was used in selecting study participants. First, 3 states were purposively selected from the 36 states of Nigeria because at the time of study design, they accounted for the highest number of COVID 19 cases in the country [10]. Subsequently, a sample frame of local community networks (estate associations, local organizations, schools, and religious organization groups) within Lagos State, Ogun State, and Federal Capital Territory Abuja was obtained by the Lagos State Ministry of Health (LSMoH) research office using cross-country partners. A list of networks was selected using a simple random sampling method. Leaders of these community networks were identified by the LSMoH research office and contacted via phone and email. Permission was obtained from community network leaders before the online survey link was shared with all members of their respective groups. In total, 9 leaders were approached, and 7 provided consent to share the link to the survey via email and on the social media pages (WhatsApp and Twitter) of their groups. Subsequently, consenting participants on these groups indicated individual consent by clicking the “I consent” button on the first page of the online form before proceeding to answer the survey questions. This data collection approach was used due to the movement restrictions and social/physical distancing measures enforced by the federal government of Nigeria to curb COVID-19 transmission.

Assessment Tools

Data collection was conducted using an online survey tool, SurveyMonkey (SurveyMonkey Inc). Completing the survey took an average of 7 minutes.

Study Measures

Sociodemographic Variables

Information was obtained on respondents’ sex (male or female), age, marital status, education, and income level (in naira, NGN). Marital status was grouped as single, married, and previously married (ie, widow, widower, divorced, and separated). The educational level of respondents was categorized in three groups: high school or less, university or polytechnic, and postgraduate education. Income level was categorized into three groups based on earnings per month: less than NGN 20,000 (<US $52); NGN 20,000-70,000 (US $52-$181); NGN 70,000-120,000 (US $181-$310); and greater than NGN 120,000 (>US $310).

Sources of COVID-19 Information

The study categorized COVID-19 information sources into the following broad groups: (1) family members and friends; (2) places of worship; (3) health care providers (ie, doctors, nurses, pharmacists); (4) the internet, comprising three broad groups: (a) social media (Facebook, WhatsApp, Instagram, and Twitter), (b) news websites, and (c) non–news-related, non–social media websites (eg, blogs, websites of health regulatory organizations such as the WHO and the NCDC); (5) workplace; (6) traditional media (television, radio, or newspapers); and (7) public posters and banners. The total number of sources consulted was determined for each respondent.

Perceived Accuracy of COVID-19 Information Source(s)

The perceived accuracy of COVID-19 information was assessed by asking the question "Do you think your source of information is accurate?" to which respondents were given the response options *Yes*, *No*, or *I don't know*. An open-ended question asking "How do you differentiate between accurate and inaccurate COVID-19 information?" was used to explore how respondents determined the accuracy of COVID-19 information (see survey questions in Multimedia Appendix 1).

Statistical Analysis

Responses from the survey tool were automatically converted to variables using Microsoft Excel (Microsoft Corp). Demographic data, source(s), perceived accuracy, and the total number of COVID-19 information sources were expressed using descriptive statistics. The association between sociodemographic variables, source(s), perceived accuracy, and the total number of COVID-19 information sources was investigated using bivariate and multivariate logistic regression analysis. Multiple linear regressions were also used to assess the association between the total number of information sources and factors related to COVID-19 information (age, sex, marital status, education, income, and perceived accuracy). *P* values <.05 were considered significant, and tests were two-tailed. Statistical analysis was done using STATA 15.0 software (StataCorp LLC). Qualitative analysis using R (The R Foundation) was used to explore themes and subthemes explaining the determinants of perceived accuracy of COVID-19 information.

Ethics

Ethics approval was obtained from the Health Research Ethics Committee of the Lagos State University Teaching Hospital (number LREC/06/10/1347).

Data Sharing

The data sets used and/or analyzed during the study are available from the corresponding author upon reasonable request.

Results

Overview

A total of 719 respondents completed the online survey (94.1% response rate); 45 respondents were excluded due to incomplete responses. The mean age of the respondents was 26.9 years (SD 8.8, range 15-69 years). Respondents aged less than 35 years accounted for 88% (n=633), and females made up 54% (n=390) of all study participants. The majority of respondents were single, with a university/polytechnic education and a monthly income of less than NGN 20,000 (Table 1).

Table 1. Sociodemographic factors and sources of information.

Variable	Respondents, n (%)
Age (years)	
<35	633 (88.0)
≥35	86 (12.0)
Sex	
Male	329 (45.8)
Female	390 (54.2)
Marital status	
Single	571 (79.4)
Married	141 (19.6)
Previously married	7 (1.0)
Education	
Secondary school or less	75 (10.4)
University/polytechnic	516 (71.8)
Postgraduate	128 (17.8)
Income (NGN)	
<20,000	288 (40.1)
20,000-70,000	214 (29.8)
>70,000-120,000	86 (11.9)
>120,000	131 (18.2)
Sources of information	
Family and friends	269 (37.4)
Places of worship	86 (12.0)
Health care providers	210 (29.2)
Internet	642 (89.3)
Workplace	90 (12.5)
Traditional media	452 (62.9)
Public posters and banners	89 (12.4)
Total count of information sources	
1	234 (32.6)
2	149 (20.7)
3	172 (23.9)
4	88 (12.2)
5	37 (5.2)
6	20 (2.8)
7	19 (2.6)
Perceived accuracy of information	
No	102 (14.2)
Yes	617 (85.8)

Sources of COVID-19 Information

Sources of COVID-19 information among respondents ranged from family and friends to public posters and banners. The most common source of information was the internet (n=642, 89.3%), which comprised news via social media handles, websites, blogs, and social media. This was followed by traditional media comprising television, radio, and newspaper/print resources, as reported by 62.9% (n=452) of all respondents. The least common source of information was places of worship (n=86, 12.0%). Out of a total of 7 options in the survey, most respondents used only one source of information (n=234, 32.6%) (Table 1). The sociodemographic characteristics of respondents using each source type is illustrated in Table S1 in Multimedia Appendix 1.

Family and Friends

More than one-third of respondents (n=269, 37.4%) obtained COVID-19 information from family and friends (Table 1). Bivariate logistic regressions showed that married respondents had lower odds of obtaining COVID-19 information from family and friends compared to single respondents (odds ratio [OR] 0.66, 95% CI 0.44-0.98) (Table S2 in Multimedia Appendix 1). No significant association was obtained between age, sex, education, income status, perceived accuracy, and family and friends as a source of COVID-19 information.

Places of Religious Worship

Only 12% (n=86) of respondents obtained COVID-19–related information from their respective places of worship (Table 1). There was no significant association between sociodemographic factors (age, sex, marital status, education, and income) and perceived accuracy with place of worship as a source of COVID-19 information (Tables S2 and S3 in Multimedia Appendix 1).

Health Care Providers

In total, 29% (n=210) of respondents obtained COVID-19 information from their health care providers (Table 1). Married respondents had lower odds of obtaining COVID-19 information from health care providers compared to single respondents (OR 0.60, 95% CI 0.38-0.93). In addition, respondents with a university or postgraduate degree had lower odds of obtaining COVID-19 information from health care providers compared to respondents with a secondary school degree or less (Table S2 in Multimedia Appendix 1). Respondents earning a monthly income of NGN 70,000-120,000 had 1.75 higher odds of obtaining COVID-19 information from health care providers compared to respondents earning less than NGN 20,000 (95% CI 1.05-2.91).

Table S3 in Multimedia Appendix 1 shows multivariate logistic regressions assessing the effect of sociodemographic factors on information source. Being married and having a university/polytechnic education was significantly associated with reduced odds compared to being single and with a secondary school education or less, respectively. In addition, respondents with a monthly income above NGN 70,000 had significantly higher odds of obtaining COVID-19 information from health care providers compared to respondents earning less than NGN 20,000 (adjusted OR [aOR] 2.21, 95% CI 1.28-3.80).

Internet

Most respondents (n=642, 89.3%) obtained COVID-19 information from the internet (Table 1). Respondents earning a monthly income between NGN 70,000-120,000 had lower odds of obtaining COVID-19 information from the internet compared to respondents earning less than NGN 20,000 (OR 0.49, 95% CI 0.24-0.98) (Table S2 in Multimedia Appendix 1). However, after adjusting for sociodemographic factors (age, sex, marital status, education) and perceived accuracy, there was no significant association between income levels and use of the internet as a source of COVID-19 information (aOR 0.55, 95% CI 0.26-1.16) (Table S3 in Multimedia Appendix 1).

Workplace

COVID-19 information from one's workplace was significantly associated with age, marital status, education, and monthly income. For every 1-year increase in age, there was a 1.04 increase in the odds of obtaining COVID-19 information from the workplace (95% CI 1.02-1.06). Similarly, married respondents had 1.91 higher odds of obtaining COVID-19 information from the workplace compared to single respondents (95% CI 1.16-3.13). Postgraduate education (OR 6.72, 95% CI 1.97-22.96) and higher income (>NGN 120,000; OR 16.78, 95% CI 7.89-35.68) were also significantly associated with the workplace as a source of COVID-19 information (Table S2 in Multimedia Appendix 1). Similarly, in the multivariate logistic regression model, income remained significantly associated with the workplace as a source of COVID-19 information. Respondents earning NGN 20,000 and above had significantly higher odds of obtaining COVID-19 information from their workplace compared to respondents earning less than NGN 20,000 (aOR 3.01, 95% CI 1.32-6.90) (Table S3 in Multimedia Appendix 1).

Traditional Media

About 62.9% (n=452) of respondents used traditional media as a source of COVID-19 information (Table 1). Older age was significantly associated with the use of traditional media as a source of COVID-19 information (OR 1.02, 95% CI 1.00-1.04). Married respondents also had 1.80 higher odds of using traditional media as a source of COVID-19 information (OR 1.80, 95% CI 1.20-2.71) compared to single respondents (Table S2 in Multimedia Appendix 1). In a multivariate logistic regression analysis, after adjusting for age, sex, education, income, and perceived accuracy, being married remained significantly associated with using traditional media as a source of COVID-19 information (aOR 1.83, 95% CI 1.04-3.25) (Table S3 in Multimedia Appendix 1).

Posters and Banners

About 12% (n=89) of the respondents surveyed reported that they obtained COVID-19 information from public posters and banners. Female respondents had 0.56 lower odds of using this source of information compared to males (95% CI 0.36-0.88). On the other hand, respondents with a postgraduate education had 2.75 higher odds of obtaining COVID-19 information from public posters and banners compared to respondents with a secondary school education or less (95% CI 0.99-7.63). In the multivariate logistic regression analysis, the use of public posters for obtaining COVID-19 information remained significantly associated with lower odds in females compared to male (aOR 0.55, 95% CI 0.34-0.88). Additionally, respondents with postgraduate education still had higher odds of obtaining information from posters and banners compared to secondary school or less (aOR 4.24, 95% CI 1.36-13.19) (Table S3 in Multimedia Appendix 1).

Association Between COVID-19 Information Sources and Perceived Accuracy

Table 2 shows a multiple linear regression model demonstrating that participants who earn greater than NGN 120,000 used 0.45 more information sources on average (β=.45, 95% CI 0.07-0.83), compared to those who earn less than NGN 20,000 per month. Therefore, respondents who earn more are more likely to seek information from various sources. Moreover, we devised a Poisson regression model with sources of information as the dependent variable, and sociodemographic factors and perception of accuracy as independent variables (Table S4, Multimedia Appendix 1). Compared to respondents earning less than NGN 20,000, the difference in the log count of the number of information sources increased by 0.16 for respondents earning NGN 120,000 or more, while holding age, sex, marital status, educational level, and perceived accuracy constant (95% CI 0.02-0.32). No significant association was seen between age, sex, marital status, education, and the number of COVID-19 information sources.

Table 2. Association between sociodemographic factors and the number of information sources using a multiple linear regression analysis.

Variable	β (95% CI)	*P* value
Age (years)	–0.00 (–0.02 to 0.02)	.99
Sex		
Male	1	
Female	–0.11 (–0.33 to 0.13)	.37
Marital status		
Single	1	
Married	–0.30 (–0.70 to 0.09)	.14
Previously married	–0.32 (–1.48 to 0.83)	.58
Education		
Secondary school or less	1	
University/polytechnic	–0.10 (–0.47 to 0.28)	.62
Postgraduate	0.73 (–0.42 to 0.57)	.77
Income (NGN)		
<20,000	1	
20,000-70,000	0.08 (–0.19 to 0.36)	.56
>70,000-120,000	0.19 (–0.19 to 0.57)	.33
>120,000	0.45 (0.07 to 0.83)	.02[a]
Perceived accuracy of information		
Inaccurate	1	
Accurate	0.06 (–0.26 to 0.38)	.73

[a]*P*<.05.

The majority (n=617, 85.8%) of respondents reported that their source(s) of information was accurate (Table 1). However, there was no statistically significant association between perceived accuracy and each source of information or the number of sources (Tables S2 and S3 in Multimedia Appendix 1; Table 2).

One major theme emerged by assessing respondents' means of differentiating between accurate and inaccurate COVID-19 information. Respondents (n=129) used information from recognized local and international health regulatory organizations (ie, reputable sources), such as the NCDC, the COVID-19 Presidential Task Force in Nigeria (PTF), the LSMoH, the WHO, and reference to a government or official agencies to determine accuracy. Some open-ended responses in this cohort of respondents include:

> *I rely on confirmed media accounts of Government agencies NCDC, LSMOH, COVID19 PTF.* [56-year-old married male]

> *I cross check with the verified source of information e.g. WHO, NCDC, LSMOH.* [25-year-old single female]

Among respondents who perceived their source(s) of information as accurate, 83 were confident of the accuracy of the information from the NCDC:

> *Any news different from the NCDC's is not always accurate.* [27-year-old married female]

In addition, for some respondents (n=38), information from the WHO website was deemed accurate. This could be inferred from statements like:

> *When it's from a verified source, like WHO, I know it's accurate.* [25-year-old single male]

Other ways of determining accuracy include cross-checking multiple sources (n=37):

> *When same news repeats itself in different places I see it as accurate.* [36-year-old married female]

In a subgroup analysis on the use of reputable sources of information, respondents earning between NGN 20,000-70,000 had 2.03 higher odds of using reputable sources compared to respondents earning less than NGN 20,000 (95% CI 1.21-3.38) (Table S5, Multimedia Appendix 1). In addition, Table S6 (Multimedia Appendix 1) distinguishes between the sociodemographic characteristics of respondents who used reputable sources of COVID-19 information and those who used other sources.

Discussion

Principal Findings

This study assessed common sources of COVID-19 information for a Nigerian study population during the early stages of the pandemic. Findings from this study suggest that most respondents used the internet as a source of COVID-19 information. Further, more than two-thirds of respondents considered their source(s) of information as accurate, and one-third depended on only 1 out of the 7 potential sources of COVID-19 information. However, there was no significant association between any of the potential sources of COVID-19 information and perceived accuracy. In addition, high-income earners (>NGN 120,000) had a greater likelihood of using more than one source of COVID-19 information.

The findings have shown that the internet was the most common source of information among the respondents. This contrasts with findings from the EVD outbreak in Nigeria 6 years ago, where traditional media (television and radio) were the main sources of EVD information at that time [11]. Similarly, a study in Vietnam listed mass media and peer educators as the most common sources of COVID-19 information as opposed to the internet [12]. Nonetheless, several studies in Nigeria and Asia have identified social media (a component of the internet information source) as an important source of information, particularly serving as the first source of information during the 2014 EVD outbreak in Nigeria [7,11-13]. Given the worldwide advancement in technology, the internet may be a common source of COVID-19 information now more than before because of the increased access to smartphones and internet networks in the country [14-17]. Aside from greater access to the internet, access to real-time health information, with audio-visual tools such as YouTube, which can enhance user attention, and respondent demographics (ie, young, educated, and living in major cities) could have contributed to increased internet use during the present pandemic in this study [16,17]. Despite this, there is broad consensus that misinformation is highly prevalent on the internet, and challenges such as limited reach to underserved populations may caution against overreliance on the internet for health communication [18]. Nonetheless, using the internet—through social media handles and websites—for public health messaging accomplishes several of the goals of successful health communication. These goals include reaching a broad audience, creating interactive and ongoing community engagement, and broadening the transmission of urgent public health information [6,14].

Of note in this study is the relatively low proportion of respondents who depended on health care providers for COVID-19–related information. This was particularly common in more educated respondents (university education and above) compared to those with secondary school education or below. On the other hand, respondents with a higher income were more likely to obtain COVID-19 information from health care providers compared to the lowest category of earners in the population studied. A possible explanation for this finding could be that higher-income earners are more likely to have access to health care providers compared to low-income earners through health insurance or their own ability to pay out of pocket. In addition, higher-income earners are more likely to be able to afford medical consultations, facilitating direct access to COVID-19 information from health care providers. Similarly, higher-income earners were more likely to obtain COVID-19 information from their workplace. This finding could be ascribed to a likelihood of having more stable white- or blue-collar jobs compared to lower-income earners who are mostly students on stipends, small business entrepreneurs, and daily wage earners.

Perceived accuracy was not significantly associated with any particular source of COVID-19 information; however, respondents determined the accuracy of their COVID-19 information by cross-referencing with perceived reputable sources of information (ie, internet sources run by national or multinational health regulatory organizations), such as the NCDC and the WHO digital media handles. In an analysis of sources of information, high-income earners had more than 2 times higher odds of using reputable sources of information compared to low-income earners. Overall, our findings correspond with the literature, which indicates that a large proportion of the population relies on the media, as well as family and friends, to inform their perception of health risks during an outbreak [6].

Limitations

Limitations in the interpretation of findings from this study could be the mode and language of data collection. Since the study used an online survey in the English language, respondents were individuals with access to the internet who could communicate in English, alienating segments of the population with limited access to the internet or without English language fluency. Nonetheless, it is worth noting that individuals with internet access can serve as sources of information for those with limited or no access to the internet. In addition, respondents were predominantly youth, with university/polytechnic education, and earning less than NGN 20,000 per month, therefore limiting the generalizability of our findings.

Conclusion

The internet was the most common source of COVID-19 information, and the population sampled had a relatively high level of perceived accuracy for the COVID-19 information received. Further, high-income earners had a higher likelihood of using multiple sources of COVID-19 information. To determine the accuracy of COVID-19 information, a significant fraction of respondents cross-referenced any information received with news from official government regulatory bodies (eg, the NCDC) and the WHO.

The dissemination of timely and accurate health information to support public health interventions is crucial during a pandemic. Therefore, targeted and evidence-based approaches must be implemented for effective communication. Findings from this study can inform health communication measures to mitigate the effects of the COVID-19 pandemic on the population and reduce the burden and spread of the disease.

Effective communication requires sensitivity to social perceptions and dissemination of information via relevant communication channels [19,20], as identified from this study. Policymakers responsible for COVID-19 risk communication in Nigeria may consider measures such as an increased focus on the internet (eg, use of NCDC social media handles and local traditional media stations with an online presence). In addition, white- and blue-collar employers can be encouraged to promote anti–COVID-19 health behaviors by conducting health education exercises for employees using official digital channels. Future studies should examine COVID-19 content across various potential sources to better understand the information obtained by the public.

Acknowledgments

The authors would like to acknowledge the role of Mr Babajide Sanwo-Olu, the Governor of Lagos State, and Dr Kadiri Hamza, the Deputy Governor of Lagos State, for providing leadership through the Incident Command System established to respond to the COVID-19 outbreak. The authors would also like to acknowledge the Directorate of Planning, Research and Statistics of the Lagos State Ministry of Health for substantial contributions to data collection.

This work was supported by funding from the Lagos State Government. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 45 KB - publichealth_v7i1e22273_app1.docx]

References

1. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization. 2020 Feb 28. URL: https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19) [accessed 2020-07-06]
2. WHO Director-General's opening remarks at the media briefing on COVID-19 — 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-07-06]
3. Coronavirus COVID-19 Global Cases. Johns Hopkins CSSE / Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-07-06]
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
5. Coronavirus disease (COVID-19) pandemic. Nigeria Centre for Disease Control. URL: https://covid19.ncdc.gov.ng/ [accessed 2020-07-06]
6. Abrams EM, Greenhawt M. Risk Communication During COVID-19. J Allergy Clin Immunol Pract 2020 Jun;8(6):1791-1794 [FREE Full text] [doi: 10.1016/j.jaip.2020.04.012] [Medline: 32304834]
7. Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-Related Information Sources and the Relationship With Confidence in People Coping with COVID-19: Facebook Survey Study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
8. Gamhewage G. An Introduction to Risk Communication. World Health Organization. 2014. URL: https://www.who.int/risk-communication/introduction-to-risk-communication.pdf?ua=1 [accessed 2020-06-19]
9. Chile N, Akwagyiram A. Nigeria reopens main cities Lagos and Abuja as lockdowns phased out. Reuters. 2020 May 4. URL: https://www.reuters.com/article/us-health-coronavirus-nigeria-lockdown/nigeria-reopens-main-cities-lagos-and-abuja-as-lockdowns-phased-out-idUSKBN22G225 [accessed 2020-06-19]

10. An update of COVID-19 outbreak in Nigeria: COVID-19 Outbreak in Nigeria Situation Report 12. Nigeria Centre for Disease Control. 2020 Mar 15. URL: https://ncdc.gov.ng/diseases/sitreps/?cat=14&name=An%20update%20of%20COVID-19%20outbreak%20in%20Nigeria [accessed 2020-06-19]
11. Adebimpe WO, Adeyemi DH, Faremi A, Ojo JO, Efuntoye AE. The relevance of the social networking media in Ebola virus disease prevention and control in Southwestern Nigeria. Pan Afr Med J 2015 Oct 10;22(Supp 1):7-11. [doi: 10.11604/pamj.supp.2015.22.1.6165]
12. Tran BX, Dang AK, Thai PK, Le HT, Le XTT, Do TTT, et al. Coverage of Health Information by Different Sources in Communities: Implication for COVID-19 Epidemic Response. Int J Environ Res Public Health 2020 May 20;17(10):3577 [FREE Full text] [doi: 10.3390/ijerph17103577] [Medline: 32443712]
13. Maduka O, Maleghemi S, Komakech W, Nwaduito I, Green P, Ikpe A, et al. Effective risk communication and contact tracing for Ebola virus disease prevention and control - Experiences from Port Harcourt, Nigeria. Public Health 2016 Jun;135:140-143. [doi: 10.1016/j.puhe.2015.10.037] [Medline: 26976486]
14. Heldman AB, Schindelar J, Weaver JB. Social Media Engagement and Public Health Communication: Implications for Public Health Organizations Being Truly “Social”. Public Health Rev 2013 Jun 3;35(1). [doi: 10.1007/bf03391698]
15. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for Health Promotion: Reviewing the Current Evidence. Am J Public Health 2013 Jan;103(1):e9-e18. [doi: 10.2105/ajph.2012.301071]
16. Lefebvre RC, Bornkessel AS. Digital Social Networks and Health. Circulation 2013 Apr 30;127(17):1829-1836. [doi: 10.1161/circulationaha.112.000897]
17. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May 14;5(5):e002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
18. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
19. Kalongo E. Communicating with Ebola-affected communities. World Health Organization. 2020 Oct. URL: https://www.who.int/risk-communication/communicating-ebola-affected-communities-unsp-oct-2014.pdf?ua=1 [accessed 2020-07-01]
20. Pillar 2: Risk Communication and Community Engagement. World Health Organization. 2020 May 01. URL: https://covid19partnersplatform.who.int/pillar/2 [accessed 2020-07-01]

Abbreviations

aOR: adjusted odds ratio
EVD: Ebola virus disease
LSMoH: Lagos State Ministry of Health
NCDC: Nigeria Centre for Disease Control
OR: odds ratio
PTF: COVID-19 Presidential Task Force
WHO: World Health Organization

Edited by T Sanchez; submitted 08.07.20; peer-reviewed by H Li Oi- Yee, CF Yen; comments to author 05.08.20; revised version received 12.08.20; accepted 17.08.20; published 25.01.21.

Please cite as:
Erinoso O, Wright KO, Anya S, Kuyinu Y, Abdur-Razzaq H, Adewuya A
Predictors of COVID-19 Information Sources and Their Perceived Accuracy in Nigeria: Online Cross-sectional Study
JMIR Public Health Surveill 2021;7(1):e22273
URL: http://publichealth.jmir.org/2021/1/e22273/
doi:10.2196/22273
PMID:33428580

Original Paper

Diet, Nutrition, Obesity, and Their Implications for COVID-19 Mortality: Development of a Marginalized Two-Part Model for Semicontinuous Data

Naser Kamyari[1], MSc; Ali Reza Soltanian[2], PhD; Hossein Mahjub[3], PhD; Abbas Moghimbeigi[4], PhD

[1]Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

[2]Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

[3]Research Center for Health Sciences, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

[4]Department of Biostatistics and Epidemiology, School of Health & Determinants of Health Research Center, Alborz University of Medical Sciences, Karaj, Iran

Corresponding Author:
Ali Reza Soltanian, PhD
Modeling of Noncommunicable Diseases Research Center
Hamadan University of Medical Sciences
Daneshgah-e-Bu Ali Sina
Hamadan
Iran
Phone: 98 81 38380025
Email: soltanian@umsha.ac.ir

Abstract

Background: Nutrition is not a treatment for COVID-19, but it is a modifiable contributor to the development of chronic disease, which is highly associated with COVID-19 severe illness and deaths. A well-balanced diet and healthy patterns of eating strengthen the immune system, improve immunometabolism, and reduce the risk of chronic disease and infectious diseases.

Objective: This study aims to assess the effect of diet, nutrition, obesity, and their implications for COVID-19 mortality among 188 countries by using new statistical marginalized two-part models.

Methods: We globally evaluated the distribution of diet and nutrition at the national level while considering the variations between different World Health Organization regions. The effects of food supply categories and obesity on (as well as associations with) the number of deaths and the number of recoveries were reported globally by estimating coefficients and conducting color maps.

Results: The findings show that a 1% increase in supplementation of pulses reduced the odds of having a zero death by 4-fold (OR 4.12, 95% CI 11.97-1.42). In addition, a 1% increase in supplementation of animal products and meat increased the odds of having a zero death by 1.076-fold (OR 1.076, 95% CI 1.01-1.15) and 1.13-fold (OR 1.13, 95% CI 1.0-1.28), respectively. Tree nuts reduced the odds of having a zero death, and vegetables increased the number of deaths. Globally, the results also showed that populations (countries) who consume more eggs, cereals excluding beer, spices, and stimulants had the greatest impact on the recovery of patients with COVID-19. In addition, populations that consume more meat, vegetal products, sugar and sweeteners, sugar crops, animal fats, and animal products were associated with more death and less recoveries in patients. The effect of consuming sugar products on mortality was considerable, and obesity has affected increased death rates and reduced recovery rates.

Conclusions: Although there are differences in dietary patterns, overall, unbalanced diets are a health threat across the world and not only affect death rates but also the quality of life. To achieve the best results in preventing nutrition-related pandemic diseases, strategies and policies should fully recognize the essential role of both diet and obesity in determining good nutrition and optimal health. Policies and programs must address the need for change at the individual level and make modifications in society and the environment to make healthier choices accessible and preferable.

(JMIR Public Health Surveill 2021;7(1):e22717) doi:10.2196/22717

KEYWORDS

COVID-19; diet; nutrition; obesity; marginalized two-part model; semicontinuous data

Introduction

Transmission of COVID-19 began in Wuhan, Hubei Province, China on December 31, 2019 [1,2]. According to the latest World Health Organization (WHO) report on July 3, 2020, there were 11,188,120 confirmed cases and 528,431 deaths worldwide, with 1505 total cases and 69.3 deaths per 1 million population [3]. The WHO named it a global pandemic because of the rapid outbreak of the disease worldwide [4,5].

The COVID-19 epidemic started during winter in areas of the world where the consumption of wildlife is not uncommon. Coronavirus is one of the viruses causing the common cold, a disease that has never had a cure nor any effective prevention or vaccine. However, there are relatively consistent data suggesting that the risk of contracting the common cold is high under inadequate sleep, psychosocial or physical stress including exposure to cold temperatures, inadequate nutrition, and any condition that compromises the body's immune system [6].

A high percentage of COVID-19 deaths worldwide are associated with one or more chronic conditions. It is also evident that older people are at a higher risk for severe illness with this pandemic [7,8]. Nutrition is not a treatment for COVID-19, but it is a modifiable contributor to the development of chronic disease, which is highly associated with COVID-19 severe illness and deaths [9]. A well-balanced diet and healthy patterns of eating strengthens the immune system, improves immunometabolism, and reduces the risk of chronic disease and infectious diseases [10,11]. Furthermore, nutrition may have a positive or negative impact on COVID-19, as it may be a way to support people at higher risk for the disease (ie, older people and people with pre-existing conditions of noncommunicable diseases) [12].

It is clear in these challenging times that optimizing nutrition is important, not only for ourselves but also for every patient that goes through their own period of treatment. Every health system should be aware of the benefits of healthy eating and be able to provide sound nutritional guidance to their patients, especially those with chronic disease. Having knowledge about nutritional interventions that may help prevent chronic conditions and their associated risks is now more important than ever [13].

On the other hand, being overweight or obese are interpreted as excessive fat [14] accumulation and represent a risk to health [15]. Most of the world's populations live in countries where being overweight or obese kill more people than being underweight. However, does it cause a decrease in the immune system or severity of COVID-19? Is it dangerous toward getting an infection and the mortality of COVID-19?

This study aims to assess the effect of diet, nutrition, and obesity on COVID-19 mortality among 188 countries by using new statistical marginalized two-part (mTP) models. Hence, we globally evaluated the distribution of diet and nutrition on the national level while considering the variation between different regions. The effects of food supply categories and obesities on (as well as associations with) the number of deaths and the number of recoveries is reported worldwide by estimating coefficients and conducting color maps.

Methods

Overview

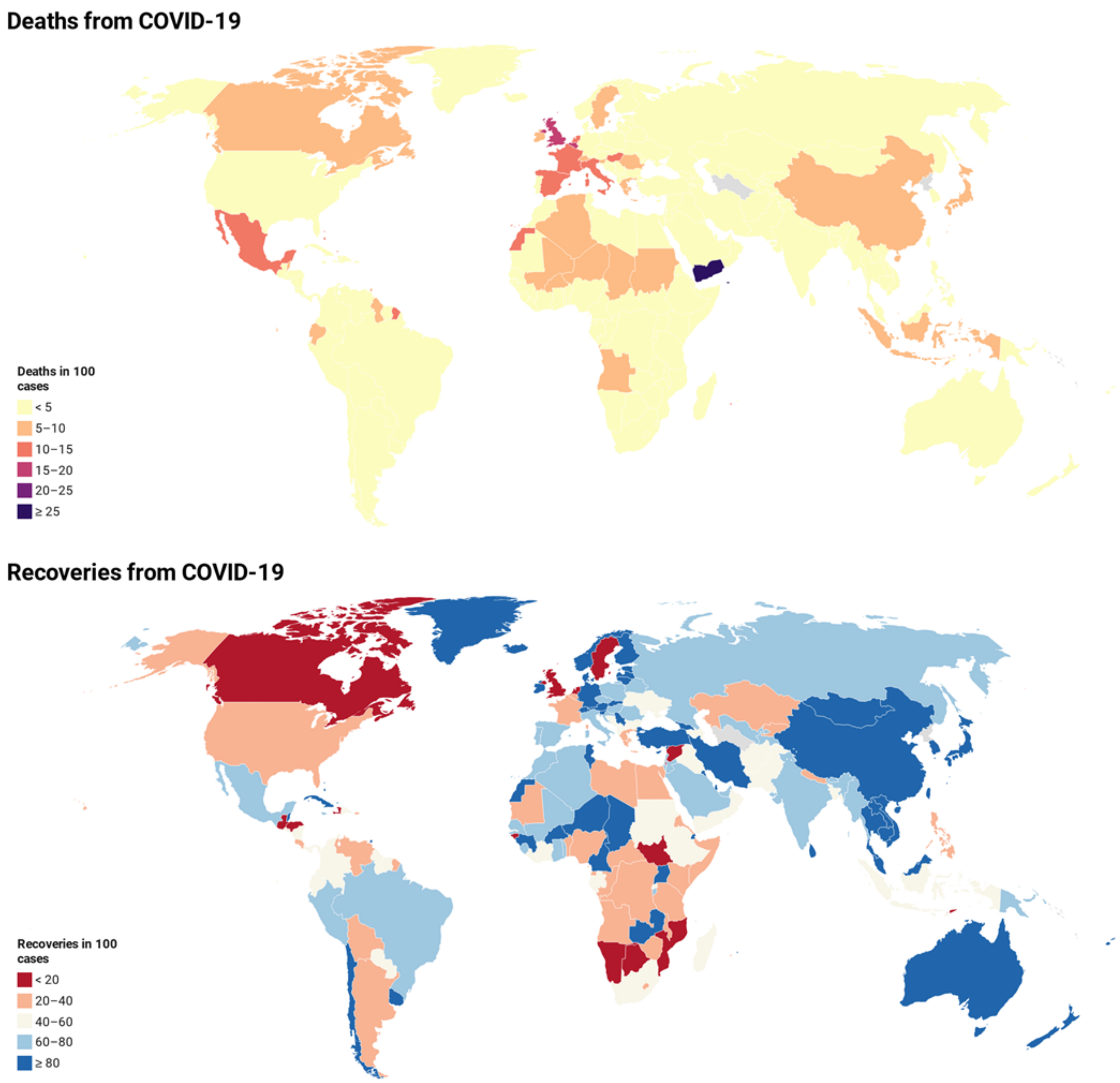
This section starts with a short description of the data set and information on relevant sources. In the following section, we introduce the conventional two-part (TP) regression model and the proposed mTP regression model for semicontinuous data. For the continuous part, we considered two flexible distributions including log-normal (LN) and beta prime (BP). We also described their properties to assess the overall impact of covariates on the marginal mean and demonstrated that the proposed model outperforms the conventional model. Finally, the proposed mTP model was applied to the healthy diet data set on fat quantity and protein to investigate the effects of nutrition categories and obesity on the number of deaths and recoveries in 100 cases of COVID-19.

Dietary, Obesity, and COVID-19 Data

Food supply data is some of the most important data in both Food and Agriculture Organization (FAO)/WHO STAT [16]. In fact, this data is the basis for estimations of global and national undernourishment assessment when it is combined with parameters and other data sets. In addition, both businesses and governments use this data for economic analysis and policy setting, and the academic community also uses this data.

In this data set, we combined data of different types of food, world population obesity and undernourished rates, and the global COVID-19 cases count from around the world (188 countries) to learn more about how a healthy eating style could help combat COVID-19. In addition, from the data set, we can gather information regarding diet patterns from countries with lower COVID-19 infection rates and adjust our own diet accordingly. The spread of the disease, deaths, recoveries, and their different distributions are shown in Figure 1, which can be evaluated according to the WHO regions.

Figure 1. World map related to the number of deaths (top) and the number of recoveries (bottom) in 100 cases of COVID-19 as of July 3, 2020.



From the data sets, accessible as Google sheets in GitHub [17], we have used fat quantity and protein for different categories of food (all calculated as percentage of total intake amount). We have also added on the obesity rate (in percentage) for comparison. The end of the data sets also included the most up-to-date confirmed infections, deaths, recoveries, and active cases (also in percentage of current population for each country). In this study, response variables were the deaths in 100 cases and the recoveries in 100 cases that were continuously (ranged 0 to 100) measured for 188 countries [18].

To synchronize results relative to interregional variations, data sets were grouped according to WHO regions (Multimedia Appendix 1), and a mTP analysis of deaths and recoveries was conducted using a random effects (regions cluster) model. Supply food data description is described in Multimedia Appendix 2. Both fat quantity and protein data sets, including 23 categories, were obtained from the FAO database [19] and were used to show the specific types of food that belong to each category for assessing influential effects of the fat quantity and protein supply.

Semicontinuous response variables such as mortality indexes are typically characterized by the presence of zeros and positive continuous outcomes that are often right skewed. In this paper, we propose a class of models for positive and zero responses by means of a zero-augmented mixed regression model. Under this class, we are particularly interested in studying positive responses whose distribution accommodates skewness. At the same time, responses can be zero, and therefore, we justified the use of a zero-augmented mixture model.

Marginalized Two-Part Models for Semicontinuous Data

Conventional Two-Part Model

We began with a review of the conventional TP model presented in Cragg [20], Manning et al [21], Duan et al [22], and elsewhere. Let Y_{ij} be a semicontinuous variable for the i-*th*

(i=1,2, ..., n) subject at cluster j (j=1,2, ..., n_i). For nonnegative data (Y_{ij}≥0) consisting of independent observations that clustered in the j-*th* level, the generic form of the conventional TP model can be written as:

where $\pi_{ij} = Pr(Y_{ij}>0)$, $1_{(.)}$ is the indicator function, and $g(y_{ij}|y_{ij}>0)$ is any density function applicable to the positive values of Y_{ij}, although the LN density is often chosen. This model is parameterized following equations (2) and (3) relevant to zero and nonzero components, respectively:

where $\boldsymbol{Z}_{ij}$ is a 1 × q covariate (used as an explanatory variable) vector, α is a q × 1 regression coefficient vector, and $\boldsymbol{b_{1i}}$ is the cluster-level random effect in the zero component. The location parameter μ_{ij} is modeled in the second part of the TP model assuming a log link:

where $\boldsymbol{X}_{ij}$ is a 1 × p covariate vector, β is a p × 1 regression coefficient vector, and $\boldsymbol{b_{2i}}$ is again the cluster-level random effect in the nonzero component. The error term ε_{ij} is assumed to be normally distributed as N(0,). Note that this TP mixed model can be extended to include additional random effects. For illustration purposes and simplicity, we restricted attention here to the TP mixed models with two levels; extensions to multilevel models are straightforward.

When fitting this model to independent responses, the binary and conditionally continuous components of the likelihood are separable, and therefore, these two parts are fit separately. The binary component is often modeled using logistic regression, and the continuous component can be fit using standard regression models such as the BP [23], LN [24], gamma [25,26], and long-skew normal [27].

The marginal mean and variance of Y_{ij} from a TP model can be derived as follows:

When LN is assumed in the continuous part, the marginal mean is:

$E(Y_{ij}) = \nu_{ij} = \pi_{ij} exp\{\mu_{ij} + \sigma^2/2\}$ **(5)**

Marginalized Two-Part Model

To obtain interpretable covariate effects on the marginal (unconditional) mean, we proposed the following mTP model that parameterizes the covariate effects directly in terms of the marginal mean, $\nu_i = E(Y_i)$, on the original (ie, untransformed) data scale. The mTP model specifies the linear predictors:

where b_{1i} represents the random effect that accounts for the within-subject correlation pertaining to the clustered measures in the zero part, $b_{1i} \sim N(0,$).

where b_{2i} represents the random effect that accounts for the within-subject correlation pertaining to the clustered measures in the continuous part, $b_{2i} \sim N(0,$).

The two random effect intercepts b_{1i} and b_{2i} in the two processes of zero and nonzero are assumed to be independent and uncorrelated. is the vector of covariates for the i-*th* subject measured at the j-*th* cluster for the binary part, and is the vector of covariates for the i-*th* subject measured at the j-*th* cluster used for the continuous part. The two parts might have common covariates or completely different ones. α is the vector of model coefficients corresponding to the binary part, and β is the vector of coefficients corresponding to the continuous part, conditional on the values being nonzero. The model can be easily extended to include higher-order random effects.

Marginalized Two-Part Log-Normal Model

When modeling semicontinuous data, the continuous component is most frequently modeled using a LN distribution. The generic form of the marginalized two-part log-normal (mTP LN) model for independent responses can be written as in equation (1), with $g(y_{ij}|y_{ij}>0)$, taking the LN density function $LN(.; \mu, \sigma^2)$ with mean μ and variance σ^2 on the log scale. The marginal mean and variance of Y_{ij} are:

$E(Y_{ij}) = \nu_{ij} = \pi_{ij} exp(\mu_{ij} + \sigma^2/2)$ **(8)**

$Var(Y_{ij}) = \pi_{ij} exp(2\mu_{ij} + \sigma^2)[exp(\sigma^2) - \pi_{ij}]$ **(9)**

The likelihood (L), parameterized in terms of π_{ij} and μ_{ij}, is:

where $\varphi(b_{1ij}, b_{2ij})$ represents the bivariate normal distribution for the random effects with a mean vector of zeros and variance-covariance matrix and for zero and nonzero parts, respectively.

To use this LN likelihood framework, the marginal mean in equation (8) can be rearranged to solve for μ_{ij}, yielding:

Noting that:

and:

Marginalized Two-Part Beta Prime Model

The BP distribution [28,29] is also known as inverted beta distribution or beta distribution of the second kind, often the model of choice for fitting semicontinuous data where the response variable is measured continuously on the positive real line (Y>0) because of the flexibility it provides in terms of the variety of shapes it can accommodate. The probability density function of a BP distributed random variable Y parameterized in terms of its mean μ and a precision parameter φ is given by:

where, B denotes the beta function μ>0, φ>0, $E(Y) = \mu$, and $Var(Y) = (\mu(1 + \mu)) / \varphi$.

To obtain interpretable covariate effects on the marginal mean, we proposed the following mTP model that parameterizes the covariate effects directly in terms of the marginal mean, $\nu_{ij} = E(Y_{ij})$, on the original (ie, untransformed) data scale. The mTP model with random (cluster) effects $\boldsymbol{Z}_{1ij}$ and $\boldsymbol{Z}_{2ij}$ for the zero and the continuous components, respectively, specifies the linear predictors:

where, and have full rank p and q for the zero and the continuous components, respectively; $\alpha_{(p+1)\times 1}$ and $\beta_{(q+1)\times 1}$ are the corresponding vectors of the regression coefficients. As seen in equations (15) and (16), the mixing probability and mean of the component of the continuous parts are linked to the independent variables through logit and logarithmic link functions. The vectors $\boldsymbol{b}_1=(b_{11}, b_{12}, ..., b_{1m})'$ and $\boldsymbol{b}_2=(b_{21}, b_{22}, ..., b_{2m})'$ denote random effects of the third level in the components of logistic and continuous, respectively. For simplicity of interpretation and mathematical calculations, the random effects (b_1, b_2) were assumed to be joint normally distributed with mean zero and variances and , respectively [30,31]. The errors term $e_{ij} \sim N(0,$) was also assumed to be of normal distribution and independent of the random effects.

Let $\psi_{ij} = I(y_{ij}>0)$ denote the indicator of Y_{ij} being nonzero. The general form of the likelihood function for the i-*th* subject can be described as follows:

where the log-likelihood for the binary part is:

and the log-likelihood for the continuous part is:

with , which can be implemented in the SAS NLMIXED procedure by quasi-Newton optimization with adaptive Gaussian quadrature techniques [32]. With the conventional model, the likelihood and score equations can be separated into two independent components: one for the binary part and one for the continuous part. In contrast, note that the score equations for the mTP model are not separable, and thus, the binary and continuous parts are fit simultaneously. Model-based asymptotic standard errors are computed using Fisher information matrix, I(α, β, σ), as:

with the maximum likelihood estimates substituted for α, β, and σ.

Results

In this section, the proposed mTP model was applied to the healthy diet data set on fat and protein to investigate the effects of supplementation categories on the number of deaths per 100 cases and recoveries per 100 cases of COVID-19. The estimations of mTP BP and mTP LN related to deaths and recoveries are shown in Tables 1 and 2, respectively. In these tables, variances (and) show the variety of responses among level 2 (ie, the WHO regions) related to each part of the zero and nonzero (ie, positive) components. Tables 1 and 2 show that almost all categories have the same effect on the number of deaths and recoveries in 100 cases. The number of deaths per 100 cases, number of recoveries per 100 cases, and the obesity rates until July 3, 2020, for all countries and split by the WHO regions is shown in Multimedia Appendix 3. Deaths are more common in Western and Southwest Europe (eg, Belgium, the United Kingdom, France, Italy, Hungary, Netherlands, and Spain), North America (eg, Mexico, Bahamas, Canada, Barbados, Belize, and the United States), and North Africa (eg, Western Sahara, Chad, Algeria, and Niger). The highest number of deaths occurred in Yemen (26.62 deaths per 100 cases), which could be due to the crises caused by the war and the poor health conditions in this country in the last years. Frequently, it seems that the northern regions of the world appear to have had more deaths, which may be due to temperature differences between the two hemispheres.

Table 1. Results of marginalized two-part BP and LN model in predicting number of deaths per 100 cases and considering the cluster effect of World Health Organization regions in the fat data set.

Fat (categories)	Fat quantity				Protein			
	Zero component		Nonzero component		Zero component		Nonzero component	
	Coefficient (α)		Coefficient (β)		Coefficient (α)		Coefficient (β)	
Alcoholic beverages								
BP[a]	—[b]	—	–7.0449	0.1145	1.3670	0.4907	0.04707	0.1286
LN[c]	—	—	–6.4512	0.1165	1.3752	0.4814	0.04918	0.1189
Animal products								
BP	0.0401	0.7457	–0.0028	0.1539	*0.0736* [d]	0.6212	0.0105	0.1236
LN	0.0405	0.6543	–0.0031	0.1495	*0.0812*	0.5965	0.0108	0.1209
Animal fats								
BP	0.1505	0.6458	0.0009	0.1223	7.6487	0.3055	0.3259	0.1212
LN	0.1489	0.6147	0.0009	0.1239	7.5461	0.2891	0.3319	0.1801
Aquatic products other								
BP	4.3474	0.7746	–13.9857	0.1123	30.9618	0.4060	–0.7270	0.1201
LN	4.5421	0.7469	–10.2576	0.1127	29.1456	0.3995	–0.7345	0.1193
Cereals excluding beer								
BP	–0.0839	0.6195	–0.0218	0.1485	–0.0548	0.4994	–0.0171	0.1163
LN	–0.0956	0.6015	–0.0221	0.1399	–0.0551	0.4861	–0.0170	0.1123
Eggs								
BP	0.3217	0.6155	–0.1291	0.1391	0.3993	0.4487	–0.0466	0.1297
LN	0.3514	0.5412	–0.1299	0.1381	0.4125	0.4912	–0.0481	0.1183
Fish seafood								
BP	–0.0556	0.6000	0.0383	0.1903	0.0542	0.4220	0.0305	01318
LN	–0.0598	0.5816	0.0401	0.1849	0.0556	0.4001	0.0351	0.1301
Fruits excluding wine								
BP	–0.4420	0.6323	0.0933	0.1633	–0.4744	0.5036	0.0783	0.1176
LN	–0.4511	0.5971	0.0931	0.1617	–0.4598	0.5121	0.0803	0.1165
Meat								
BP	0.0394	0.7340	0.0008	0.1699	*0.1246*	0.4690	0.0256	0.1174
LN	0.0391	0.7300	0.0008	0.1684	*0.1268*	0.4581	0.0221	0.1173
Miscellaneous								
BP	2.9863	0.6822	2.0655	0.0994	0.0900	0.6768	–0.0090	0.1336
LN	2.5531	0.5836	2.1510	0.0991	0.0912	0.6154	–0.0061	0.1136
Milk excluding butter								
BP	0.0589	0.6160	–0.0139	0.1667	0.5456	0.4718	0.0561	0.1259
LN	0.0512	0.6013	–0.0142	0.1561	0.5537	0.4316	0.0560	0.1241
Offals								
BP	0.0196	0.6429	–0.1844	0.1022	–0.0526	0.4808	0.0704	0.1227
LN	0.0197	0.6129	–0.1798	0.1124	–0.0541	0.4493	0.0713	0.1201
Oilcrops								
BP	–0.2833	0.5580	0.0179	0.1876	–0.1453	0.5600	0.0090	0.1245

Fat (categories)	Fat quantity				Protein			
	Zero component		Nonzero component		Zero component		Nonzero component	
	Coefficient (α)		Coefficient (β)		Coefficient (α)		Coefficient (β)	
LN	–0.2891	0.5137	0.0184	0.1773	–0.1457	0.5413	0.0094	0.1239
Pulses								
BP	*–1.4170*	0.1944	0.1150	0.1434	–0.1583	0.4308	0.1496	0.1290
LN	*–1.4242*	0.1832	0.1159	0.1400	–0.1581	0.4311	0.1499	0.1289
Spices								
BP	–0.5889	0.4622	0.0476	0.1907	–0.0916	0.4309	–0.0561	0.1161
LN	–0.5887	0.4604	0.0467	0.1891	–0.0914	0.4312	–0.0551	0.1163
Starchy roots								
BP	–0.9187	0.5559	–0.3262	0.1637	0.3208	0.4513	–0.0469	0.1246
LN	–0.9188	0.5412	–0.3301	0.1621	0.3208	0.4511	–0.0461	0.1212
Stimulants								
BP	0.6042	0.6715	–0.0287	0.1704	0.3476	0.4530	–2.4158	0.1227
LN	0.6101	0.6712	–0.0288	0.1698	0.3479	0.4530	–2.4129	0.1201
Sugar crops								
BP	–13.0537	0.5727	–6.2283	0.1063	5.7548	0.4394	1.4861	0.1217
LN	–12.5132	0.5624	–6.3120	0.1059	5.5431	0.4329	1.4869	0.1214
Sugar sweeteners								
BP	–0.8040	0.6104	10.2992	0.2136	1.7184	0.4899	0.5042	0.1328
LN	–0.8042	0.6112	9.4532	0.2013	1.7204	0.4782	0.5100	0.1236
Tree nuts								
BP	0.3739	0.6207	0.1138	0.2178	*–0.0736*	0.6211	–0.0105	0.1236
LN	0.3721	0.6211	0.1142	0.2017	*–0.0740*	0.6127	–0.0104	0.1221
Vegetal products								
BP	–0.0401	0.8313	0.0028	0.1485	39.1538	0.5764	2.8053	0.1327
LN	–0.0413	0.8219	0.0027	0.1449	27.1870	0.4365	2.8101	0.1254
Vegetable oils								
BP	–0.0008	0.6667	0.0025	0.1663	0.0228	0.4442	–0.0376	0.1228
LN	–0.0012	0.6120	0.0024	0.1659	0.0224	0.4318	–0.0354	0.1224
Vegetables								
BP	–0.5748	0.6205	–0.4784	0.1099	1.3637	0.3991	*0.7713*	0.1050
LN	–0.5739	0.5945	–0.4754	0.1098	1.3641	0.3981	*0.7716*	0.1002
Obesity								
BP	0.0228	0.5954	0.0054	0.1716	0.0228	0.4798	0.0054	0.1261
LN	0.0212	0.5871	0.0057	0.1624	0.0227	0.4821	0.0034	0.1178

[a]BP: beta prime.

[b]Empty cells related to unestimated or nonconverged values.

[c]LN: log-normal.

[d]Italics indicate statistical significance at the .05 significance level.

Table 2. Results of marginalized two-part BP and LN model in predicting number of deaths in 100 cases and considering the cluster effect of World Health Organization regions in the protein data set.

Protein (categories)	Fat quantity				Protein			
	Zero component		Nonzero component		Zero component		Nonzero component	
	Coefficient (α)		Coefficient (β)		Coefficient (α)		Coefficient (β)	
Alcoholic beverages								
BP[a]	0.3934	—[b]	–3.3321	0.0053	–0.5040	—	0.3291	0.0241
LN[c]	0.3941	0.6541	–3.3520	0.0051	-0.5139	—	0.3301	0.0240
Animal products								
BP	–0.0518	—	–0.0030	5.0390	–0.0377	—	–0.0036	—
LN	–0.0520	—	–0.0031	5.0011	–0.0378	—	–0.0035	—
Animal fats								
BP	–0.1417	0.6458	0.0037	0.1223	*–4.3929* [d]	0.6854	–0.1882	0.0291
LN	–0.1419	0.5454	0.0039	0.1221	*–4.3821*	0.6855	–0.1881	0.0297
Aquatic products other								
BP	0.1026	0.7746	0.0998	0.1123	13.6501	0.6543	–0.1588	0.0102
LN	0.9817	0.7751	0.0981	0.1123	11.5479	0.6571	–0.1581	0.0115
Cereals excluding beer								
BP	0.2108	0.6195	0.0201	0.1485	0.0935	0.5987	0.0070	—
LN	0.2107	0.5981	0.0200	0.1498	0.0992	0.4564	0.0074	0.0120
Eggs								
BP	1.0198	0.6155	0.0585	0.1391	0.3015	0.7154	0.0971	1.0199
LN	1.0211	0.6154	0.0594	0.1384	0.3101	0.6245	0.0954	0.8745
Fish seafood								
BP	2.1882	0.6000	–0.0587	0.1903	0.3784	0.6542	–0.0109	2.0553
LN	2.1452	0.6124	–0.0591	0.1914	0.3781	0.6549	–0.0117	2.4923
Fruits excluding wine								
BP	3.5529	0.6323	0.0049	0.1633	0.5023	0.6980	–0.0284	0.3869
LN	3.5504	0.6341	0.0051	0.1601	0.5001	0.7401	–0.0294	0.3881
Meat								
BP	–0.0020	0.7340	–0.0081	0.1699	–0.0720	0.7012	–0.0050	2.8965
LN	–0.0019	0.7341	–0.0078	0.1692	–0.0711	0.6984	–0.0050	2.8457
Miscellaneous								
BP	8.988	0.6822	–0.1517	0.0994	–0.1208	0.6503	–0.0057	1.8106
LN	7.5415	0.6824	–0.1540	0.0982	–0.1235	0.6511	–0.0064	1.8001
Milk excluding butter								
BP	–0.1359	0.6160	–0.0020	0.1667	–0.3806	0.4562	0.0398	0.4047
LN	–0.1314	0.5912	–0.0080	0.1724	–0.3817	0.4575	0.0410	0.4521
Offals								
BP	–3.0231	0.6429	–0.1611	0.1022	–0.1169	0.6985	0.0301	—
LN	–3.0024	0.6540	–0.1617	0.1512	–0.1141	0.7211	0.0341	—
Oilcrops								

Protein (categories)	Fat quantity				Protein			
	Zero component		Nonzero component		Zero component		Nonzero component	
	Coefficient (α)		Coefficient (β)		Coefficient (α)		Coefficient (β)	
BP	0.0091	0.5580	0.0070	0.1876	–0.1958	0.5913	–0.0269	1.2942
LN	0.0084	0.5557	0.0072	0.1868	–0.1954	0.5914	–0.0264	1.3125
Pulses								
BP	*–1.6086*	0.1944	–0.3639	0.1434	4.1312	0.2456	–0.2215	0.0303
LN	*–1.5401*	0.1984	–0.3578	0.1545	4.0128	0.2541	–0.2211	0.0311
Spices								
BP	3.1903	0.4622	–0.2388	0.1907	0.0560	0.5441	0.0208	0.6601
LN	3.1912	0.4684	–0.2491	0.1912	0.0541	0.5237	0.0214	0.6641
Starchy roots								
BP	0.4714	0.5559	0.0997	0.1637	–0.2431	0.6003	0.1089	0.0603
LN	0.4787	0.5651	0.0984	0.1746	–0.2433	0.6210	0.1146	0.0701
Stimulants								
BP	0.5780	0.6715	0.1503	0.1704	–0.0752	0.6439	2.1776	0.0100
LN	0.5417	0.6871	0.1511	0.1724	–0.754	0.7431	2.1290	0.0150
Sugar crops								
BP	0.3393	0.5727	0.9821	0.1063	0.3819	0.5589	*–4.7273*	0.1670
LN	0.3365	0.5821	0.9807	0.1163	0.3814	0.5613	*–4.7198*	0.1687
Sugar sweeteners								
BP	1.3970	0.6104	*–9.6769*	0.2136	0.5242	0.6987	–0.4085	0.2326
LN	1.4121	0.6255	*–9.5421*	0.2166	0.5420	0.6999	–0.4142	0.2401
Tree nuts								
BP	0.2521	0.6207	*–0.1732*	0.2178	0.0378	0.7432	0.0036	2.8525
LN	0.2515	0.6210	*–0.1732*	0.2245	0.0412	0.7421	0.0065	3.1450
Vegetal products								
BP	0.0517	0.8313	0.0030	0.1485	–18.9397	0.7823	0.7848	0.0053
LN	0.0521	0.8214	0.0024	0.1521	–15.2981	0.7888	0.6954	0.0055
Vegetable oils								
BP	0.0081	0.6667	–0.0007	0.1663	0.4467	0.5968	–0.0039	2.4944
LN	0.0043	0.6821	–0.0014	0.1681	0.4428	0.5960	–0.0041	2.4912
Vegetables								
BP	2.7153	0.6205	–0.1959	0.1099	1.6107	0.5935	0.1003	0.0184
LN	2.7154	0.6650	–0.1991	0.1124	1.5459	0.6520	0.1000	0.0156
Obesity								
BP	0.0042	0.5954	–0.0009	0.1716	0.0042	0.4986	–0.0009	4.1135
LN	0.0041	0.5954	–0.0005	0.1712	0.0041	0.5101	–0.0009	4.1232

[a]BP: beta prime.

[b]Empty cells related to unestimated or nonconverged values.

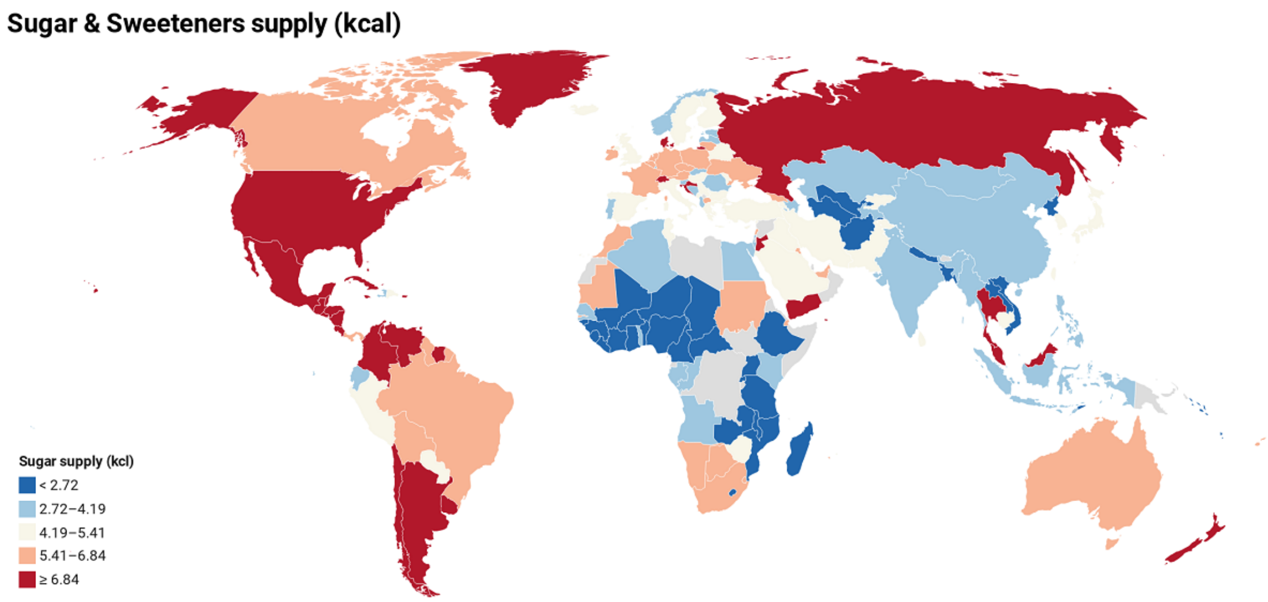
[c]LN: log-normal.

[d]Italics indicate statistical significance at the .05 significance level.

According to the results of Table 1, except pulses in fat quantity and animal products, meat, tree nuts, and vegetables in the protein data set, all categories had no significant effect on the number of deaths. A 1% increase in supplementation of pulses reduced the odds of having a zero death by 4-fold (1 / exp(–1.417) = 4.1251). In addition, a 1% increase in supplementation of animal products and meat increased the odds of having a zero death by 1.076-fold (exp(0.0736) = 1.076) and 1.133-fold (exp(0.0736) = 1.133), respectively. Tree nuts reduced the odds of having a zero death, and vegetables increased the number of deaths.

Continuously, except animal fats, sugar sweeteners, and tree nuts in fat quantity, and animal fats and sugar crops in the protein data, all categories had no significant effect on the number of recoveries (Table 2). The effect of consuming sugar products on mortality was considerable. Every 1% increment in sugar sweeteners decreased the number of recoveries by 98.17% (–9.68, 95% CI –12.6440 to –6.7098). Tree nuts in fat quantity also reduced the number of recoveries by 16.9% (–0.1732, 95% CI –0.3157 to –0.3070). In the protein data, sugar crops reduced the number of recoveries by 99.11% (1 – exp(–4.7273) = 0.9911). The world map related to sugar and sweetener supply is shown in Figure 2. Based on the results of the proposed model and estimates of the effects of sugar, our prediction for the coming days is that the countries of the Americas, with more sugar product intake, will probably face more deaths.

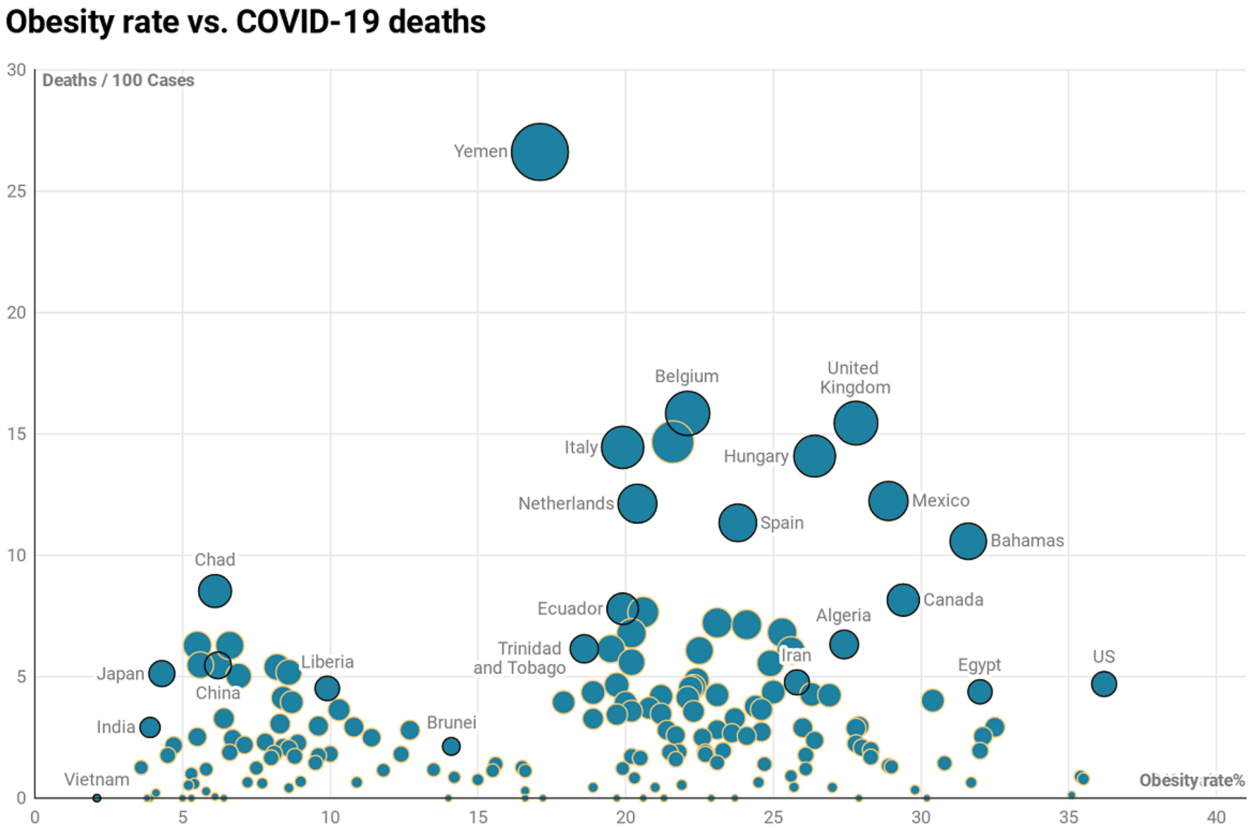
Figure 2. World map related to sugar and sweeteners supply (kcal).



For further evaluation, we calculated correlations between categories (plus obesity rate) with the number of deaths (Figure 3 A and C) and the number of recoveries (Figure 3 B and D) by using the bivariate Pearson correlation. Results of the correlations showed that, in the protein data, countries that consumed more spices, tree nuts, cereals, aquatic products, stimulants, vegetable oils, oil crops, pulses, fruit (wine), and alcoholic beverage (in order) had fewer deaths from COVID-19, and conversely, countries that consumed more meat, vegetables, vegetal products, sugar and sweeteners, animal products, animal fats, sugar crops, milk, fish, offals, miscellaneous, eggs, and starchy roots (in order) had more deaths from COVID-19. In the fat quantity data, countries that consumed more sugar and sweeteners, miscellaneous, tree nuts, meat, animal products, animal fats, offals, and fish had more deaths from COVID-19. Finally, same as the mTP model results, obesity has affected increased death rates and reduced recovery rates in all correlation analyses (Figures 3 and 4).

Figure 3. Bivariate Pearson correlation between nutrition categories (plus obesity) and the number of deaths (A and C) and the number of recoveries (B and D) in 100 cases of COVID-19.

Figure 4. Scatterplot of obesity rate versus deaths per 100 cases of COVID-19 by country. The circle's size is related to the number of deaths per 100 cases.



Discussion

Principal Results

In this study, we proposed a mTP regression model for clustered semicontinuous diet and nutrition data. This model allows investigators to obtain covariate effects on the marginal mean of the outcome (eg, deaths and recoveries). It also has an unconditional interpretation of the covariate effect on the marginal mean. Our proposed mTP model had satisfactory performance in the diet and nutrition data analysis.

Findings of this study show that populations (countries) who consume more eggs, cereals excluding beer, spices, and stimulants had the greatest impact on the recovery of patients with COVID-19. In addition, populations that consumed more meat, vegetal products, sugar and sweeteners, sugar crops, animal fats, and animal products were associated with more deaths and less recoveries in patients. The effect of consuming sugar products on mortality was considerable. In addition, obesity has affected increased death rates and reduced recovery rates.

Comparison With Prior Work

Healthy diets and physical activity are key to good nutrition and necessary for a long and healthy life and prevention of chronic disease [33]. Eating nutrition dense foods and balancing energy intake with necessary physical activity to maintain a healthy weight is essential at all stages of life. Unbalanced consumption of foods high in energy (sugar, starch, and fat) and low in essential nutrition contributes to energy excess, being overweight, and being obese. The amount of energy consumed in relation to physical activity and the quality of food are key determinants of nutrition-related chronic disease [11]. In a review study from January 2020, Zhang and Liu [13] reviewed the importance of some nutrition interventions (vitamins, minerals, immunoenhancers) in infectious and respiratory diseases. The authors suggested that the nutritional status of each patient who was infected should be evaluated before the administration of general treatments, and the current children's RNA‐virus vaccines, including the influenza vaccine, should be used for people who are not infected and health care workers. Moreover, the results of their review showed that all the potential interventions (nutritional or immunoenhancers) should be implemented to control COVID‐19 if the infection is uncontrollable [13]. Our results also confirm these associations by introducing influential diet categories, including sugar and sweeteners, animal products, animal fats, sugar crops, miscellaneous, and tree nuts as more important risk factors for death or slowing of recovery in patients with COVID-19.

Recent studies point to obesity as a critical risk factor for being hospitalized or dying from COVID-19 [34-36]. Indeed, a high prevalence of obesity has been observed in patients with COVID-19, requiring invasive mechanical ventilation [37], a robust proxy of SARS-CoV-2 severity. In patients younger than 60 years, those with obesity were at almost double the risk of being admitted to critical care when compared with patients of a normal weight [38]. Results from this study confirm previous findings on the risk of obesity and add that obesity slows down patients' recovery and treatment.

People need to eat fewer prepared foods and more complex plant-based foods [11]. Although there are differences in dietary patterns, overall, unbalanced diets are a health threat across the world and do not just affect death rates but also the quality of life. To achieve best results in preventing nutrition-related pandemic diseases, strategies and policies should fully recognize the essential role of both diet and obesity in determining good nutrition and optimal health. Policies and programs must address the need for change at the individual level as well as the modifications in society and the environment to make healthier choices accessible and preferable.

Study Limitations

We have some limitation in using the nutrition data sets. The study is based on observational data, and inevitably with 188 countries included, there were variations in how the data were collected. This study included 23 dietary attributes; some that are of interest to health such as saturated and monounsaturated fatty acids and free sugars across the diet (not just those in drinks) were not included in the analysis. The study also did not take into account lifestyle factors, such as smoking and physical activity, that can have a significant impact on the risk of the disease outcomes used in the study.

Finally, we remind all our readers to take care of themselves during this pandemic, follow the guidelines of the Centers for Disease Control and Prevention [39], and eat healthy foods with sufficient amounts of fruits and vegetables as previously discussed.

Conclusions

Good nutrition is important before, during, and after an infection. The findings of this study show that populations who consume more eggs, cereals excluding beer, spices, and stimulants had the greatest impact on the recovery of patients with COVID-19. In addition, populations that consumed more meat, vegetal products, sugar and sweeteners, sugar crops, animal fats, and animal products were associated with more deaths and less recoveries in patients. The effect of consuming sugar products on mortality is considerable. In addition, obesity has affected increased death rates and reduced recovery rates. Although there are differences in dietary patterns, overall, unbalanced diets are a health threat across the world and affect not only death rates but also the quality of life. To achieve best results in preventing nutrition-related pandemic diseases, strategies and policies should fully recognize the essential role of both diet and obesity in determining good nutrition and optimal health. Policies and programs must address the need for change at the individual level as well as the modifications in society and the environment to make healthier choices accessible and preferable.

Acknowledgments

This study was adapted from a PhD thesis by Naser Kamyari (No. 9804253260) at Hamadan University of Medical Sciences. The authors are grateful to Dr Maryam Seyedtabib for providing the data for this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Geographical distribution of the six World Health Organization regions and member states for each region (alphabetical order).
[PNG File , 859 KB - publichealth_v7i1e22717_app1.png]

Multimedia Appendix 2
The specific types of food that belong to each category for the fat quantity and protein data sets.
[DOCX File , 14 KB - publichealth_v7i1e22717_app2.docx]

Multimedia Appendix 3
Number of deaths per 100 cases, number of recoveries per 100 cases, and obesity rates until July 3, 2020, by countries and split by World Health Organization regions.
[DOCX File , 3385 KB - publichealth_v7i1e22717_app3.docx]

References

1. Zhan C, Tse CK, Fu Y, Lai Z, Zhang H. Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data. PLoS One 2020;15(10):e0241171 [FREE Full text] [doi: 10.1371/journal.pone.0241171] [Medline: 33108386]
2. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]

3. Coronavirus disease (COVID-19) situation report – 165. World Health Organization. 2020 Jul 03. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200703-covid-19-sitrep-165.pdf?sfvrsn=b27a772e_6 [accessed 2021-01-19]
4. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
5. Xu C, Zhang X, Wang Y. Mapping of health literacy and social panic via web search data during the COVID-19 public health emergency: infodemiological study. J Med Internet Res 2020 Jul 02;22(7):e18831 [FREE Full text] [doi: 10.2196/18831] [Medline: 32540844]
6. Kalantar-Zadeh K, Moore LW. Impact of nutrition and diet on COVID-19 infection and implications for kidney health and kidney disease management. J Ren Nutr 2020 May;30(3):179-181 [FREE Full text] [doi: 10.1053/j.jrn.2020.03.006] [Medline: 32291198]
7. Osama T, Pankhania B, Majeed A. Protecting older people from COVID-19: should the United Kingdom start at age 60? J R Soc Med 2020 May;113(5):169-170 [FREE Full text] [doi: 10.1177/0141076820921107] [Medline: 32315559]
8. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of symptoms prognostic of COVID-19 severity: multivariate data analysis of a case series in Henan Province. J Med Internet Res 2020 Jun 30;22(6):e19636 [FREE Full text] [doi: 10.2196/19636] [Medline: 32544071]
9. Butler MJ, Barrientos RM. The impact of nutrition on COVID-19 susceptibility and long-term consequences. Brain Behav Immun 2020 Jul;87:53-54 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.040] [Medline: 32311498]
10. Caccialanza R, Laviano A, Lobascio F, Montagna E, Bruno R, Ludovisi S, et al. Early nutritional supplementation in non-critically ill patients hospitalized for the 2019 novel coronavirus disease (COVID-19): rationale and feasibility of a shared pragmatic protocol. Nutrition 2020 Jun;74:110835 [FREE Full text] [doi: 10.1016/j.nut.2020.110835] [Medline: 32280058]
11. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
12. Muscogiuri G, Barrea L, Savastano S, Colao A. Nutritional recommendations for CoVID-19 quarantine. Eur J Clin Nutr 2020 Jun;74(6):850-851 [FREE Full text] [doi: 10.1038/s41430-020-0635-2] [Medline: 32286533]
13. Zhang L, Liu Y. Potential interventions for novel coronavirus in China: a systematic review. J Med Virol 2020 May;92(5):479-490 [FREE Full text] [doi: 10.1002/jmv.25707] [Medline: 32052466]
14. Garvey WT. Clinical definition of overweight and obesity. In: Gonzalez-Campoy JM, Hurley DL, Garvey WT, editors. Bariatric Endocrinology: Evaluation and Management of Adiposity, Adiposopathy and Related Diseases. Cham: Springer; 2019:121-143.
15. Bansal S, Zilberman D. Role of health care expenditure in countering adverse effects of obesity on health: evidence from global data. 2018 Presented at: 2018 International Association of Agricultural Economists (IAAE) Conference; July 28-August 2, 2018; Vancouver, BC. [doi: 10.22004/ag.econ.275941]
16. Leclercq C, Allemand P, Balcerzak A, Branca F, Sousa RF, Lartey A, et al. FAO/WHO GIFT (Global Individual Food Consumption Data Tool): a global repository for harmonised individual quantitative food consumption studies. Proc Nutr Soc 2019 Nov;78(4):484-495. [doi: 10.1017/S0029665119000491] [Medline: 30816080]
17. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub. 2020. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2021-01-19]
18. Reported cases and deaths by country, territory, or conveyance. Worldometer. URL: https://www.worldometers.info/coronavirus/#countries [accessed 2020-07-03]
19. FAOSTAT: FAO statistical database. Food and Agriculture Organization of the United Nations. 2016. URL: http://www.fao.org/faostat/en/#home [accessed 2021-01-19]
20. Cragg JG. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 1971 Sep;39(5):829. [doi: 10.2307/1909582]
21. Manning WG. A two-part model of the demand for medical care : preliminary results from the Health Insurance Study. In: van der Gaag J, Perlman M, editors. Health, Economics, and Health Economics: Proceedings of the World Congress on Health Economics, Leiden, The Netherlands, September 1980. Amsterdam: North-Holland Publishing Company; 1981:103-123.
22. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. J Business Econ Statistics 1983 Apr;1(2):115. [doi: 10.2307/1391852]
23. Bourguignon M, Santos-Neto M, de Castro M. A new regression model for positive data. arXiv Preprint posted online April 20, 2018.
24. Smith VA, Preisser JS, Neelon B, Maciejewski ML. A marginalized two-part model for semicontinuous data. Stat Med 2014 Dec 10;33(28):4891-4903. [doi: 10.1002/sim.6263] [Medline: 25043491]
25. Nieto-Barajas LE, Bandyopadhyay D. A zero-inflated spatial gamma process model with applications to disease mapping. J Agricultural Biol Environ Statistics 2013 Feb 7;18(2):137-158. [doi: 10.1007/s13253-013-0128-z]

26. Voronca DC, Gebregziabher M, Durkalski VL, Liu L, Egede LE. Marginalized two part models for generalized gamma family of distributions. arXiv Preprint posted online November 18, 2015.
27. Chai HS, Bailey KR. Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. Stat Med 2008 Aug 15;27(18):3643-3655 [FREE Full text] [doi: 10.1002/sim.3210] [Medline: 18186536]
28. Keeping ES. Introduction to Statistical Inference. North Chelmsford, MA: Courier Corporation; 1995.
29. McDonald JB. Some generalized functions for the size distribution of income. In: Chotikapanich D, editor. Modeling Income Distributions and Lorenz Curves. New York, NY: Springer; 2008:37-55.
30. Li X, Hedeker D. A three-level mixed-effects location scale model with an application to ecological momentary assessment data. Stat Med 2012 Nov 20;31(26):3192-3210 [FREE Full text] [doi: 10.1002/sim.5393] [Medline: 22865663]
31. Liu L, Ma J, Johnson B. A multi-level two-part random effects model, with application to an alcohol-dependence study. Stat Med 2008 Aug 15;27(18):3528-3539. [doi: 10.1002/sim.3205] [Medline: 18219701]
32. Tooze J, Grunwald G, Jones R. Analysis of repeated measures data with clumping at zero. Stat Methods Med Res 2002 Aug;11(4):341-355. [doi: 10.1191/0962280202sm291ra] [Medline: 12197301]
33. Ng R, Sutradhar R, Yao Z, Wodchis W, Rosella L. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. Int J Epidemiol 2020 Feb 01;49(1):113-130 [FREE Full text] [doi: 10.1093/ije/dyz078] [Medline: 31329872]
34. Dietz W, Santos-Burgoa C. Obesity and its implications for COVID-19 mortality. Obesity (Silver Spring) 2020 Jun;28(6):1005. [doi: 10.1002/oby.22818] [Medline: 32237206]
35. Finer N, Garnett SP, Bruun JM. COVID-19 and obesity. Clin Obes 2020 Jun;10(3):e12365 [FREE Full text] [doi: 10.1111/cob.12365] [Medline: 32342637]
36. Kassir R. Risk of COVID-19 for patients with obesity. Obes Rev 2020 Jun;21(6):e13034 [FREE Full text] [doi: 10.1111/obr.13034] [Medline: 32281287]
37. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, LICORN and the Lille COVID-19 and Obesity study group. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obesity (Silver Spring) 2020 Jul;28(7):1195-1199 [FREE Full text] [doi: 10.1002/oby.22831] [Medline: 32271993]
38. Lighter J, Phillips M, Hochman S, Sterling S, Johnson D, Francois F, et al. Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission. Clin Infect Dis 2020 Jul 28;71(15):896-897 [FREE Full text] [doi: 10.1093/cid/ciaa415] [Medline: 32271368]
39. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html?CDC_%20AA_refVal5https%3A%2F%202Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%%202Fprepare%2Fprevention.html [accessed 2020-03-28]

Abbreviations

BP: beta prime
FAO: Food and Agriculture Organization
LN: log-normal
mTP: marginalized two-part
TP: two-part
WHO: World Health Organization

Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by I Mircheva, M Shamsizadeh, Z Shayan; comments to author 03.08.20; accepted 18.08.20; published 26.01.21.

Please cite as:
Kamyari N, Soltanian AR, Mahjub H, Moghimbeigi A
Diet, Nutrition, Obesity, and Their Implications for COVID-19 Mortality: Development of a Marginalized Two-Part Model for Semicontinuous Data
JMIR Public Health Surveill 2021;7(1):e22717
URL: http://publichealth.jmir.org/2021/1/e22717/
doi:10.2196/22717
PMID:33439850

Original Paper

Using the Novel Mortality-Prevalence Ratio to Evaluate Potentially Undocumented SARS-CoV-2 Infection: Correlational Study

Sheng-Hsuan Lin[1], ScM, MD, ScD; Shih-Chen Fu[1], PhD; Chu-Lan Michael Kao[1], PhD

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

Corresponding Author:
Chu-Lan Michael Kao, PhD
Institute of Statistics
National Chiao Tung University
Assembly Building I, 4th Floor
1001 University Road
Hsinchu, 30010
Taiwan
Phone: 886 35712121 ext 56822
Email: chulankao@gmail.com

Abstract

Background: The high prevalence of COVID-19 has resulted in 200,000 deaths as of early 2020. The corresponding mortality rate among different countries and times varies.

Objective: This study aims to investigate the relationship between the mortality rate and prevalence of COVID-19 within a country.

Methods: We collected data from the Johns Hopkins Coronavirus Resource Center. These data included the daily cumulative death count, recovered count, and confirmed count for each country. This study focused on a total of 36 countries with over 10,000 confirmed COVID-19 cases. Mortality was the main outcome and dependent variable, and it was computed by dividing the number of COVID-19 deaths by the number of confirmed cases.

Results: The results of our global panel regression analysis showed that there was a highly significant correlation between prevalence and mortality (ρ=0.8304; P<.001). We found that every increment of 1 confirmed COVID-19 case per 1000 individuals led to a 1.29268% increase in mortality, after controlling for country-specific baseline mortality and time-fixed effects. Over 70% of excess mortality could be attributed to prevalence, and the heterogeneity among countries' mortality-prevalence ratio was significant (P<.001). Further, our results showed that China had an abnormally high and significant mortality-prevalence ratio compared to other countries (P<.001). This unusual deviation in the mortality-prevalence ratio disappeared with the removal of the data that was collected from China after February 17, 2020. It is worth noting that the prevalence of a disease relies on accurate diagnoses and comprehensive surveillance, which can be difficult to achieve due to practical or political concerns.

Conclusions: The association between COVID-19 mortality and prevalence was observed and quantified as the mortality-prevalence ratio. Our results highlight the importance of constraining disease transmission to decrease mortality rates. The comparison of mortality-prevalence ratios between countries can be a powerful method for detecting, or even quantifying, the proportion of individuals with undocumented SARS-CoV-2 infection.

(JMIR Public Health Surveill 2021;7(1):e23034) doi:10.2196/23034

KEYWORDS

COVID-19; prevalence; mortality; undocumented infection; mortality-prevalence ratio; China

Introduction

The first cluster of cases of pneumonia, which was later identified as COVID-19, a disease caused by the SARS-CoV-2 virus [1], was reported in Wuhan, China on December 31, 2019 [2]. The disease outbreak in China eventually developed into a pandemic, which forced widespread changes throughout the world and added substantial disease and economic burden worldwide. As of May 2, 2020, more than 36 countries have reported at least 10,000 cases of COVID-19. A total of around 4 million cases and 274,000 deaths have been reported [2,3]. Numerous studies have been conducted to investigate the biological and epidemiologic characteristics of COVID-19 [4-6].

Most results have been derived from traditional epidemiological models, wherein both COVID-19 mortality (ie, the "case fatality rate" in some literature) and recovery rates were assumed to be constants. However, in a study conducted by Bialek et al [7], heterogeneity in mortality rates was found among countries and cities, but this has been attributed to the assumed underlying medical conditions within an area [8-10]. The trend in mortality over time is also controversial [11-13]. Although results from an exponential growth model have shown an overall exponential decay in mortality within China since the disease outbreak [13], there has been evidence that shows disease prevalence influences disease mortality to a considerable extent. The rapid increase in the number of infections may result in the collapse of the health care system, leading to a sharp rise of mortality [11,12]. Despite the inconsistencies in mortality characteristics between studies, previous analyses have been performed with data that were collected before March, 2020. Up until then, only a few countries reported the number of COVID-19 deaths, whereas most areas were not majorly affected by COVID-19.

This study aims to sophisticatedly quantify the relationship between COVID-19 prevalence and mortality, by using data that have been updated up until May 2, 2020. A linear relationship between prevalence and mortality was observed, and this was referred to as the mortality-prevalence ratio. The global mortality-prevalence ratio was estimated after adjusting for country-specific baseline mortality and time-fixed effects. Country-specific mortality-prevalence ratio values can be used as a powerful index for identifying countries with a substantial number undocumented infections or overburdened health care systems.

Methods

COVID-19–related data [14] was downloaded from the Johns Hopkins Coronavirus Resource Center. These data included the cumulative number of confirmed cases (C_{it}), death cases (D_{it}), and recovered cases (R_{it}) of the i^{th} country from January 22 to May 2, 2020. We then matched each country with their respective national population data, which were provided by World Population Review [15]. Countries without a matched population were excluded from this study. After exclusion, 174 countries remained in our dataset. We later aggregated the remaining countries to obtain the corresponding global counts.

For each country and each time point, we computed the following 3 metrics, along with the global data: (1) the number of cases still in treatment (CT_{it}), which represents the total number of COVID-19 cases that involved medical assistance at time t; (2) the prevalence of COVID-19 in country i at time t (P_{it}); and (3) COVID-19 mortality in country i at time t (M_{it}). For the sake of model stability, the analyses were only performed on countries with a C_{it} of ≥10,000. The following equations were used to calculate each metric:

$$CT_{it} = C_{it} – D_{it} − R_{it} \text{ .....}\mathbf{(1)}$$

$$P_{it} = C_{it}/\text{total population of country i} \text{ .....}\mathbf{(2)}$$

$$M_{it} = D_{it}/C_{it} \text{ .....}\mathbf{(3)}$$

To investigate the association between mortality and prevalence after adjusting for the baseline mortality in each country and the effect of regular fluctuation over time, we built the following panel regression model (ie, Model 1):

$$M_{it} = \beta_{country} + \beta_t + \gamma P_{it} + \varepsilon_{it} \text{ .....}\mathbf{(4)}$$

In this model, $\beta_{country}$ represents the country-specific baseline mortality; β_t is the time-fixed effect on the mortality; γ represents the global association between P_{it} and M_{it}, which we referred to as the global mortality-prevalence ratio; and ε_{it} is the residual. To meet the assumption that the mortality-prevalence ratio varies in each country, we built a panel regression model (ie, Model 2), in which the global mortality-prevalence ratio was replaced with the country-specific mortality-prevalence ratio, $\gamma_{country}$. Model 2 is described as follows:

$$M_{it} = \beta_{country} + \beta_t + \gamma_{country} P_{it} + \varepsilon_{it} \text{ .....}\mathbf{(5)}$$

In this model, $\gamma_{country}$ is the country-specific association between P_{it} and M_{it}, which we referred to as the country-specific mortality-prevalence ratio. Furthermore, we tested whether $\gamma_{country}$ differed between each country with an analysis of variance test. We also tested whether the difference could be treated as the random effect of a normal population with the Shapiro-Wilk normality test. All analyses were conducted with R version 3.5.2. The approval of an institutional review board was not required because no individual-level/personal data were used.

Results

Table 1 shows the population and the total number of confirmed cases, death cases, and recovered cases for countries that reported at least 10,000 confirmed cases by May 2, 2020. Figure 1 shows the association between COVID-19 prevalence and mortality among these countries. The Spearman correlation coefficient was 0.8304 (*P*<.001) and the Pearson correlation coefficient was 0.3385 (*P*=.04). These values indicated a significant positive correlation between prevalence and mortality. COVID-19 mortality and prevalence were relatively high in the United Kingdom and Belgium, while the United States had a high prevalence and a relatively low mortality compared to countries with similar prevalence levels, such as China and Canada.

It is worth mentioning that the positive correlation between mortality and prevalence is not restricted to COVID-19. For example, when considering the prevalence and mortality of severe acute respiratory syndrome (SARS) on July 31, 2003 based on data from the World Health Organization, the Spearman correlation coefficient was 0.3915 (*P*=.03). Since the number of countries involved with the COVID-19 pandemic is considerably larger than those involved with the SARS pandemic, the correlation between COVID-19 mortality and prevalence is statistically more significant than the correlation between SARS mortality and prevalence.

The relationship between global COVID-19 prevalence and mortality can also be observed when time is considered (Figure 2). Both prevalence and mortality increased over time.

Table 1. Total population and the total number of confirmed cases, death cases, and recovered cases for countries that reported at least 10,000 confirmed cases by May 2, 2020.

Country	Total population, N	Confirmed cases, n	Deaths, n	Recovered cases, n
Austria	9,006,398	15,558	596	13,180
Belarus	9,449,323	15,828	97	3117
Belgium	11,589,623	49,517	7765	12,211
Brazil	212,559,417	97,100	6761	40,937
Canada	37,742,154	57,926	3684	23,814
Chile	19,116,201	18,435	247	9572
China	1,439,323,776	83,959	4637	78,586
Ecuador	17,643,054	27,464	1371	2132
France	65,273,511	168,518	24,763	50,663
Germany	83,783,942	164,967	6812	129,000
India	1,380,004,385	39,699	1323	10,819
Indonesia	273,523,615	10,843	831	1665
Iran	83,992,949	96,448	6156	77,350
Ireland	4,937,786	21,176	1286	13,386
Israel	8,655,535	16,185	229	9593
Italy	60,461,826	209,328	28,710	79,914
Japan	126,476,461	14,571	474	3205
Mexico	128,932,753	22,088	2061	12,377
Netherlands	17,134,872	40,434	5003	138
Pakistan	220,892,340	19,103	440	4817
Peru	32,971,854	42,534	1200	12,434
Poland	37,846,611	13,375	664	3762
Portugal	10,196,709	25,190	1023	1671
Qatar	2,881,053	14,872	12	1534
Romania	19,237,691	12,732	771	4547
Russia	145,934,462	124,054	1222	15,013
Saudi Arabia	34,813,871	25,459	176	3765
Singapore	5,850,342	17,548	17	1347
Spain	46,754,778	216,582	25,100	117,248
Sweden	10,099,265	22,082	2669	1005
Switzerland	8,654,622	29,817	1762	24,200
Turkey	84,339,067	124,375	3336	58,259
Ukraine	43,733,762	11,411	279	1498
United Arab Emirates	9,890,402	13,599	119	2664
United Kingdom	67,886,011	183,500	28,205	896
United States	331,002,651	1,132,539	66,369	175,382

Figure 1. COVID-19 mortality and prevalence of all countries (ρ=0.8304; *P*<.001). Only the top 20 countries with the highest prevalence are shown.

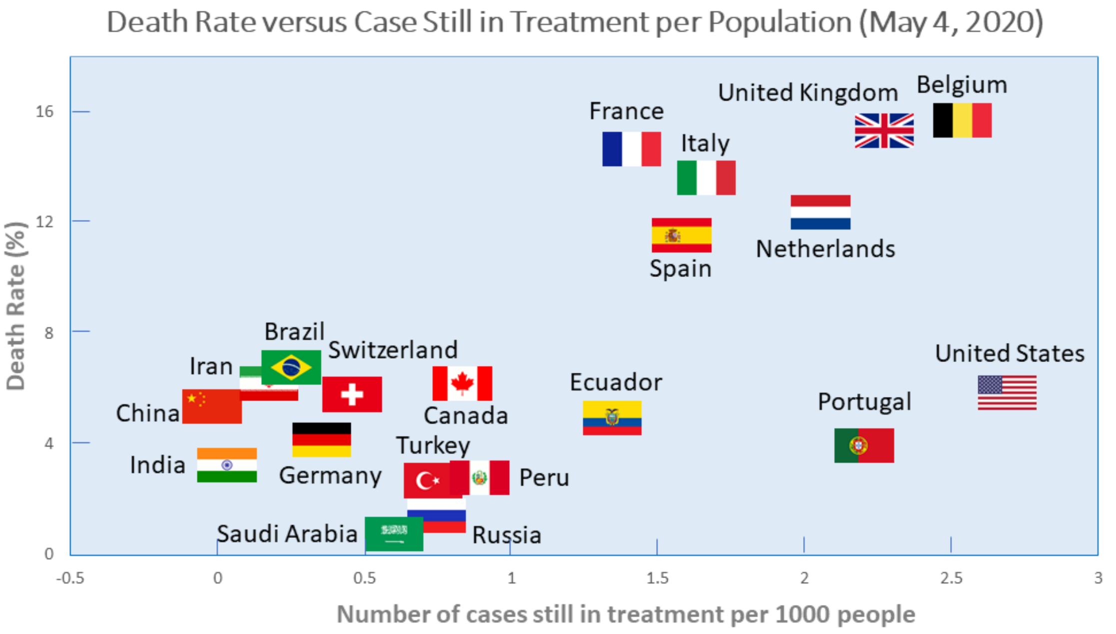
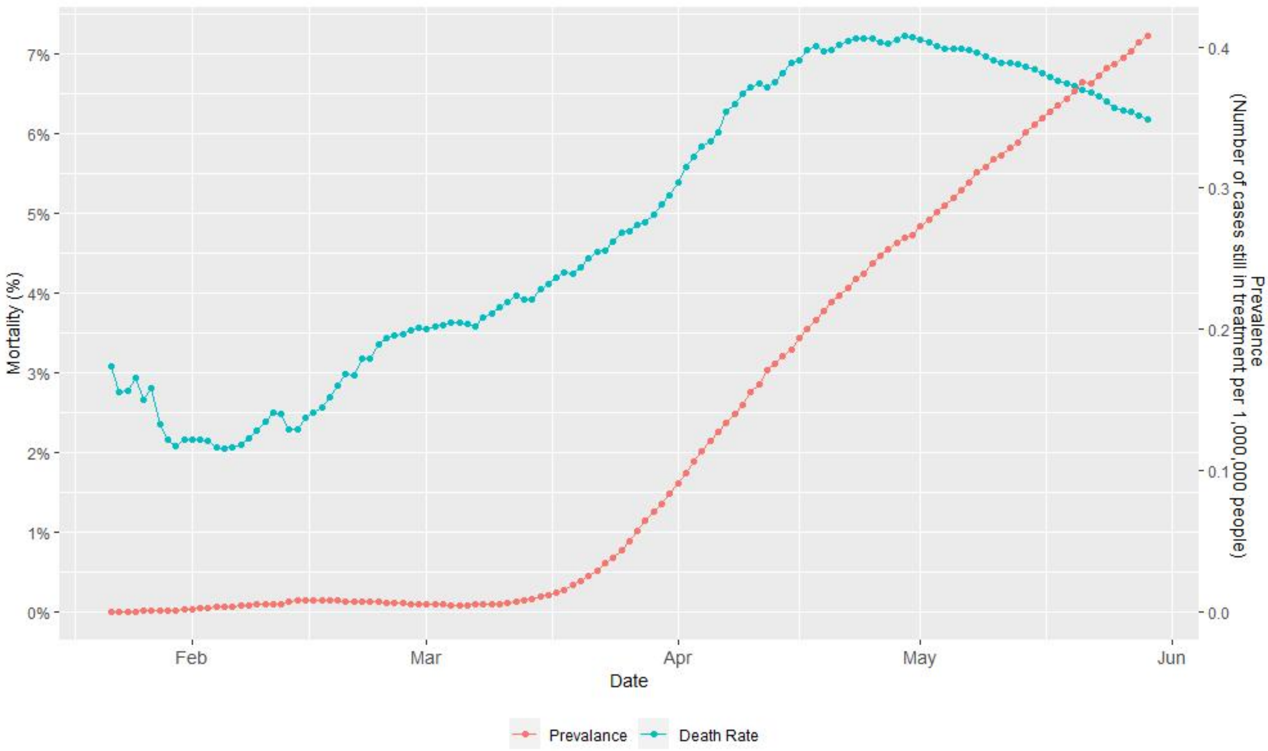


Figure 2. Trends of global COVID-19 mortality and prevalence over time.



In order to sophisticatedly estimate the relationship between mortality and prevalence, time and country-specific baseline mortalities in Model 1 were adjusted. The estimations for all coefficients are shown in Table 2. The global mortality-prevalence ratio, which was represented by γ in Model 1, was estimated to be 12.9268 (*P*<.001). This number can be interpreted as follows: an increment of 1 COVID-19 case per 1000 people is coupled with a 1.29268% (ie, 12.9268 × 1/1000 × 100) increase in mortality. The R^2 value that was calculated from Model 1 was 98.11%, and the partial R^2 value for was 70.41%. These values indicated that COVID-19 prevalence could roughly explain the 70% heterogeneity in excess mortality after controlling for country-specific baseline mortality and time-fixed effects. The analysis of variance test showed potential heterogeneity in the mortality-prevalence ratios among different countries (*P*<.001). Therefore, we performed a panel regression analysis based on Model 2, as shown in Table 2. It should be noted that the partial R^2 value for the mortality-prevalence ratio increased to 89.37% in Model 2.

Table 2. Estimation of all coefficients for Model 1 and Model 2.

Model	Estimation	*P* value	Partial R^2
Model 1[a]			
Mortality-prevalence ratio (ie, γ)	12.9268	<.001	0.7041
Model 2[b]			
Country-specific mortality-prevalence ratio (ie, $\gamma_{country}$)			0.8937
All data			
Austria	–30.8171	<.001	
Belarus	–19.2428	.27	
Belgium	45.4706	<.001	
Brazil	71.4636	.002	
Canada	65.5696	<.001	
Chile	–27.4605	.47	
China	347.7652	<.001	
Ecuador	–33.2373	<.001	
France	43.1863	<.001	
Germany	–22.7914	<.001	
India	–341.5505	.34	
Indonesia	–1205.3198	.79	
Iran	–52.9484	<.001	
Ireland	–23.1711	<.001	
Israel	–14.2313	.22	
Italy	13.0634	<.001	
Japan	334.2415	.17	
Mexico	–42.3179	.79	
Netherlands	4.1811	.08	
Pakistan	11.9388	.96	
Peru	–9.3371	.16	
Poland	286.3706	.28	
Portugal	–5.4107	.10	
Qatar	–14.4505	.006	
Romania	81.1284	.65	
Russia	–14.0904	.03	
Saudi Arabia	–26.3058	.09	
Singapore	–14.4186	.01	
Spain	7.1163	<.001	
Sweden	23.1690	<.001	
Switzerland	–32.3043	<.001	
Turkey	–17.0113	<.001	
Ukraine	–82.0183	.88	
United Arab Emirates	–14.6387	.60	
United Kingdom	14.4444	<.001	
United States	–4.1818	<.001	

Model	Estimation	*P* value	Partial R^2
All data excluding those collected from China after February 17, 2020			
Austria	–18.4144	<.001	
Belarus	–5.9724	.73	
Belgium	58.4075	<.001	
Brazil	89.1904	<.001	
Canada	80.0715	<.001	
Chile	–12.7334	.73	
China	–1.6094	.99	
Ecuador	–20.0146	<.001	
France	53.6918	<.001	
Germany	–17.5007	.001	
India	–265.0415	.46	
Indonesia	–1179.8199	.79	
Iran	–62.5286	<.001	
Ireland	–10.4239	<.001	
Israel	–3.2505	.78	
Italy	20.4366	<.001	
Japan	359.2586	.14	
Mexico	–31.1910	.85	
Netherlands	17.1263	<.001	
Pakistan	33.3642	.88	
Peru	–4.5690	.48	
Poland	308.5810	.24	
Portugal	8.3993	.01	
Qatar	–1.3688	.79	
Romania	98.2757	.58	
Russia	0.4826	.94	
Saudi Arabia	–12.5797	.42	
Singapore	–1.2272	.81	
Spain	16.7545	<.001	
Sweden	37.2237	<.001	
Switzerland	–19.2162	<.001	
Turkey	–3.9981	.34	
Ukraine	–67.5598	.90	
United Arab Emirates	–1.1999	.97	
United Kingdom	27.0183	<.001	
United States	7.5391	<.001	

[a]The R^2 value for Model 1 was 0.9811 (*P*<.001).

[b]The R^2 value for Model 2 was 0.9931 (*P*<.001).

We obtained estimated country-specific mortality-prevalence ratios that ranged from −1205 to 348 from the 36 countries that were included in our analysis (Figure 3). Absolute mortality-prevalence ratio values of >100 were found in 5 countries (ie, Indonesia, India, Poland, Japan, and China), of which China was the only country that had a significantly different mortality-prevalence ratio (348; *P*<.001). The results of our Shapiro-Wilk normality test meant that we could reject the hypothesis that all significant country-specific mortality-prevalence ratios came from a normal distribution (*P*<.001). As we further investigated the pattern of China's mortality-prevalence ratio over time, we noted that the correlation had turned from positive to negative after February 17, 2020 (Figure 4). This disparity was not observed if the data that was collected after February 17, 2020 was excluded (Figure 3) (Shapiro-Wilk normality test: *P*=.78).

Figure 3. Countries with significant country-specific mortality-prevalence ratios based on (A) all data and (B) all data excluding those collected from China after February 17, 2020.

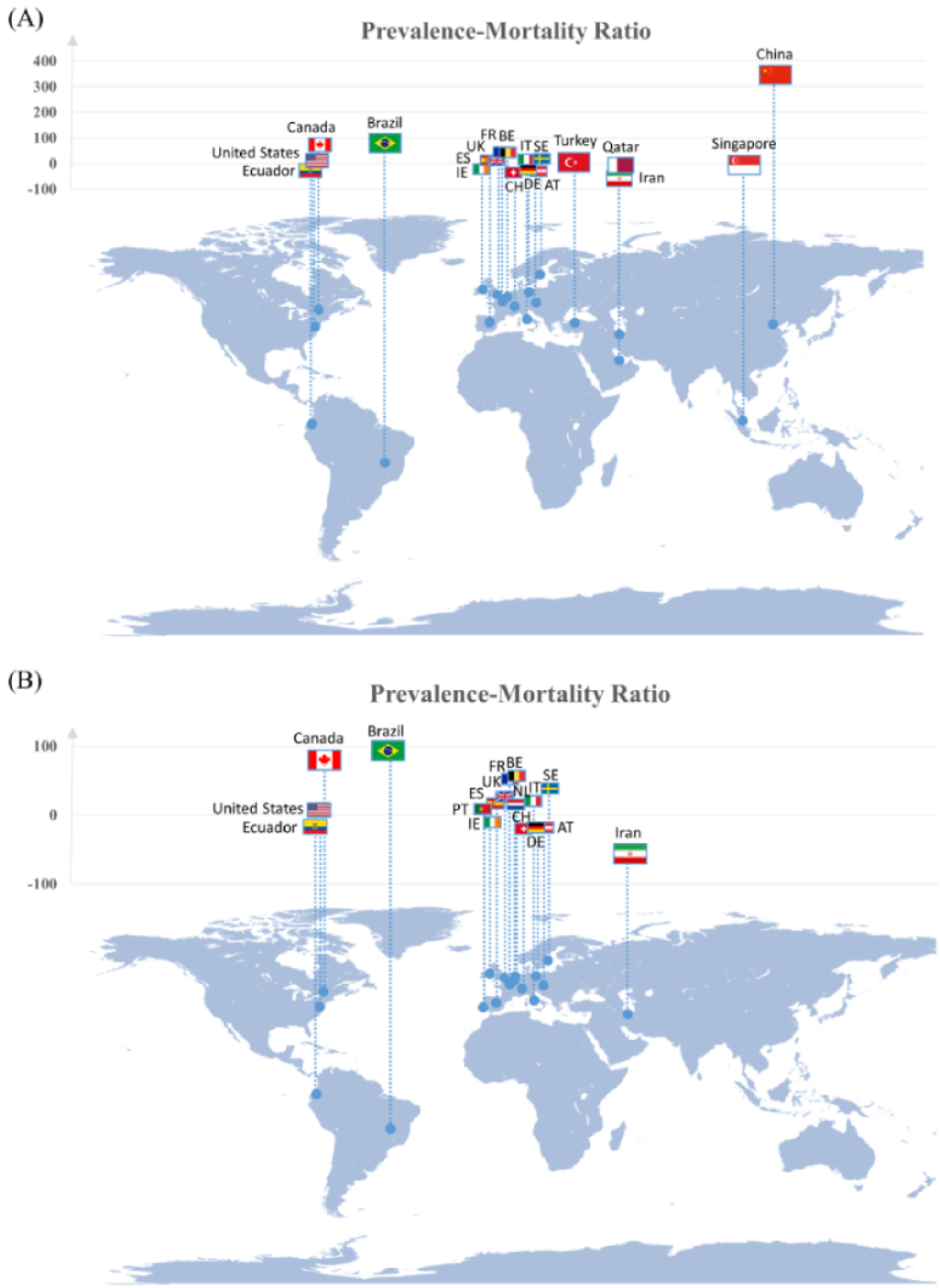
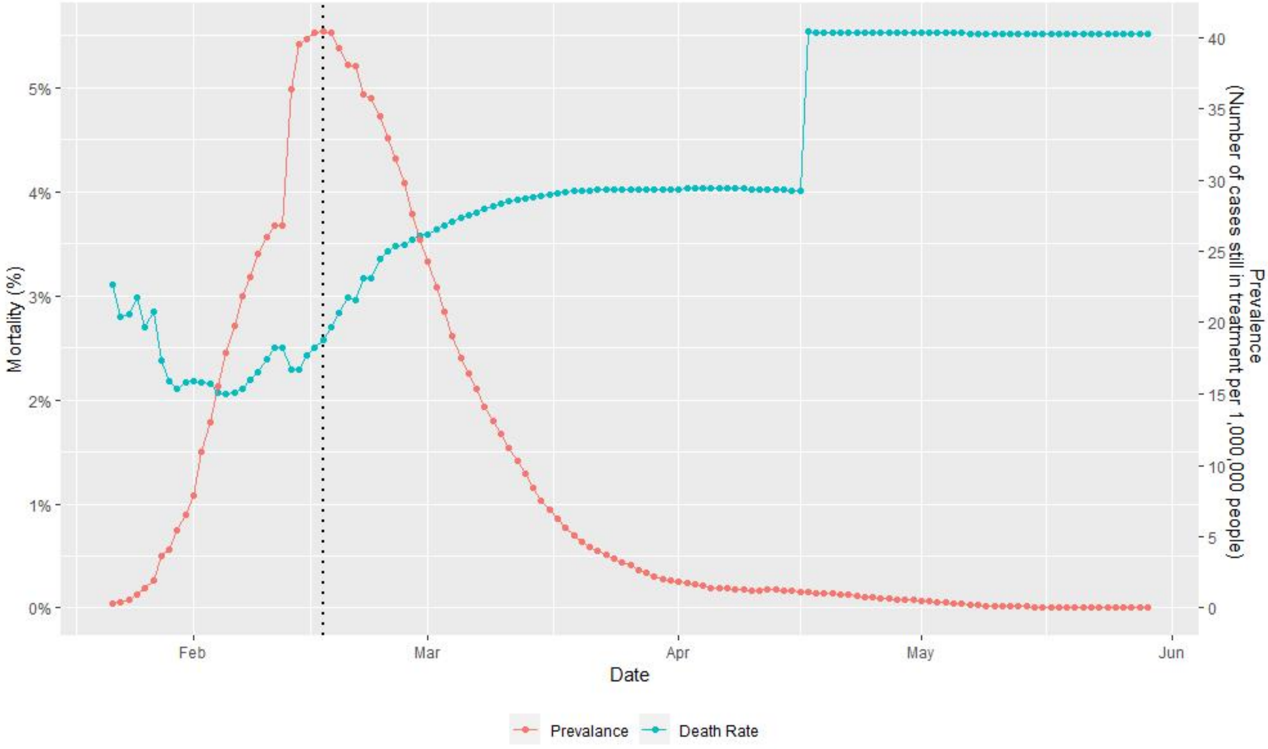


Figure 4. COVID-19 prevalence and mortality reported by China over time.



Discussion

This is the first study to assess the correlation between COVID-19 prevalence and mortality after adjusting for time-fixed effects and country-specific baseline mortality. We proposed the mortality-prevalence ratio as a novel characteristic for an infectious disease pandemic because of the high association between disease mortality and prevalence. In addition, a disparity in the mortality-prevalence ratios of 5 countries was observed; China was the only country with a significant mortality-prevalence ratio (348; *P*<.001). The disparity of China's mortality-prevalence ratio was due to the data reported after February 17, 2020. Although the mortality was proportional to the prevalence, the mortality-prevalence ratio was relatively robust to changes in prevalence (Figure 3). A high peak in mortality-prevalence ratios could be explained by a high proportion of undocumented infections within a country, which might be attributed to the limited number of diagnostic kits or changes in surveillance policies. An alternative explanation for the sudden rise of mortality could be that the health care system in China was relatively weak after February 17. However, this argument contradicts the fact that China's overall baseline country-specific mortality was typically followed by a steady increase in disease prevalence after February 17. The evolution of the pathogenicity and transmissibility of SARS-CoV-2 within China during this period could be another alternative reason for the disparity in mortality-prevalence ratios. Further studies are required to determine the underlying cause of this sharp increase in the mortality-prevalence ratio.

This study revealed the importance of public policies that aim to prevent disease transmission. These policies include social distancing, restricting travel, encouraging the wearing of facial masks and hand washing, and cancelling large events. Although the mortality rate of a certain infectious disease is traditionally assumed to be a constant in an infectious dynamic model [16], it is conceivable that a highly infectious disease affects the quality and availability of a health care system. The fast consumption of ventilation machines and the decline of nurse-to-patient ratios accelerate mortality. Prevention policies not only lower the financial burden on COVID-19 diagnosis and treatment, but also reduce COVID-19 mortality. Therefore, when future cost-effectiveness analyses are performed with respect to the balance between economic recovery and public health, it is crucial to consider the positive association between disease prevalence and mortality and the costs that come with it.

There are several limitations in this study. First, all results were based on ecological and panel data. Such data lack individual-level information. Therefore, the ecological fallacy would occur when trying to infer causality at the individual level [17]. The temporal effects of prevalence on mortality should also be confirmed to verify country-level causality. Second, although the prevalence of COVID-19 can generally be interpreted as an acute burden of health care, this relationship can be better verified when data on the actual insufficiencies of health care systems are available. Third, disease prevalence relies on accurate diagnoses and comprehensive surveillance, which can be difficult to achieve due to practical or political concerns. This was especially true at the beginning of the COVID-19 pandemic, which was when tests for COVID-19 were not accurate and data on people who died from COVID-19 may not have been captured. In this study, although countries with undocumented infections can be partially inferred with disparities in mortality-prevalence ratios, a more direct index merits further study.

In conclusion, we observed the relationship between COVID-19 mortality and prevalence and quantified this relationship as mortality-prevalence ratios. Our results highlight the benefit of constraining disease transmission to reduce mortality. Disparities in mortality-prevalence ratios can also be a powerful tool to detect, or even quantify, the proportion of undocumented infections.

Acknowledgments

We thank Ms Kai-Fen Wong for editing the figures. This study was supported by the Ministry Of Science and Technology, Taiwan (grant numbers 107-2118-M-009-003-MY2 and 108-2636-B-009-001).

Authors' Contributions

SHL and CLMK came up with the original idea. SHL and CLMK wrote the first version of manuscript. SCF edited the manuscript. CLMK built the model, wrote all the software, and conducted the data analysis.

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Coronavirusdisease (COVID-19) Situation Report–105. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200504-covid-19-sitrep-105.pdf?sfvrsn=4cdda8af_2 [accessed 2021-01-05]
3. Johns Hopkins Coronavirus Resource Center. Johns Hopkins University & Medicine. URL: https://coronavirus.jhu.edu/ [accessed 2021-01-05]
4. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
5. Weitz JS, Beckett SJ, Coenen AR, Demory D, Dominguez-Mirazo M, Dushoff J, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. Nat Med 2020 Jun;26(6):849-854. [doi: 10.1038/s41591-020-0895-3] [Medline: 32382154]
6. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
7. CDC COVID-19 Response Team. Geographic Differences in COVID-19 Cases, Deaths, and Incidence - United States, February 12-April 7, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):465-471 [FREE Full text] [doi: 10.15585/mmwr.mm6915e4] [Medline: 32298250]
8. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
10. Lescure FX, Bouadma L, Nguyen D, Parisey M, Wicky PH, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis 2020 Jun;20(6):697-706 [FREE Full text] [doi: 10.1016/S1473-3099(20)30200-0] [Medline: 32224310]
11. Lai CC, Wang CY, Wang YH, Hsueh SC, Ko WC, Hsueh PR. Global epidemiology of coronavirus disease 2019 (COVID-19): disease incidence, daily cumulative index, mortality, and their association with country healthcare resources and economic status. Int J Antimicrob Agents 2020 Apr;55(4):105946 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105946] [Medline: 32199877]
12. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020 Apr;8(4):e480 [FREE Full text] [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]
13. Zhang Z, Yao W, Wang Y, Long C, Fu X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. J Infect 2020 Jul;81(1):147-178 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.018] [Medline: 32209384]

14. COVID-19/time_series_covid19_confirmed_global.csv. GitHub. URL: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv [accessed 2021-01-06]
15. Countries by Density 2020. World Population Review. URL: https://worldpopulationreview.com/countries/countries-by-density/ [accessed 2021-01-06]
16. Wang XS, Wu J, Yang Y. Richards model revisited: validation by and application to infection dynamics. J Theor Biol 2012 Nov 21;313:12-19. [doi: 10.1016/j.jtbi.2012.07.024] [Medline: 22889641]
17. Rothman KJ. Modern epidemiology. Boston, MA: Little Brown & Co; 1986.

Abbreviations

SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 31.07.20; peer-reviewed by LA Lee, W Zhang; comments to author 20.11.20; revised version received 26.11.20; accepted 14.12.20; published 27.01.21.

Please cite as:
Lin SH, Fu SC, Kao CLM
Using the Novel Mortality-Prevalence Ratio to Evaluate Potentially Undocumented SARS-CoV-2 Infection: Correlational Study
JMIR Public Health Surveill 2021;7(1):e23034
URL: http://publichealth.jmir.org/2021/1/e23034/
doi:10.2196/23034
PMID:33332282

Original Paper

Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study

Jay Amol Bapaye[1,2*], MBBS, MD; Harsh Amol Bapaye[2,3*]

[1]Department Of Internal Medicine, Rochester General Hospital, Rochester, NY, United States

[2]Foundation for Research and Education in Endoscopy, Pune, India

[3]Byramjee Jeejeebhoy Medical College and Sassoon General Hospital, Pune, India

[*]all authors contributed equally

Corresponding Author:
Jay Amol Bapaye, MBBS, MD
Department Of Internal Medicine
Rochester General Hospital
1425 Portland Avenue
Rochester, NY, 14621
United States
Phone: 1 5855370133
Email: jaybapaye@gmail.com

Abstract

Background: The risks of misinformation on social networking sites is a global issue, especially in light of the COVID-19 infodemic. WhatsApp is being used as an important source of COVID-19–related information during the current pandemic. Unlike Facebook and Twitter, limited studies have investigated the role of WhatsApp as a source of communication, information, or misinformation during crisis situations.

Objective: Our study aimed to evaluate the vulnerability of demographic cohorts in a developing country toward COVID-19–related misinformation shared via WhatsApp. We also aimed to identify characteristics of WhatsApp messages associated with increased credibility of misinformation.

Methods: We conducted a web-based questionnaire survey and designed a scoring system based on theories supported by the existing literature. Vulnerability (K) was measured as a ratio of the respondent's score to the maximum score. Respondents were stratified according to age and occupation, and K_{mean} was calculated and compared among each subgroup using single-factor analysis of variance and Hochberg GT2 tests. The questionnaire evaluated the respondents' opinion of the veracity of coronavirus-related WhatsApp messages. The responses to the false-proven messages were compared using z test between the 2 groups: coronavirus-related WhatsApp messages with an attached link and/or source and those without.

Results: We analyzed 1137 responses from WhatsApp users in India. Users aged over 65 years had the highest vulnerability (K_{mean}=0.38, 95% CI 0.341-0.419) to misinformation. Respondents in the age group 19-25 years had significantly lower vulnerability (K_{mean}=0.31, 95% CI 0.301-0.319) than those aged over 25 years (*P*<.05). The vulnerability of users employed in elementary occupations was the highest (K_{mean}=0.38, 95% CI 0.356-0.404), and it was significantly higher than that of professionals and students (*P*<.05). Interestingly, the vulnerability of healthcare workers was not significantly different from that of other occupation groups (*P*>.05). We found that false CRWMs with an attached link and/or source were marked true 6 times more often than false CRWMs without an attached link or source (*P*<.001).

Conclusions: Our study demonstrates that in a developing country, WhatsApp users aged over 65 years and those involved in elementary occupations were found to be the most vulnerable to false information disseminated via WhatsApp. Health care workers, who are otherwise considered as experts with regard to this global health care crisis, also shared this vulnerability to misinformation with other occupation groups. Our findings also indicated that the presence of an attached link and/or source falsely validating an incorrect message adds significant false credibility, making it appear true. These results indicate an emergent need to address and rectify the current usage patterns of WhatsApp users. This study also provides metrics that can be used by health care organizations and government authorities of developing countries to formulate guidelines to contain the spread of WhatsApp-related misinformation.

(JMIR Public Health Surveill 2021;7(1):e19858) doi:10.2196/19858

KEYWORDS

coronavirus; COVID-19; SARS–CoV–2; WhatsApp; social media; misinformation; infodemiology; infodemic; pandemic; medical informatics

Introduction

COVID-19, commonly referred to as the novel coronavirus disease, was first reported in Wuhan city, Hubei Province, in the People's Republic of China in late December 2019 [1]. Since then, it has spread to over 3.4 million people across 212 countries around the world and claimed the lives of over 239,892 people as of May 2, 2020 [2]. The first case of COVID-19 in India was reported in the state of Kerala on January 30, 2020 [3], and as of May 2, 2020, the number of active COVID-19 cases was 26,167 with 1218 deaths reported [4]. A simple Google search about "coronavirus" yields over 5 billion results. These numbers are in line with the fact that COVID-19 has been declared as not only a pandemic but also an infodemic [5].

The established treatment options available for COVID-19 are limited, and the disease outcomes are worrisome. Such sensitive scenarios may lead people to believe or, even worse, act on information such as unverified remedies for COVID-19 without confirming its authenticity [6]. This might lead to serious adverse events, which mandates limiting misinformation on public platforms.

Social networking sites (SNS) are an important source of communication as well as information due to their easy accessibility, rapid transmission, simple user interface, dependability, and widespread outreach. Along with its advantages, social media also has a dark side of propagation of rumors or misleading information, to the extent that Goolsby [7] demonstrated that social media can be potentially used to orchestrate crime. Kumar et al [8] has defined misinformation as the spread of false information without the intent to deceive. Several studies have evaluated the resourcefulness of Facebook and Twitter during natural disasters and health care crises, including the COVID-19 pandemic [9-11]. Studies have also described problems related to misinformation on Facebook and Twitter, especially during health crises. Sharma et al [12] reported that misleading posts were far more popular than the posts dispersing accurate relevant public health information about the Zika virus disease. Kouzy et al [10] reported alarming rates of COVID-19–related misinformation being disseminated on Twitter [10]. The social media platform WhatsApp has over 2 billion users across 180 countries, enabling global peer-to-peer and mass communication [13], with over 400 million users in India alone [14]. One in two Indians received some sort of misinformation via WhatsApp or Facebook within the span of merely 30 days in the year 2019 [15]. This kind of misinformation, especially related to health care, on a platform with minimal surveillance such as WhatsApp could have far-reaching adverse consequences. Thus, efforts to better identify and potentially rectify this issue are crucial.

Dodda et al [16] reported that certain age groups could be more vulnerable to misinformation via WhatsApp. Our study aimed to identify specific demographic groups that may be more vulnerable towards misinformation via WhatsApp. Vulnerability of users towards content on WhatsApp may be influenced by several factors, ranging from the mere appearance of a message to different types of evidence included in a message [8].

Our study aimed to analyze the effect of evidence included in a message on the credibility of a message shared on WhatsApp.

Methods

Study Design

This study involved a web-based questionnaire survey conducted at Pune, Maharashtra, India. The study was specifically designed to perform a comparative analysis of vulnerability of user groups to misinformation based on a function of 9 parameters. Our methodology was novel and not based on the existing relevant literature.

Research Model (Objective 1)

The study was designed to calculate a value K, defined as the risk of a WhatsApp user to get exposed to, accept, act on, or share unverified data regarding COVID-19 via WhatsApp. K was a value defined specifically for the purpose of this study. In a diverse population, several parameters may affect K. We identified 9 such parameters (P) framed by theories based on the existing literature, as detailed below.

P_1—Duration of WhatsApp Usage Affects K

According to an article published in Business Today, India reported an 87% increase in the use of SNS during the first week of the COVID-19 lockdown as compared to the previous week [17]. Ünal [18] and Junco [19] reported a correlation between the duration of the use of SNS and the user's investment of physical and psychological energies. In accordance with these reports, we posit that the duration of WhatsApp usage affects K.

P_2—Priority of WhatsApp for Communication Affects K

Vosoughi et al [20] reported that even false information reaches more people if it is shared on a peer-to-peer basis rather than being broadcasted by a few users.

The social normative theory states that accepting and sharing content aligned with one's peers' beliefs is attractive regardless of its veracity [8,21], implying that users tend to believe more in information shared through peers. In the context of WhatsApp, sharing of information on a peer-to-peer basis is equivalent to its use as a social messenger. Hence, we posit that the priority of WhatsApp for communication affects K.

P_3—Priority of WhatsApp for News and Information Affects K

Kumar and Shah [8] reported that the role of SNS for news and information has been on the rise as compared to traditional news sources. Out of 67% US adults using SNS as a source of news, 20% reported a relatively high usage frequency [22]. This confirms the dominance and dependence of SNS as news sources. Although SNS are great tools for disseminating information, they have a potential disadvantage of limited verification [23]. Thus, the level of priority placed in the use of SNS for news and information affects K. Hence, we posit that the priority of WhatsApp for news and information affects K and that the use of SNS for COVID-19–related updates affects K.

P_4—Use of SNS for News Updates Regarding COVID-19 Affects K

Misinformation about COVID-19 is prevalent across all SNS [5]. Individuals using WhatsApp for COVID-19–related news updates are at risk of being exposed to misinformation and of propagating it via WhatsApp, thus affecting K. Even if a user does not use WhatsApp for news updates regarding COVID-19 but uses other SNS for the same purpose, they could still be exposed to COVID-19–related misinformation via these platforms. This misinformation could be further shared by the user via WhatsApp, thus affecting K.

P_5—Trust Placed in COVID-19–Related Information Received via WhatsApp Affects K

The American Psychological Association defines trust as "reliance on or confidence in the dependability of someone or something" [24]. Reliance on or confidence in WhatsApp directly affects the acceptance of information provided therein. Hence, we posit that trust placed in COVID-19–related information received via WhatsApp affects K.

P_6—Fact Check Rate Affects K

Misinformation has become a major menace on all SNS, including WhatsApp. SNS are reported to be untrustworthy during critical situations [25]. Multiple organizations such as boomlive.in [26] and factcheck.org [27] aid users to distinguish between true and false information received on SNS. Fact-checking could thus limit blind acceptance of false information and prevent its inadvertent dissemination. Actions based on such misinformation could also be prevented. Hence, we posit that the fact check rate affects K.

P_7—Forward Rate Affects K

SNS like Facebook and Twitter implement determinants such as likes, comments, or shares of posts to analyze user engagement [19]. WhatsApp is primarily a messenger with a user interface designed specifically for communication; hence, determinants of user engagement of other SNS cannot be directly applied to it. The user interface of WhatsApp allows people to be members of groups comprising as many as 256 persons. Users can thus forward messages in a peer-to-peer fashion and simultaneously to multiple groups. This allows WhatsApp users the potential to rapidly disseminate information, regardless of its authenticity, to a staggering number of individuals. Hence, we posit that the forward rate of WhatsApp messages affects K.

P_8—User's Ability to Discern a WhatsApp Message as True or False Affects K

Several studies have conducted experiments to measure the ability of humans to detect false information and have shown that humans are not particularly good at discerning false information from true. Kumar and Shah [8] reported that humans correctly identified a hoax merely 66% of times. False information would not have any influence if readers were able to tell that it is false [8].

Hence, we posit that the user's ability to discern a WhatsApp message as true or false affects K.

P_9—Actions Taken in Response to a Message Affects K

SNS have the potential to motivate users to act based on any information provided to them. For instance, a rumor on WhatsApp claiming that public transport will be made available to transport over 1000 workers back to their hometowns resulted in a gathering of 700-800 people on a railway station in India [28]. Based on these reports, we posit that actions taken in response to a message affects K.

These 9 parameters P_1-P_9 were used to formulate the study questionnaire.

Data Collection

We conducted a web-based questionnaire survey. Its design was influenced by models used by Oyero et al [23] and Dodda et al [16]. The questionnaire was created using Google Forms in 2 languages—English and vernacular (Marathi), as the target population was fluent in either of these languages (see Multimedia Appendix 1).

It consisted of 28 questions: 5 about demographics, 12 regarding general and COVID-19–related WhatsApp usage, and 10 regarding coronavirus related WhatsApp messages (CRWMs) that the respondents were asked to identify as true or false. These 10 messages were selected randomly from a pool of 30 WhatsApp messages commonly shared in India, sourced from AlJazeera, Reuters, and Google Images [29,30]. These questions were specifically included to reduce the risk of biased reporting through the survey and increase its accuracy. The final question was open-ended and enquired the respondent's opinion about WhatsApp as an information tool in the current COVID-19 pandemic. Of the 28 questions, 21 (Q_K) were framed to evaluate the 9 parameters (P_1-P_9) that affect K as detailed in the section "Research Model (Objective 1)." The questionnaire was initially distributed among a WhatsApp group comprising users from diverse demographics. The first 10 respondents were consulted to check for errors and obtain feedback regarding understandability of the selected questions. Necessary modifications were made accordingly, and pilot responses were excluded from the analysis. The final questionnaire was shared on the following social media platforms: WhatsApp, Facebook, Twitter, Instagram, and LinkedIn. All responses were voluntary and anonymous. The survey was conducted from April 8 to April 13, 2020. Data were compiled using Microsoft Excel (Microsoft Corp.).

Scoring System for Calculation of K

The 21 questions (Q_K) evaluating the 9 parameters (P_1-P_9) were sorted into 9 groups based on the parameter being tested (Table 1). Numerical scores were assigned to the response options of each of the 9 question groups. Responses that were assigned a score "0" had no impact on K. For other responses, higher the value of the score, greater was the impact of the response on K.

Table 1. Scoring system for calculation of K values.

Question group no.[a]	Response options (score)				
1. Duration of WhatsApp use	0-30 minutes (1)	30 minutes to 1 hour (2)	1-2 hours (3)	>2 hours (4)	N/A[b]
2. Communication[c]	Low (1)	Moderate (2)	High (3)	N/A	N/A
3. Information[d]	Low (1)	Moderate (2)	High (3)	N/A	N/A
4. Source for COVID-19–related updates[e]	Neither social media nor WhatsApp (0)	Social Media but no WhatsApp (2)	WhatsApp (4)	N/A	N/A
5. Trust[f]	0% (0)	25% (1)	50% (2)	75% (3)	100% (4)
6. Forwards[g]	0-2 (1)	3-5 (2)	6-8 (3)	More than 8 (4)	N/A
7. Fact-check[h]	75-100% (1)	50-75% (2)	25-50% (3)	0-25% (4)	Never (5)
8. Incorrect responses[i]	0-2 (1)	3-5 (2)	6-8 (3)	More than 8 (4)	N/A
9. Actions or opinions[j]	Never considered using (0)	Considered but not used (1)	Used once (2)	Using regularly and recommending (3)	N/A

[a]Question group numbers 1-9 represent questions testing parameters P_1-P_9, respectively.

[b]Not applicable.

[c]Priority of WhatsApp usage for communication.

[d]Priority of WhatsApp usage for information.

[e]Sources used by respondents to stay updated with COVID-19–related information.

[f]Trust placed in COVID-19–related WhatsApp messages.

[g]Number of forwards about COVID-19 per day.

[h]Percentage of COVID-19–related messages that were fact-checked.

[i]Number of incorrect responses provided while discerning a WhatsApp message as true or false.

[j]Actions taken by respondents in line with unverified information received via WhatsApp.

The 7 parameters (P_1-P_7) were tested by question groups 1-7, each consisting of 1 question. Each of these questions was assigned an individual score. Question group 4 was assigned a different scoring system as it recorded the source for COVID-19–specific information, which has a stronger impact on K. COVID-19–related misinformation available on traditional news sources is negligible. Hence, respondents referring to only these sources (ie, sources other than social media and WhatsApp for COVID-19–related information) are not at risk of misinformation. As this does not affect K, the score assigned to this response was "0." Respondents referring to WhatsApp as a source of COVID-19–related information have a high impact on K; their responses were therefore assigned a score of "4." Those referring to social media but not WhatsApp still carry the risk of sharing misinformation via WhatsApp, thereby moderately impacting K; a score of "2" was therefore assigned for their responses.

A total of 10 questions out of the 21 Q_K tested user ability to discern a WhatsApp message as true or false (P_8). They were grouped together in question group 8. The total number of incorrect responses was calculated for these 10 questions, and respondents were thus assigned scores.

Four of the 21 Q_K questions recorded the actions taken by a user in response to unverified treatment options (herbal, homeopathic, ayurvedic, and home remedies) for COVID-19 received via WhatsApp. These questions were grouped together as they measured P_9. Taking an action in response to misinformation can lead to adverse health effects. A higher risk is therefore anticipated in this case in comparison with only exposure, acceptance, or sharing of misinformation. Since each of the 4 questions had a higher impact on K, they were scored individually.

One numerical score represented each question group (1-8), and 4 scores represented question group 9. A total of 12 scores were thus recorded, and their sum was calculated (S_{cal}) for each respondent. The maximum possible score for any respondent was 46 (S_{max}). The scoring system intended to compare subgroups of the same sample. S_{cal} of one respondent in relation to the S_{cal} of another respondent was measured, whereas the standalone value of S_{cal} had negligible relevance.

Classification of Age and Occupation Subgroups

The study sample was classified into 6 groups according to the respondents' age in years: under 18, 19-25, 26-35, 36-50, 51-65, and over 65. In addition, occupations of working adults were classified based on The International Standard Classification of Occupations [31]. Students and retired individuals were considered as separate groups.

This study was focused on WhatsApp usage during the COVID-19 pandemic—a global health care crisis. Therefore, health care workers (HCWs) were considered a separate occupation subgroup, as their usage patterns could have been unique.

Computation of K

K was calculated as a ratio S_{cal}/S_{max} for each respondent, and K_{mean} (±95% CI) was calculated for each age and occupation subgroup. K_{mean} was compared for age groups and occupations separately. Since K_{mean} was compared within the same sample, relative values of K_{mean} and not its absolute value were relevant in this study.

Research Model (Objective 2)

Dodda et al [32] demonstrated that the presence of background evidence may increase the credibility of a message on WhatsApp. The Merriam-Webster dictionary defines credibility as "the quality or power of inspiring belief" [32]. In this study, we aimed to analyze factors that could affect the credibility of messages on WhatsApp during the COVID-19 pandemic.

The questionnaire included 10 commonly shared WhatsApp messages related to COVID-19, which the respondents were asked to rate as follows: definitely true, maybe true, maybe false, or definitely false. These were the same 10 questions that were used to test P8. The authenticity of these 10 messages was confirmed with data from authorized sources [29,30,33-44]. Of the 10 messages analyzed, 8 were false messages; these were selected and subcategorized as those with background evidence containing link or source: (N_s) and those without: (N_x).

The total number of instances that a false message was marked by the respondents as definitely true or maybe true was measured in each group (N_S and N_X). The number of such instances was recorded as a_s in group N_S and a_x in group N_X.

a_s/N_s=Number of instances that a false message was marked definitely true or maybe true for every message with background evidence.

a_x/N_x=Number of instances that a false message was marked definitely true or maybe true for every message without background evidence.

Comparison of these ratios detected whether an association exists between the presence of background evidence in a false WhatsApp message and the number of instances of it being marked as true.

Statistical Analyses

All calculations were performed on SPSS software (Build no. 1.0.0.1347; IBM Corp.). The level of significance was set at P=.05 for all variables.

For research objective 1, a single-factor analysis of variance (ANOVA) was applied to compare K_{mean} values in the age and occupation subgroups separately [45]. Hochberg GT2 test was conducted to perform a post hoc analysis since variances were homogenous according to Hartley's F_{max} test and sample sizes in each subgroup were different [46,47].

For research objective 2, a_s/N_s and a_x/N_x values were calculated and compared using 2-tailed z-test.

Results

Demographic Characteristics of the Study

We obtained a total of 1191 responses to our survey. Of these, 54 responses were not considered as the respondents either did not use WhatsApp or were not presently living in India. These responses were eliminated, and the remaining 1137 responses obtained from respondents across 20 Indian states and union territories were further analyzed. Demographic characteristics of the study sample are described in Table 2. No specific measures were taken to control the demographics of the respondents. Hence, the sample distribution across subgroups was varied.

Table 2. Demographic characteristics of the study sample.

Demographic	Value (N=1137), n (%)
Age (years)	
Below 18	25 (2.20)
19-25	395 (34.74)
26-35	160 (14.07)
36-50	291 (25.59)
51-65	235 (20.67)
Above 66	31 (2.73)
Gender	
Men	648 (56.99)
Women	488 (42.92)
Gender fluid	1 (0.09)
Occupation	
HCW[a]	291 (25.59)
Other professionals	276 (24.27)
Manager	18 (1.58)
Service and sales	89 (7.83)
Elementary occupation	72 (6.33)
Clerical occupation	7 (0.62)
Student	345 (30.34)
Retired and unemployed	39 (3.43)

[a] HCW: health care worker.

Of the 1137 respondents, 648 (56.99%) were men, 488 (42.92%) were women, and 1 (0.09%) was gender fluid. Age group distribution showed that 25 of the 1137 (2.20%) respondents were under 18 years, 395 (34.74%) were between 19 and 25 years, 160 (14.07%) were between 26 and 35 years, 291 (25.59%) were between 36 and 50 years, 235 (20.67%) were between 51 and 65 years, and 31 (2.73%) were over 65 years. Occupation-wise distribution showed 345 of the 1137 (30.34%) were students, 291 (25.59%) were HCWs, 276 (24.27%) were professionals, 89 (7.83%) were service and sales workers, 72 (6.33%) were involved in elementary occupations, 39 (3.43%) were retired and unemployed, 18 (1.58%) were managers, and 7 (0.62%) were clerical workers.

Results for Objective 1

Usage patterns of WhatsApp in the sample population are detailed in Table 3. Active use of WhatsApp as a source of information regarding COVID-19 was confirmed by 355 of the 1137 (31.22%) respondents. Moreover, 657 (57.78%) respondents demonstrated trust in CRWMs, of which 139 (21.15%) trusted more than 50% of CRWMs. The presence of an attached link and/or source demonstrated an increase in CRWM credibility for 859 of the 1137 (75.55%) respondents. Furthermore, of the 1137 respondents, 151 (13.28%) never fact-checked CRWMs before forwarding them and 164 (14.43%) forwarded 3 or more CRWMs per day. The various actions taken or considered in response to CRWMs are detailed in Table 4.

Table 3. WhatsApp usage patterns in the context of COVID-19.

Variable	Value (N=1137), n (%)
Duration of WhatsApp use	
0-30 minutes	136 (11.96)
30 minutes to 1 hour	327 (28.76)
1-2 hours	391 (34.39)
>2 hours	283 (24.89)
WhatsApp as a source of information regarding COVID-19	
Yes	355 (31.22)
No	782 (68.78)
Trust in CRWM[a] (%)	
0	479 (42.13)
25	519 (45.65)
50	121 (10.64)
75	16 (1.41)
100	2 (0.18)
Factors accompanying CRWM[b]	
Attached link and/or source	859 (75.55)
Sender's authenticity	374 (32.89)
Attached multimedia	77 (6.77)
None of the above	164 (14.42)
Fact-check rate[c] (%)	
Never	151 (13.28)
1-25	114 (10.03)
25-50	73 (6.42)
50-75	132 (11.61)
75-100	667 (58.66)
Forward rate[d]	
0-2	973 (85.58)
3-5	83 (7.30)
6-8	26 (2.29)
More than 8	55 (4.84)

[a]CRWM: COVID-19–related WhatsApp message.

[b]Users could select more than 1 factor which increased their trust in a CRWM.

[c]CRWMs fact-checked before forwarding.

[d]CRWMs forwarded per day.

Table 4. Actions taken in response to information regarding COVID-19 received vis WhatsApp.

Actionable measures	Value (N=1137), n (%)
Social distancing	
Never considered using	46 (4.05)
Considered but not used	138 (12.14)
Used once	291 (25.59)
Using regularly and recommending	662 (58.22)
Masks	
Never considered using	50 (4.40)
Considered but not used	188 (16.53)
Used once	329 (28.94)
Using regularly and recommending	570 (50.13)
Allopathic remedies	
Never considered using	449 (39.49)
Considered but not used	398 (35)
Used once	131 (11.52)
Using regularly and recommending	159 (13.98)
Herbal remedies	
Never considered using	776 (68.25)
Considered but not used	271 (23.83)
Used once	58 (5.10)
Using regularly and recommending	32 (2.81)
Ayurvedic remedies	
Never considered using	740 (65.08)
Considered but not used	302 (26.56)
Used once	62 (5.45)
Using regularly and recommending	33 (2.9)
Homeopathic remedies	
Never considered using	788 (69.31)
Considered but not used	268 (23.57)
Used once	60 (5.28)
Using regularly and recommending	21 (1.85)
Home remedies	
Never considered using	803 (70.62)
Considered but not used	200 (17.59)
Used once	88 (7.74)
Using regularly and recommending	46 (4.05)

Recommended practices such as social distancing and use of masks were not being followed by 184 (16.19%) and 238 (20.93%) of the 1137 respondents, respectively. Responses regarding the use of unverified treatment options revealed that 90 of the 1137 (7.91%) respondents had used herbal, 81 (7.13%) had used homeopathic, 95 (8.35%) had used ayurvedic, and 134 (11.79%) had used home remedies at least once.

K_{mean} value was found to be the lowest for the under-18 years age subgroup (0.31, 95% CI 0.264-0.356) and the highest for the over-65 years age subgroup (0.38, 95% CI 0.341-0.419) (Figure 1). Statistically significant differences were found among K_{mean} values for the 6 age subgroups (P<.001, single-factor ANOVA). Post hoc analysis revealed that respondents in the age group 19-25 years had a significantly lower K_{mean} value (0.31, 95% CI 0.301-0.319) than all respondents aged over 25 years (Table 5).

Figure 1. Graphical representation of K_{mean} values (± 95% CI) across age subgroups.

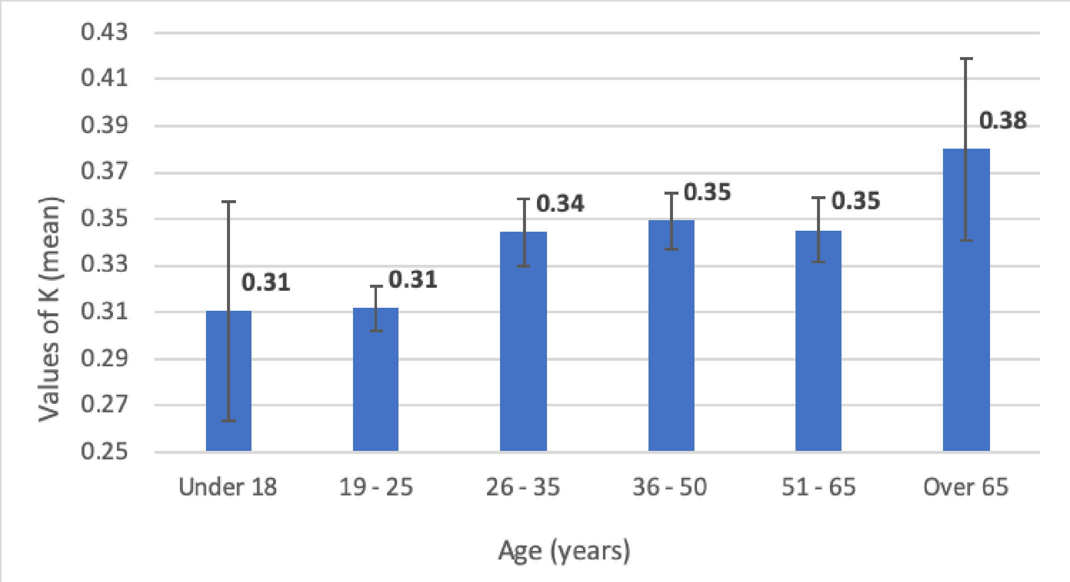


Table 5. Post hoc comparisons (*P* values) among Kmean values of age subgroups using Hochberg GT2 test.

Age group (years)	Under 18	19-25	26-35	36-50	51-65	Over 65
Under 18	>.99	>.99	.85	.65	.80	.15
19-25	—[a]	>.99	.009	0	.001	.005
26-35	—	—	>.99	>.99	>.99	.68
36-50	—	—	—	>.99	>.99	.82
51-65	—	—	—	—	>.99	.68
Over 65	—	—	—	—	—	>.99

[a]Not applicable.

K_{mean} value was the lowest for managers (0.33, 95% CI 0.286-0.374) and the highest for elementary occupations (0.38, 95%, CI 0.356-0.404) (Figure 2). Statistically significant differences were observed among K_{mean} values measured for the 8 occupation groups (*P*=.023, single-factor ANOVA). Post hoc analysis revealed a significantly higher K_{mean} value in the elementary occupation subgroup than in the professionals and students' subgroups (Table 6).

Figure 2. Graphical representation of K_{mean} values (± 95% CI) across occupation subgroups.

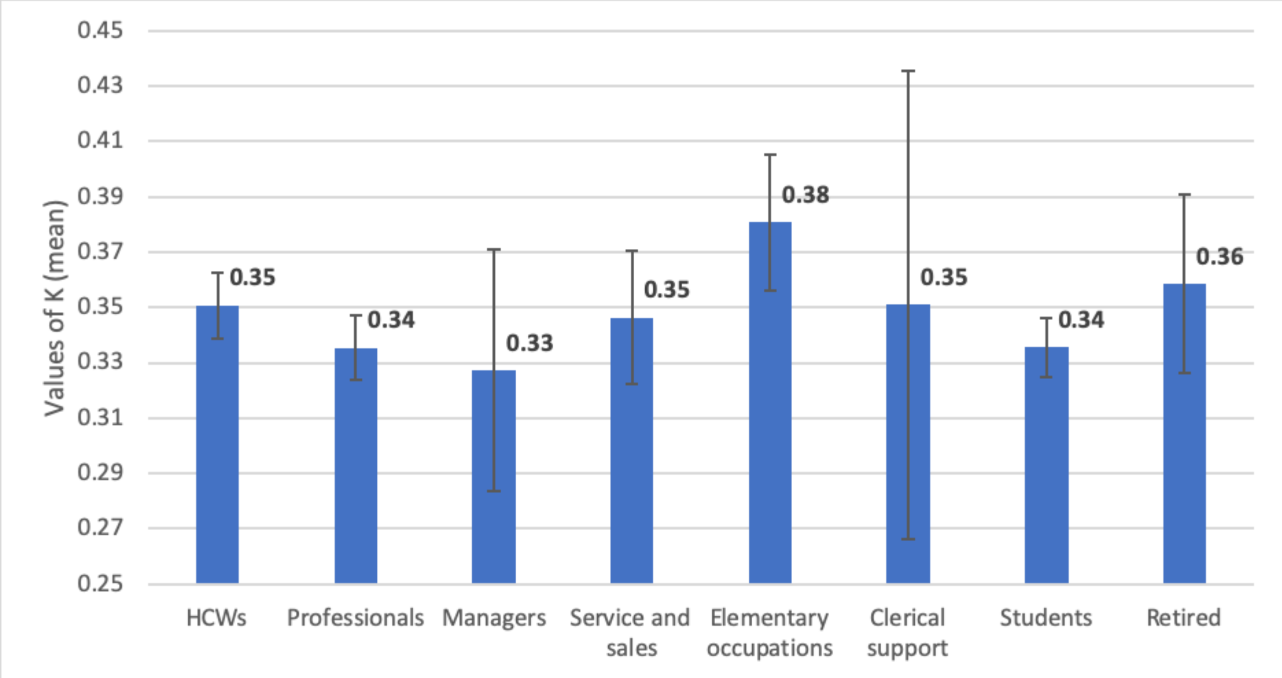


Table 6. Post hoc comparisons (*P* values) among Kmean values of occupation subgroups using Hochberg GT2 test.

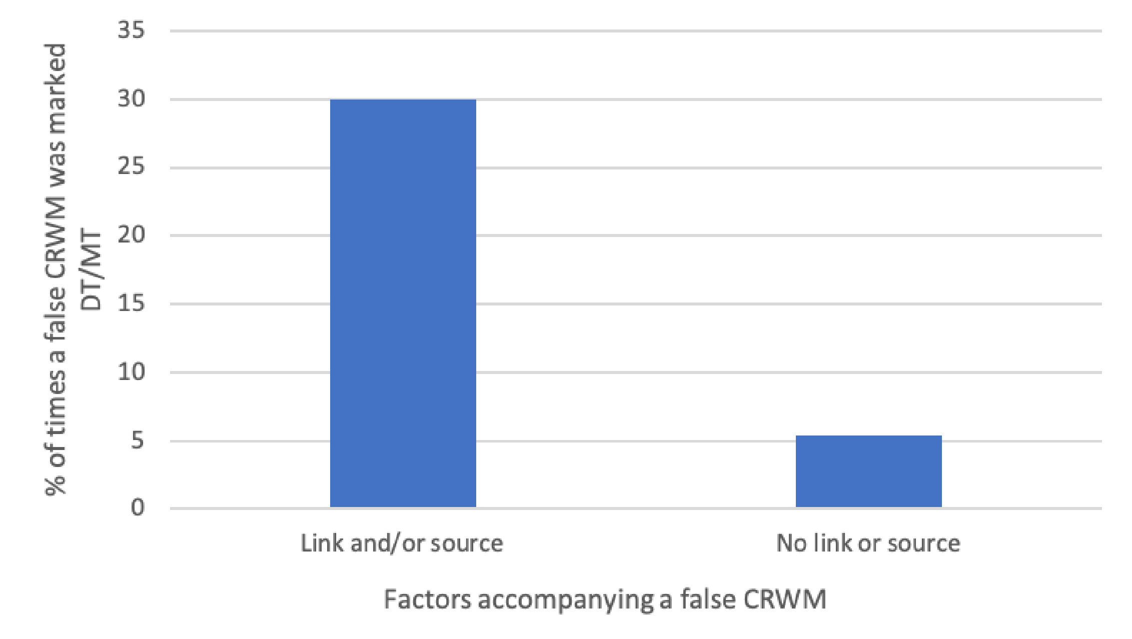
Group	HCW[a]	Professionals	Service[b]	Managers	Elementary[c]	Clerical[d]	Students	Retired
HCW	>.99	.87	>.99	>.99	.52	>.99	.83	>.99
Professionals	__[e]	>.99	>.99	>.99	.02	>.99	>.99	.99
Service	—	—	>.99	>.99	.62	>.99	>.99	>.99
Managers	—	—	—	>.99	.74	>.99	>.99	>.99
Elementary	—	—	—	—	>.99	>.99	.02	>.99
Clerical	—	—	—	—	—	>.99	>.99	>.99
Students	—	—	—	—	—	—	>.99	.997
Retired	—	—	—	—	—	—	—	>.99

[a]HCW: health care worker.
[b]Service and sales workers.
[c]Elementary occupations.
[d]Clerical support workers.
[e]Not applicable.

Results for Objective 2

Presence of background evidence in a false message demonstrated a significant impact on the perceived credibility of the WhatsApp user (a_s/N_s=0.299 vs a_x/N_x=0.053; *P*<.001, 2-tailed *z* test). The number of times false messages with background evidence were marked definitely true or maybe true was significantly (nearly 6 times) higher than that for false messages without background evidence (Figure 3).

Figure 3. Number of times a false CRWM was marked as true for every false CRWM with and without background evidence (link and/or source). CRWM: coronavirus-related WhatsApp message, DT: definitely true, MT: maybe true.



Discussion

COVID-19 has caused an unprecedented biopsychosocial crisis. It has been declared not only as a pandemic but also as a worldwide infodemic [5]. Mentions about coronavirus in the form of news, updates, and especially misinformation on SNS are rising exponentially. In this study, we broadly analyzed factors influencing misinformation related to the disease (COVID-19) on WhatsApp. van Velsen et al [25] studied the reliability of Facebook and Twitter during an infective health crisis. They concluded that the majority of their study sample did not find these platforms to be reliable sources of information, and only 11% of the participants used them passively for information [25]. However, 39.92% (454/1137) of our sample considered WhatsApp as a useful information tool during this health crisis. Moreover, our study demonstrates active WhatsApp use by only 31.22% (355/1137) of the sample for COVID-19–related information. This finding suggests an almost 4-fold increase in reliance of individuals on SNS, particularly WhatsApp for news and information in recent times. Although trust levels in traditional sources of news have been higher, 57.78% (657/1137) of our study sample demonstrated a certain level of trust in CRWMs. With every 1 in 2 Indians receiving some false news on WhatsApp or Facebook, the consequences of the trust placed in these platforms could be serious [15].

Community preventive measures against COVID-19, particularly social distancing and using face masks, have proven to be effective and have been strongly recommended by the Government of India [48]; however, our study demonstrated that 16.18% (184/1137) and 20.93% (238/1137) of our study participants, respectively, did not implement these measures. This finding indicates that a considerable proportion of individuals demonstrate reluctance towards authorized recommendations.

Currently, no pharmacological agent has been proven to reduce mortality in COVID-19 cases [6]. Nevertheless, our findings confirmed that 1.85%-7.74% of the respondents were using or had used some form of complementary or alternative medicines for COVID-19 (Table 4). This finding suggests an interesting paradox wherein there is reluctance to accept evidence-based recommendations in favor of easy acceptance of novel but unverified remedies. This could result in serious health-related adverse events [49].

In addition, our study identified certain age groups were at a higher risk of getting influenced by COVID-19–related misinformation via WhatsApp. In particular, users over 65 years demonstrated the highest vulnerability to misinformation, whereas those under 18 years demonstrated the lowest. Vulnerability to misinformation progressively increased with users' age. Limited knowledge of fact-checking resources, unawareness about existing misinformation on WhatsApp, and positive reinforcement from echo chambers could possibly explain these results [8]. This is particularly worrisome considering that older users tend to be more susceptible to the adverse effects of unsupervised remedies and are at a higher risk of severe COVID-19. Dodda et al [16] reported that WhatsApp users aged below 20 years and those above 50 years were the most susceptible to fake news. In contrast, our study findings showed that users aged under 25 years were the least susceptible to misinformation. Users aged between 19 and 25 years were at a significantly lower risk than those over 25 years ($P<.05$). Users in this age group tend to be well versed with SNS, including WhatsApp. Awareness about fact-checking resources and existence of misinformation on WhatsApp could possibly have reduced their vulnerability.

Among the different occupation groups, those employed in elementary occupations demonstrated the highest vulnerability to false news (K_{mean}=0.38). These users were significantly more vulnerable to misinformation than students and professionals

(K_{mean}=0.34 for both, *P*<.05). A possible explanation for this observation could be the lower levels of education among the elementary occupation subgroup. Managers demonstrated the least vulnerability among the sample. However, a statistically significant difference could not be established, presumably due to the small sample size of this subgroup.

The vulnerability among HCWs was found to be consistent with all other occupation groups. During a health care crisis like the COVID-19 pandemic, HCWs are considered reliable sources of information by the community. Kumar et al [8] concluded that if false information is purposefully made to appear genuine, it can deceive trained and casual readers alike. COVID-19 is an evolving global health crisis wherein the mortality rate is high, established treatment options are limited, and HCWs are at a high risk of contracting the disease [6,50]. This vulnerability, coupled with a desire to treat patients, could possibly lead even HCWs to accept novel but unverified information. Such medical misinformation inadvertently propagated by HCWs may have unbeknownst consequences on the community.

Dodda et al [16] reported that background evidence and trust in organizations and individuals can make people believe in information received on SNS. Kumar et al [8] reported that that lengthy and well-referenced hoaxes were frequently misjudged to be true. In our study, for more than half the sample (659/1137, 57.95%), an attached link or source was a factor associated with increasing trust in a CRWM. Presence of background evidence was strongly associated with an erroneously higher trust of a false CRWM (*P*<.001). This finding indicates that individuals are likely to place emphasis on the presence of background evidence in a false CRWM in addition to its content while judging its veracity. Identification of such factors and ways to limit seemingly credible misinformation on SNS are pertinent questions that need to be evaluated in this digital age.

SNS usage patterns in developing countries are broadly comparable, although regional preferences regarding the choice of SNS may be different [51]. WhatsApp is a popular SNS globally with over 2 billion users worldwide, of which 400 million users are from India [13,14]. In our web-based survey study on WhatsApp users from India, all survey responses obtained were anonymous and voluntary. The study sample was representative of all age groups and majority of the occupation groups. We, therefore, are of the opinion that the findings of this study are representative of WhatsApp usage patterns in developing countries.

Our study had certain limitations. First, it was a survey-based study wherein data was self-reported; therefore, the sample may be tainted by observer bias (Hawthorne effect [52]). Second, the number of respondents in certain age and occupation subgroups were fewer, which could increase the risk of type 2 error during subgroup analysis. Third, the respondents' opinions regarding CRWMs may have been formed through sources other than WhatsApp. The questionnaire was specifically designed to include the terms "WhatsApp" and "coronavirus" in the questions to address this limitation. Fourth, this study tested 4 message-based factors that could affect users' trust in CRWMs. However, it was possible that other personal, regional, or community-based factors could also affect the users' trust. To ensure generalizability of the results of this study, we chose not to analyze these belief systems. Fifth, the research model has limited support from the existing literature owing to its specific nature. It could be implemented in further similar studies to improve its validity. Finally, although well-referenced, the parameters P_1-P_9 used in this study are based on theories supported by the existing literature; hence, the results should be interpreted as correlational.

The results of this study present several challenges regarding the current COVID-19 pandemic for WhatsApp, fact-checking bodies, health organizations, and government authorities in developing countries. Cuan-Baltazar et al [53] found that the use of the internet during a pandemic is a risk to public health and recommend that the authorities should develop strategies to regulate online health information without censuring the public. Disasters mandate the circulation of timely information, which may at times compromise its genuineness. SNS have traditionally faced the issue of credibility at the cost of rapid transmission of information [23,54]. This may reduce the resourcefulness of a highly efficient communication tool such as WhatsApp, as seen in recent trends of decreasing trust in SNS in developed countries [51,55].

To address these challenges, health and government authorities in developing countries could collaborate with WhatsApp to develop methods to authenticate and tamper-proof messages from official sources. Leadership in health care organizations could actively work towards addressing digital awareness among health care workers who are the anchor points of information for the rest of the community. Fact-checking organizations could increase their presence on and integrate with SNS to improve their resourcefulness. Authorities could undertake awareness campaigns to educate users of SNS to recognize misinformation. WhatsApp could allow users to report messages containing suspected misinformation to allow necessary measures to be undertaken. WhatsApp, if used as a platform to broadcast validated information to a userbase as large as 400 million while providing measures enabling them to discern genuine news from misinformation, could revolutionize disaster management in a developing country.

In conclusion, our study analyzed factors influencing COVID-19–related misinformation among WhatsApp users in a developing country. Older adults (above 65 years) were more vulnerable to misinformation circulated on WhatsApp, as were individuals employed in elementary occupations. HCWs were also not spared from the influence of misinformation and were found to be as vulnerable as any other occupations. Presence of background evidence in a false CRWM was strongly linked to an increase in its credibility. These findings may provide important insights to health organizations and government authorities of developing countries to formulate suitable guidelines to contain the COVID-19 infodemic. Further experimental studies about parameters tested in this study could be considered.

Acknowledgments

The authors express gratitude towards the Foundation for Research and Education in Endoscopy for critical scientific review of the manuscript and for providing financial support to conduct this study. They are also grateful to Ms. Sonia Patwardhan for providing assistance with statistical analyses. JB was affiliated with the Foundation for Research and Education in Endoscopy at the time of the study and is currently affiliated with the Department of Internal Medicine, Rochester General Hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaire.
[PDF File (Adobe PDF File), 2905 KB - publichealth_v7i1e19858_app1.pdf]

References

1. Li H, Liu L, Zhang D, Xu J, Dai H, Tang N, et al. SARS-CoV-2 and viral sepsis: observations and hypotheses. The Lancet 2020 May;395(10235):1517-1520. [doi: 10.1016/s0140-6736(20)30920-x]
2. World - Coronavirus Cases. Worldometer. 2020 May 02. URL: https://www.worldometers.info/coronavirus/ [accessed 2020-05-02]
3. Update on Novel Coronavirus: one positive case reported in Kerala. Ministry of Health and Family Welfare. 2020 Jan 30. URL: https://pib.gov.in/PressReleasePage.aspx?PRID=1601095 [accessed 2021-01-18]
4. India - Coronavirus. Worldometer. 2020 May 03. URL: https://www.worldometers.info/coronavirus/country/india/ [accessed 2020-05-03]
5. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
6. Zagury-Orly I, Schwartzstein RM. Covid-19 — A reminder to reason. N Engl J Med 2020 Jul 16;383(3):e12. [doi: 10.1056/nejmp2009405]
7. Goolsby R. Social media as crisis platform: The future of community maps/crisis maps. ACM Trans Intell Syst Technol 2010 Oct;1(1):1-11. [doi: 10.1145/1858948.1858955]
8. Kumar S, Shah N. False information on web and social media: A survey. arXiv. Preprint posted online April 23, 2018 2018 [FREE Full text]
9. Kavota J, Kamdjoug J, Wamba S. Social media and disaster management: Case of the north and south Kivu regions in the Democratic Republic of the Congo. Int J Inf Manage 2020 Jun;52:102068. [doi: 10.1016/j.ijinfomgt.2020.102068]
10. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
11. Niles MT, Emery BF, Reagan AJ, Dodds PS, Danforth CM. Social media usage patterns during natural hazards. PLoS One 2019;14(2):e0210484 [FREE Full text] [doi: 10.1371/journal.pone.0210484] [Medline: 30759111]
12. Sharma M, Yadav K, Yadav N, Ferdinand K. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
13. About WhatsApp. WhatsApp. URL: https://www.whatsapp.com/about/ [accessed 2020-05-02]
14. Tech Desk. WhatsApp now has 400 million active users in India. The Indian Express. 2019 Jul 26. URL: https://indianexpress.com/article/technology/social/whatsapp-now-has-400-million-active-users-in-india-confirms-company-5854881/ [accessed 2020-06-20]
15. 1 in 2 Indians receiving fake news via Facebook, WhatsApp. The Economic Times. 2019 Apr 09. URL: https://economictimes.indiatimes.com/news/elections/lok-sabha/india/1-in-2-indians-receiving-fake-news-via-facebook-whatsapp/articleshow/68798051.cms [accessed 2020-05-04]
16. Dodda TP, Rakesh D. Countering misinformation (fake news) in India: solutions and strategies. Factly Media & Research (Factly) and the Internet and Mobile Association of India (IAMAI). 2019 Feb 22. URL: https://factly.in/wp-content/uploads//2019/02/Countering-Misinformation-Fake-News-In-India.pdf [accessed 2021-01-18]
17. Coronavirus: 87% increase in social media usage amid lockdown; Indians spend 4 hours on Facebook, WhatsApp. Business Today. 2020 Mar 30. URL: https://www.businesstoday.in/technology/news/coronavirus-87-percent-increase-in-social-media-usage-amid-lockdown-indians-spend-4-hours-on-facebook-whatsapp/story/399571.html [accessed 2020-03-31]
18. Ünal S. The effect of social media use to the time spent with family members. International Journal of Eurasia Social Sciences 2018 Mar;9(31):01-78 [FREE Full text]
19. Junco R. The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. Computers & Education 2012 Jan;58(1):162-171 [FREE Full text] [doi: 10.1016/j.compedu.2011.08.004]

20. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
21. Reiss S. Intrinsic and extrinsic motivation. Teaching of Psychology 2012 Mar 20;39(2):152-156. [doi: 10.1177/0098628312437704]
22. Shearer E, Gottfried J. News use across social media platforms 2017. Pew Research Centre - Journalism & Media Internet. 2017. URL: https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/ [accessed 2020-05-02]
23. Oyero O. The use and believability of social networks news among Nigerian youths. Covenant Journal of Communication 2013 Jun;1(1):1-55 [FREE Full text]
24. "Trust". APA Dictionary Of Psychology.: American Psychological Association URL: https://dictionary.apa.org/trust [accessed 2020-05-03]
25. van Velsen L, van Gemert-Pijnen JEWC, Beaujean DJMA, Wentzel J, van Steenbergen JE. Should health organizations use web 2.0 media in times of an infectious disease crisis? An in-depth qualitative study of citizens' information behavior during an EHEC outbreak. J Med Internet Res 2012 Dec 20;14(6):e181 [FREE Full text] [doi: 10.2196/jmir.2123] [Medline: 23257066]
26. BOOM FACT Check. 2020. URL: https://www.boomlive.in/ [accessed 2020-05-03]
27. FactCheck - A Project of The Annenberg Public Policy Center. 2020. URL: https://www.factcheck.org/ [accessed 2020-05-03]
28. Hindustantimes. Mumbai: Inside story of how 'malicious' rumours caused migrant's frenzy. The Hindustan Times. 2020 Apr 15. URL: https://www.hindustantimes.com/videos/coronavirus-crisis/mumbai-frenzy-inside-story-of-how-malicious-whatsapp-rumours-caused-mess/video-0MxBikEfs1T1DwCNfHzUwJ.html [accessed 2021-01-20]
29. Purohit K. Misinformation, fake news spark India coronavirus fears. AlJazeera - News (Coronavirus pandemic). 2020 Mar 10. URL: https://www.aljazeera.com/news/2020/03/misinformation-fake-news-spark-india-coronavirus-fears-200309051731540.html [accessed 2020-05-03]
30. False claim: Say goodbye to coronavirus with a daily regimen of vitamins C and E, sunlight, resting, drinking water, an egg, and an alkaline diet. Thomson Reuters. 2020 Apr 16. URL: https://www.reuters.com/article/uk-factcheck-goodbye-coronavirus/false-claim-say-goodbye-to-coronavirus-with-a-daily-regimen-of-vitamins-c-and-e-sunlight-resting-drinking-water-an-egg-and-an-alkaline-diet-idUSKBN21Y2IQ [accessed 2020-05-03]
31. International Standard Classification of Occupations 08 (ISCO-08). In: International Labour Office. Geneva, Switzerland: United Nations Statistics Division; 2012.
32. "Credibility". Merriam-Webster.com Dictionary. 2020 Apr 27. URL: https://www.merriam-webster.com/dictionary/credibility [accessed 2020-05-03]
33. Hoax report claims China sought Supreme Court approval to euthanise 20,000 coronavirus patients. AFP Fact Check. 2020 Feb 11. URL: https://factcheck.afp.com/hoax-report-claims-china-sought-supreme-court-approval-euthanise-20000-coronavirus-patients [accessed 2020-05-03]
34. Coronavirus: The fake health advice you should ignore. BBC News. 2020 Mar 08. URL: https://www.bbc.com/news/world-51735367 [accessed 2020-05-03]
35. Cascella MR, Rajnik M, Cuomo A, Dulebohn SC, Napoli RD. Features, evaluation, and treatment of coronavirus. StatPearls [Internet]. 2020 Oct 04. URL: https://www.ncbi.nlm.nih.gov/books/NBK554776/ [accessed 2021-01-25]
36. Dahiya H. Burning candles kills coronavirus? PM's message given false spin. The Quint. 2020 Apr 03. URL: https://www.thequint.com/news/webqoof/candle-heat-will-kill-covid-19-fake-claims-viral-after-pms-message [accessed 2020-05-03]
37. Desai M. Corona Virus Disease (Covid-19) [Report]. Plexus MD. 2020. URL: https://s3-ap-southeast-1.amazonaws.com/plexusmd/Images/HostedResources/8b3a348307c34bfd995d8831715cd366.pdf?utm_source=Whatsapp&utm_medium=SMS&utm_campaign=COVID19 [accessed 2021-01-25]
38. Isalkar U. Pune: Two private labs to start novel coronavirus tests. The Times Of India. 2020 Mar 26. URL: https://timesofindia.indiatimes.com/city/pune/pune-two-private-labs-to-start-novel-coronavirus-tests/articleshow/74821695.cms [accessed 2020-05-03]
39. Khanna M. This is fake! NASA satellite hasn't shown coronavirus retreat in India, Please stay safe. Indiatimes - Technology. 2020 Mar 22. URL: https://www.indiatimes.com/technology/science-and-future/this-is-fake-nasa-satellite-hasnt-shown-coronavirus-retreat-in-india-please-stay-safe-509003.html [accessed 2020-05-03]
40. Kurup R. COVID-19: Govt of India launches a WhatsApp chatbot. The Hindu - Businessline. 2020 Mar 25. URL: https://www.thehindubusinessline.com/info-tech/covid-19-india-launches-a-whatsapp-chatbot/article31127438.ece [accessed 2020-05-03]
41. Menon S. Coronavirus: Herbal remedies in India and other claims fact-checked. BBC Reality Check. 2020 May 19. URL: https://www.bbc.com/news/world-asia-india-51910099 [accessed 2020-05-03]
42. Studman A. Coronavirus: how to protect yourself and others, plus what protective measures don?t work. Which.co.uk. 2020 Mar 23. URL: https://www.which.co.uk/news/2020/03/coronavirus-how-you-can-protect-yourself/ [accessed 2020-05-03]

43. Sutaria S. Viral Coronavirus 'Advisory' Is Not From UNICEF2020. Boomlive.in. 2020 Mar 09. URL: https://www.boomlive.in/health/viral-coronavirus-advisory-is-not-from-unicef-7171 [accessed 2021-01-18]
44. Fake News Alert: WhatsApp message on WHO lockdown protocol for COVID-19 in India. World Health Organization. 2020. URL: https://www.who.int/india/emergencies/coronavirus-disease-(covid-19)/fake-news-alert [accessed 2021-01-20]
45. ANOVA (Analysis of Variance). StatisticsSolutions. URL: https://www.statisticssolutions.com/manova-analysis-anova/ [accessed 2020-05-03]
46. Stephanie G. Fmax / Hartley's Test: Definition, Step by Step Example, Table. StatisticsHowTo.com: Elementary Statistics for the rest of us!. 2016 May 25. URL: https://www.statisticshowto.com/fmax-hartleys-test/ [accessed 2020-05-03]
47. Field A. Discovering Statistics using IBM SPSS Statistics. In: Sage Publications Ltd, 4th edition. Newbury Park, California: Sage Publications; Jan 2013.
48. Ministry of Health and Family Welfare, Government of India. 2020 Apr 03. URL: https://www.mohfw.gov.in/pdf/Advisory&ManualonuseofHomemadeProtectiveCoverforFace&Mouth.pdf [accessed 2021-01-25]
49. Chan K. Some aspects of toxic contaminants in herbal medicines. Chemosphere 2003 Sep;52(9):1361-1371 [FREE Full text] [doi: 10.1016/s0045-6535(03)00471-5]
50. The Lancet. COVID-19: protecting health-care workers. Lancet 2020 Mar 21;395(10228):922 [FREE Full text] [doi: 10.1016/S0140-6736(20)30644-9] [Medline: 32199474]
51. Edelman. Trust Barometer Special Report: Brands and Social Media. Edelman. 2018 Jun 18.
52. Allen RL, Davis AS. Hawthorne Effect. In: Goldstein S, Naglieri JA, editors. Encyclopedia of Child Behavior and Development. Boston, MA: Springer; 2011.
53. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
54. Mehta AM, Bruns A, Newton J. Trust, but verify: social media models for disaster management. Disasters 2017 Jul;41(3):549-565. [doi: 10.1111/disa.12218] [Medline: 27652523]
55. Hern A. Britons less trusting of social media than other major nations. The Guardian Internet. 2019 May 03. URL: https://www.theguardian.com/world/2019/may/03/britons-less-trusting-of-social-media-than-other-major-nations-facebook-twitter [accessed 2020-05-03]

Abbreviations

ANOVA: analysis of variance
CRWM: coronavirus-related WhatsApp message
HCW: health care worker
SNS: social networking sites
P: parameter

Edited by G Eysenbach; submitted 05.05.20; peer-reviewed by C Jacob, A Amresh, MA Bahrami, A Louren; comments to author 08.06.20; revised version received 20.06.20; accepted 10.01.21; published 30.01.21.

Please cite as:
Bapaye JA, Bapaye HA
Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study
JMIR Public Health Surveill 2021;7(1):e19858
URL: http://publichealth.jmir.org/2021/1/e19858/
doi:10.2196/19858
PMID:33444152

Original Paper

Use of Technology to Access Health Information/Services and Subsequent Association With WASH (Water Access, Sanitation, and Hygiene) Knowledge and Behaviors Among Women With Children Under 2 Years of Age in Indonesia: Cross-sectional Study

Heidi Jane Niedfeldt[1], MPH; Emmalene Beckstead[1], MPH; Emily Chahalis[1], MPH; Mindy Jensen[1], BA; Britton Reher[1], BS; Scott Torres[2], BS; Cut Novianti Rachmi[3], MD, MIPH, PhD; Hafizah Jusril[3], MHC; Cougar Hall[1], PhD; Joshua H West[1], MPH, PhD; Benjamin T Crookston[1], MPH, PhD

[1]Department of Public Health, Brigham Young University, Provo, UT, United States
[2]RTI International, Washington, DC, United States
[3]Reconstra, Jakarta, Indonesia

Corresponding Author:
Heidi Jane Niedfeldt, MPH
Department of Public Health
Brigham Young University
2137 LSB
Provo, UT, 84602
United States
Phone: 1 801 422 3143
Email: heidiniedfeldt@gmail.com

Abstract

Background: Water access, sanitation, and hygiene (WASH) remain a public health concern in Indonesia. Proper WASH practices can decrease risk of stunting, wasting, and disease in children under the age of 2.

Objective: The purpose of our study is to examine if using technology to access health information and services among Indonesian women affects knowledge and behaviors regarding handwashing and defecation practices.

Methods: Our study is an interview-based cross-sectional survey. Participants included 1734 mothers of children under 2 years of age. These women were randomly selected and interviewed as part of a 3-stage cluster sampling technique. Our study uses data regarding WASH knowledge which includes benefits of handwashing with soap, 5 critical times of handwashing, risks of open defecation, media of disease transmission, defecation locations, and risks of open defecation. Data regarding WASH behaviors were also included: handwashing with soap, type of latrine used at home, and where defecation took place. This investigation used adjusted and unadjusted logistic and linear regression models to determine differences in WASH outcomes between those who use technology to access health information and services and those who did not.

Results: One result is that Indonesian women with children under 2 years of age who use technology to access health information and services are more likely to know the advantages of proper handwashing (odds ratio [OR] 2.603, 95% CI 1.666-4.067) and know the 5 critical times of handwashing (OR 1.217, 95% CI 0.969-1.528). Women who use technology to access health information are also more likely to know the risks of open defecation (OR 1.627, 95% CI 1.170-2.264) and use a type of toilet (such as a gooseneck or squat toilet) that limits risk (OR 3.858, 95% CI 2.628-5.665) compared to women who did not use technology to access health information.

Conclusions: Using technology to access health information and services was associated with an increase in handwashing and defecation knowledge. In the future, promoting mothers of children under 2 years of age to access health information through technology might be used to increase handwashing and defecation knowledge as well as safe defecation practices. However, further research should be done to determine how technology may increase the frequency of recommended handwashing behaviors.

(JMIR Public Health Surveill 2021;7(1):e19349) doi:10.2196/19349

KEYWORDS
technology; defecation; handwashing; WASH; stunting

Introduction

Water access, sanitation, and hygiene (WASH) remain a global public health concern. As of 2014, about 2.5 billion individuals worldwide did not use an improved sanitation facility (a sanitation facility that keeps human excreta and human contact hygienically separate) and about 1 billion of these individuals practice open defecation [1]. Additionally, about 700 million individuals did not have access to improved water (water that is protected from outside contamination such as fecal matter) [1]. In Indonesia about 30 million individuals practice open defecation [2].

WASH practices play an important role in the healthy development of children, especially those under the age of 2. In urban areas of East Jakarta, Indonesia, children living in a house with more sewage have a higher prevalence of diarrhea than those who live in houses with less sewage [3]. Diarrheal disease is the second leading cause of death among children under the age of 5 worldwide [4]. Households in 112 rural districts in India that had access to a toilet facility, compared with open defecation, had 39% reduced odds of childhood stunting in the first 24 months [5]. Stunting is of particular concern to child development as stunted children have reduced cognitive function, adult economic productivity, as well as increased mortality and morbidity [6,7]. Stunting is also a major challenge in Indonesia where approximately 37% of all children are stunted [8].

Individuals can obtain health information from a variety of traditional print, radio, or television media, which have been effective components of health communication interventions [9,10]. In recent years, interventions addressing health behaviors have used emerging technologies, starting with SMS text messages. For example, in a review of 13 studies addressing health disparities, researchers found that using SMS text messaging interventions can have positive short-term behavioral outcomes [11]. Cormick et al [12] found that 96% of women in urban Argentina would like to receive health information via SMS text messages about prenatal care. In addition to SMS text messages, the emergence of smartphones has increased opportunities for individuals to access health information more readily, which may influence dietary intake, reduce stunting rates, and address other health disparities in children. For instance, one study in the Changning District of Shanghai, China, found that 26.2% of pregnant mothers used an app to learn about nutrition and to record their diet [13]. A review of 4 studies of women in urban and rural low and lower middle-income countries (Indonesia, Kenya, and 2 in India) shows that mobile health interventions improve nutrient intake of pregnant women [10,14]. Specific to Indonesia, smartphone users are estimated to be more than 150 million, or approximately 56% of the population [15]. While smartphone ownership is highest in urban areas (71%), ownership in rural areas has grown rapidly over the past decade and was estimated at 42% in 2018. Chatting and SMS text messaging are the most commonly reported smartphone activities among Indonesians, followed by social media use, image and video searches, and gaming [16]. Household computer and laptop use for Internet access are comparatively low at just 31% in urban areas and 24% in rural areas [16].

While there is research about mobile phones, there is a lack of studies on using smartphones, tablets, and computers to access information about WASH and safe defecation practices to improve childhood health in rural Indonesia. The purpose of our study was to explore the relationship of accessing WASH-related knowledge using a smartphone, tablet, or computer with improvements in WASH knowledge and behaviors among women in Indonesia. In particular, our study sought to understand how accessing knowledge via a smartphone, tablet, or computer may impact safe defecation knowledge and practices.

Methods

Design

Our study included an analysis of cross-sectional data collected in rural Indonesia following the 2014-2018 National Nutrition Communication Campaign (NNCC) intervention and represents a collaborative effort between IMA (Interchurch Medical Assistance)–World Health, the University of Indonesia’s Center for Nutrition and Health Studies, and the Ministry of Health in Indonesia. The NNCC was designed to address the health condition of stunting that impacts numerous children in Indonesia. Of the 34 provinces in Indonesia, the NNCC utilized mass media campaigns, advocacy interventions, and interpersonal communication in 3 provinces, working in over 688 (approximately 74,000 nationwide) villages and 11 districts (approximately 7000 nationwide).

Sample

The study sample consisted of 1734 mothers of children under 2 years of age from 3 rural districts (Banyuasin, Kubu Raya, and Katingan) located in 3 provinces (South Sumatra, West Kalimantan, and Central Kalimantan) in Indonesia. One district was randomly selected from each of the 3 provinces. A multilevel sampling strategy was used to construct the study sample. Within each of the 3 rural districts, 30 villages were randomly selected, and each represented a cluster unit. At a more local level, 4 subvillages were randomly selected from within each of the 30 villages, in each of the 3 districts. Finally, in each of these subvillages a list was compiled from a local health center of mothers of children under the age of 2 and 5 mothers were selected randomly from the list from each subvillage. After using the formula for a hypothesis testing between 2 population proportions, the target sample size from each of the 3 districts was determined to be 600 mothers, 1800 overall. The final study sample included 1734 mothers from 90 villages, 3 districts, and 3 province; 1740 mothers were originally approved but 6 refused to participate.

Procedure

Ethical approval was obtained from the Ethical Research Committee by the Faculty of Public Health, Indonesia University. Reconstra, a research firm from Jakarta, conducted the data collection. Signed informed consent was sought from each participant prior to the interview and participation of all women was voluntary and no compensation was provided. Survey data were collected using an electronic tablet by experienced interviewers and field coordinators. Each interviewer interviewed approximately 6 respondents per day and reported to field coordinators who then verified the responses and uploaded survey data daily. A data manager checked data and noted any errors. Data cleaning was done prior to analysis.

Variables

Demographics

Demographic information was gathered from each participant and included mother's age, mother's highest level of education attained, and total household income (Multimedia Appendix 1). A measure of mothers' use of technology was assessed by identifying respondents who used either a tablet, computer/laptop, or smartphone to access health services.

Knowledge and Behaviors Related to Handwashing and Defecation

Knowledge and behavior of respondents related to WASH were assessed by asking respondents to identify benefits of proper handwashing using soap (options include prevent germ transmission, reduce diarrhea, and prevent infection), 5 critical handwashing times (options include after defecation, after cleaning baby who defecated, before preparing meals, before eating meals, and before breastfeeding), when they used soap while handwashing in the previous 24 hours (options include after defecation, after cleaning baby who defecated, before preparing meals, before eating meals, and before breastfeeding), risks of open defecation (options include transmission of germs and diarrhea), media for disease transmission from stool to child (options include flies, water, and dirt), proper location for defecation (options include hygienic latrine), and where the respondent defecated at home (options include gooseneck toilet, squat toilet with or without floor, and pit latrine). Each response was coded as yes or no for the knowledge or behavior indicator. For example, each critical handwashing time identified was coded as a separate variable with a yes for those who identified each particular critical time. Proportions of respondents that reported each knowledge or behavior are provided.

Index indicators were created to summarize individual indicators described above. Indices created included proportion of respondents who could identify at least one correct benefit of handwashing with soap (yes vs no); the number of critical handwashing times identified; number of times soap was used while washing hands in the previous 24 hours (0-5); proportion of respondents who could identify at least one risk of open defecation (yes vs no); proportion of respondents who could identify at least one medium of stool to child disease transmission (yes vs no); number of correct places to defecate (0-5); and proportion of respondents who reported defecating in a gooseneck toilet, squat toilet with no floor, or squat toilet with floor, or discarded feces in a septic tank or a closed ground hole (yes vs no).

Use of Technology to Access Health Services and Impact on Knowledge and Behavior Related to Hygiene and Defecation

Respondents were asked whether they accessed health services using modern tools for communication, and if so, which technology they used (options include tablet, computer/laptop, or smartphone). An index indicator of access to health services was constructed by identifying respondents who used any of the 3 technologies to access health services (yes vs no). The relationship of using technology to access health services with knowledge and behavior indices related to hygiene and defecation was then assessed.

Statistical Analysis

SAS version 9.4 (IBM) was used to calculate descriptive statistics. Regression models were used to assess the association between use of technology to access health services and each individual indicator of WASH-related knowledge and behavior. Adjusted models were also constructed to control for mother's age, mother's education, and total household income. These controls were added because of the association each variable has been shown to have on similar outcomes in previous studies [13,14,17,18]. Hence, the use of technology to access health services was examined with each WASH variable individually with regression analysis and then again with the standard controls using regression modeling. The health services technology variable was always used as the model predictor while the WASH variable was used as the outcome. Logistic regression was used to assess the association between health services technology and dichotomous WASH variables while linear regression was used when WASH variables were continuous. Logistic models included odds ratios (ORs) and 95% CI while linear regression models included point estimates and *P*-values.

Results

There were a total of 1734 mothers with children under the age of 2 (Table 1). Most mothers had a primary school education, while few had tertiary education. Being unemployed or a housewife were the most common occupations. Other occupations included small trader, civil servant, and private employee. The mean total annual household income was €131.05 (US $160.04). Almost one-fifth of respondents have access to and used a phone, computer, or tablet to access health information and services.

Table 1. Participant demographics in Banyuasin, Kubu Raya, and Katingan in 2018 (N=1734).

Demographics	Values
Mean (SD) age	28.9 (6.30)
Education, n (%)	
None	97 (5.59)
Primary school	670 (38.64)
Junior high school	423 (24.39)
Senior high school	434 (25.03)
Tertiary education	110 (6.34)
Occupation, n (%)	
Unemployed or Housewife	1461 (84.26)
Farmer	49 (2.83)
Light traders/Shop owner	79 (4.56)
Other	145 (8.36)
Religion, n (%)	
Islam	1640 (94.58)
Other	94 (5.42)
Technology use, n (%)	
Phone	265 (15.28)
Computer	14 (0.81)
Tablet	16 (0.92)
Any technology	276 (15.92)
Mean (SD) total household income (Euros)[a]	131.05 (116.17)[b]

[a]Indonesian Rupiah (official currency of Indonesia) was converted to Euros.

[b]US $160.04 (141.85).

In most cases of handwashing knowledge and behaviors, participants who used any technology to seek health information were better off than those who did not use technology (Table 2). For example, 91.7% (253/276) of participants who used technology knew at least one benefit of handwashing compared to 80.86% (1179/1458) of those who did not use technology. Further, 88.41% (244/276) of households who used technology for health reported using a hygienic location for defecation while 66.39% (968/1458) of households who did not use technology reported using a hygienic location.

Table 2. Knowledge and behaviors of handwashing and defecation by use of health technology among Indonesian women in Banyuasin, Kubu Raya, and Katingan in 2018.

Knowledge and behavior	Health technology use		*P*-value[a]
	Yes, n (%) N=276	No, n (%) N=1458	
Handwashing			
Know at least one benefit of proper handwashing	253 (91.67)	1179 (80.86)	<.001
Know that handwashing can			
Prevent germ transmission	239 (86.59)	1076 (73.80)	<.001
Decrease diarrhea	49 (17.75)	188 (12.89)	.03
Prevent infection	19 (6.88)	98 (6.72)	.92
Know handwashing should occur			
After defecation	206 (74.64)	994 (68.18)	.03
After cleaning baby/infant who defecated	108 (39.13)	638 (43.76)	.15
Before preparing meals	147 (53.26)	758 (51.99)	.70
Before eating meals	249 (90.22)	1199 (82.24)	.001
Before breastfeeding/feeding child	135 (48.91)	594 (40.74)	.012
Mean (SD) number of critical handwashing times participant identified (0-5)	3.1 (1.20)	2.87 (1.40)	.02[b]
Mean (SD) number of times soap was used for handwashing since yesterday until today (0-5)	2.6 (1.30)	2.5 (1.30)	.25[b]
Defecation			
Know the risks of open defecation	227 (82.25)	1079 (74.01)	.004
Know about transmission of germs/*Escherichia coli* bacteria	176 (63.77)	775 (53.16)	<.001
Know about causes of diarrhea	91 (32.97)	399 (27.37)	.06
Know mode of disease transmission from			
Stool	191 (69.20)	791 (54.25)	<.001
Flies	149 (53.99)	562 (38.55)	<.001
Water	59 (21.38)	258 (17.70)	.15
Dirt	22 (7.97)	77 (5.28)	.08
Know hygienic location for defecation	273 (98.91)	1377 (94.44)	.002
Household uses gooseneck toilet or squat toilet with or without floor to defecate or septic tank or closed ground to discard feces	244 (88.41)	968 (66.39)	<.001

[a]Used chi-square test unless otherwise noted.

[b]Used *t* test.

Participants who used a health technology to access health services were more likely to know the benefits of proper handwashing as opposed to those who did not use health technology to access health services (OR 2.603, 95% CI 1.666-4.067; Table 3). After controlling for maternal age, maternal education level, and total household income, the use of technology to access health information and services was associated with knowledge of proper handwashing benefits (*P*=.004). Those who used technology to access health services were more likely to understand the media of disease transmission from stool to child.

Table 3. Use of technology to access health services and its impact on knowledge of hygiene and defecation in Banyuasin, Kubu Raya, and Katingan in 2018.[a]

Knowledge and behaviors	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Handwashing		
Know at least one benefit of proper handwashing	2.60 (1.67-4.07)[b]	2.07 (1.26-3.41)[b]
Know that handwashing can		
Prevent germ transmission	2.29 (1.59-3.31)[b]	1.93 (1.28-2.91)[b]
Decrease diarrhea	1.46 (1.03-2.06)[b]	1.17 (0.78-1.75)
Prevent infection	1.03 (0.62-1.71)	0.69 (0.38-1.25)
Know handwashing should occur		
After defecation	1.37 (1.03-1.84)[b]	1.04 (0.75-1.44)
After cleaning baby/infant who defecated	0.83 (0.64-1.08)	0.65 (0.05-0.88)[b]
Before preparing meals	1.05 (0.81-1.36)	0.93 (0.70-1.25)
Before eating meals	1.99 (1.31-3.03)[b]	1.66 (1.04-2.66)[b]
Before breastfeeding/feeding child	1.39 (1.08-1.80)[b]	1.19 (0.89-1.59)
Number of critical handwashing times participant identified (0-5)	0.19 (.0361)[c]	–0.01 (.8850)[d]
Defecation		
Know the risks of open defecation	1.63 (1.17-2.26)[b]	1.21 (0.83-1.75)
Know about transmission of germs/*Escherichia coli* bacteria	1.60 (1.23-2.09)[b]	1.51 (1.11-2.03)[b]
Know about causes of diarrhea	1.31 (0.99-1.72)	0.96 (0.70-1.33)
Know mode of disease transmission from		
Stool	1.89 (1.44-2.50)[b]	1.57 (1.15-2.14)[b]
Flies	1.87 (1.44-2.42)[b]	1.52 (1.13-2.03)
Water	1.27 (0.92-1.74)	1.08 (0.75-1.55)
Dirt	1.55 (0.95-2.54)	1.38 (0.77-2.46)
Know hygienic location for defecation	5.35 (1.68-17.07)	2.22 (0.66-7.46)

[a]All adjusted models include maternal age, maternal education level, and total household income. Point estimates are derived from linear regression models while all odds ratios (ORs) are derived from logistic regression models.

[b]*P*<.05.

[c]Unadjusted point estimate (*P*-value).

[d]Adjusted point estimate (*P*-value).

Mothers with children under the age of 2 and who use technology to access health information and services have a greater chance of performing appropriate defecation behaviors (OR 3.85, 95% CI 2.62-5.66; Table 4). After adjusting for maternal age, maternal education level, and total household income, the use of technology to access health information and services was positively associated with using a gooseneck toilet or squat toilet with or without floor to defecate or a septic tank or closed ground to discard feces. The association of the use of technology to access health information and services with more hygienic handwashing behaviors was not statistically significant (*P*=.77).

Table 4. Use of technology to access health information and services and impact on behavior of hygiene and defecation in Banyuasin, Kubu Raya, and Katingan in 2018.[a]

Knowledge and behaviors	Value
Handwashing	
Number of times soap was used for handwashing since yesterday until today (0-5)	
Unadjusted point estimate (*P*-value)	0.10 (.249)
Adjusted point estimate (*P*-value)	–0.03 (.771)
Defecation	
Household uses gooseneck toilet or squat toilet with or without floor to defecate or septic tank or closed ground to discard feces	
Unadjusted OR (95% CI)	3.86 (2.63-5.67)[b]
Adjusted OR (95% CI)	2.32 (1.50-3.60)[b]

[a]All adjusted models include maternal age, maternal education level, and total household income. Point estimates are derived from linear regression models while all odds ratios (ORs) are derived from logistic regression models.

[b]*P*<.001.

Discussion

Principal Findings

The purpose of our study was to determine if using technology to access health information was associated with increased WASH knowledge and optimal behaviors regarding proper handwashing and defecation practices. After controlling for age, education, and income, the study findings show that mothers of children under the age of 2 who used technology to access health information and services were more likely to be aware of the benefits of proper handwashing and proper defecation practices. While the difference in sufficient handwashing behaviors between those who used technology to access health information and those who did not was nearly non-existent, the most significant finding of our study was that these mothers have a much higher likelihood of using appropriate defecation behaviors. This factor alone has the ability to reduce illness and could provide continual positive benefits for children and families. A campaign in India to decrease open defecation by promoting community latrine use concluded that communities that used latrines experienced reduced fecal contamination in the community and improved child arm circumference, weight, and height. Households also saved time [19]. Children from villages in India with community latrine coverage had significantly higher cognitive scores 10 years later [20] and children, especially girls, were less likely to drop out of school [21]. Another study in India showed that 30%-55% of the average differences in stunting between districts could be due to differences in open defecation [22].

Our study found that different media sources were not only associated with increased WASH knowledge but also associated with WASH behaviors. While this is the first finding of this type in Indonesia, the positive relationship between technology usage and WASH knowledge has been highlighted in previous research conducted in other countries. Previous research in rural Tanzania evaluated how media access impacts WASH knowledge and behaviors [10]. Media access in the Tanzanian study included listening to the radio, watching television, or having WhatsApp on a smartphone. Exposure to media was measured based on when the media was accessed. Participants could select from 5 options: today, yesterday, in the last week, in the last month, or more than a month ago. Like our study, results from the Tanzanian study showed a similar positive trend regarding technology access and increased handwashing. Specific findings from the Tanzanian study showed that participants who watched television had a positive correlation with increased WASH knowledge [10]. One potential reason for similar findings is that IMA–World Health sponsored the WASH media campaign in both countries. While the media campaign was adapted for cultural differences, the campaigns were likely similar and resulted in similar outcomes in both countries.

Mothers who used technology had a higher likelihood of knowing when it was appropriate and necessary to use proper handwashing but did not necessarily follow through with the appropriate behavior. This is valuable information for health agencies and service providers as it highlights where the implementation gaps are, and that increasing the use of technology may be a way to promote this information, at least for some topics. For example, many participants were able to identify at least one risk of open defecation, but less had specific knowledge about the mode of disease transmission. It is also surprising that technology use was associated with mothers' knowledge of handwashing before eating a meal, but not with knowledge of handwashing before preparing a meal. It may be an indicator of a need for more emphasis on handwashing before meal preparation, whereas handwashing before eating a meal has been a consistent message. Another study that resulted in increased handwashing knowledge with no or minimal change in handwashing behavior was an interactive campaign in India. The interventions focused on using toilets and washing hands with soap. Those who participated in the campaign increased their knowledge about the benefits of handwashing by about half a standard deviation, but the change in intention to wash their hands was small [23]. Conversely, another study found that in India television advertising and SMS text messages using mobile phones increased the likelihood of mothers washing their hands [24]. While media has the potential to improve handwashing behaviors, barriers must be addressed. Three new television campaigns to increase handwashing among Australian Aboriginal communities were widely viewed and understood. However, 75% of participants indicated they would purchase more handwashing supplies if they were less expensive [25].

Our study results showed that more accurate knowledge regarding the probability of disease arising due to open defecation and proper defecation procedures could have an effect on behavior practices. Further research could be done as to how to allow more women to access health information through technology to improve the overall health and well-being of their families. This research might include exploring the technology needs and capacities of mothers. The percentage of mothers in this study that used technology to access health information was low. It is not clear if this is a function of cultural norms, income, or some other influence. Nevertheless, the promising association between mothers' use of technology and knowledge is such that a study of this nature is warranted.

A study conducted in 7 of the 8 provinces of Kenya also found a positive correlation between accessing technology and handwashing behavior [26]. Participants were chosen from 7 of the 8 provinces in Kenya. Sources of media were divided into 2 categories: media ownership and media exposure occurring in the last month. Possible media sources in this study included, but were not limited to, television, radio, newspapers, and movies. Exposure to media was gauged by determining how many household items the participants owned or the amount of various media sources they were exposed to. Results indicated that each variable directly corresponded to increased handwashing practices. Additionally, both variables had a positive association with handwashing behavior that involved soap [26]. Data came from a cross-sectional survey conducted prior to a media and community-based handwashing program organized by the World Bank and the Kenyan Ministry of Health. The handwashing results from the Kenya study correspond with results from our study. Although defecation practices were included as sociodemographic characteristics, the paper did not address whether access to technology was associated with defecation behaviors. Further research is needed to determine whether other countries experience any relationship between technology and handwashing but not defecation behaviors.

Study Limitations

A few limitations of our study should be considered when reviewing the results. First, our study did not utilize an asset index to measure poverty; rather, it used a total household income indicator. Second, the broader study, from which our data were derived, did not intend to examine the indicator for access to technology and how it relates to handwashing and defecation behaviors. These indicators were not a key focus of the broader study; however, they remain valuable to our research because of the relationship identified between access to technology and the practices of handwashing and defecation. However, the association between the use of technology to access health information and the increased WASH knowledge in mothers of children aged under 2 are imperative discoveries.

Conclusions

In conclusion, the use of technology to access health information was associated with correct WASH knowledge, and with the use of safe methods of eliminating feces. However, using technology was not associated with an increase in the number of times of handwashing with soap. The findings of our study suggest several potential opportunities for furthering knowledge and creating behavior change as these relate to handwashing and defecation practices, thereby improving health.

Acknowledgments

Our study was made possible by IMA–World Health, and supported by funding through MCA-Indonesia. We are grateful for the contribution of all the interviewers and study participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey Instrument.
[PDF File (Adobe PDF File), 7699 KB - publichealth_v7i1e19349_app1.pdf]

References

1. World Health Organization. Progress on Sanitation and Drinking Water. Geneva, Switzerland: World Health Organization; 2013.
2. World Health Organization, United Nations Children's Fund (UNICEF). Safely Managed Drinking Water: Thematic Report on Drinking Water. Geneva, Switzerland: World Health Organization; 2017.
3. Agustina R, Sari TP, Satroamidjojo S, Bovee-Oudenhoven IMJ, Feskens EJM, Kok FJ. Association of food-hygiene practices and diarrhea prevalence among Indonesian young children from low socioeconomic urban areas. BMC Public Health 2013 Oct 19;13:977 [FREE Full text] [doi: 10.1186/1471-2458-13-977] [Medline: 24138899]

4. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 2015 Jan 31;385(9966):430-440. [doi: 10.1016/S0140-6736(14)61698-6] [Medline: 25280870]
5. Rah JH, Cronin AA, Badgaiyan B, Aguayo VM, Coates S, Ahmed S. Household sanitation and personal hygiene practices are associated with child stunting in rural India: a cross-sectional analysis of surveys. BMJ Open 2015 Feb 12;5(2):e005180 [FREE Full text] [doi: 10.1136/bmjopen-2014-005180] [Medline: 25678539]
6. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. Paediatr Int Child Health 2014 Nov;34(4):250-265 [FREE Full text] [doi: 10.1179/2046905514Y.0000000158] [Medline: 25310000]
7. Dewey K, Begum K. Long-term consequences of stunting in early life. Matern Child Nutr 2011 Oct;7 Suppl 3:5-18 [FREE Full text] [doi: 10.1111/j.1740-8709.2011.00349.x] [Medline: 21929633]
8. Depkes RI. Badan penelitian dan pengembangan Kesehatan. Riset Kesehatan Dasar. 2011. URL: http://perpustakaan.kemkes.go.id/inlislite3_kemkes/uploaded_files/temporary/DigitalCollection/MmJjNTc3Mjk4ZGI2MzNlNmJiYWE2MjM1YWUwN2YxZTU0YjFiNGVlNg==.pdf [accessed 2020-12-21]
9. Belew AK, Ali BM, Abebe Z, Dachew BA. Dietary diversity and meal frequency among infant and young children: a community based study. Ital J Pediatr 2017 Aug 15;43(1):73 [FREE Full text] [doi: 10.1186/s13052-017-0384-6] [Medline: 28810887]
10. Alexander CC, Shrestha S, Tounkara MD, Cooper S, Hunt L, Hoj TH, et al. Media Access is Associated with Knowledge of Optimal Water, Sanitation and Hygiene Practices in Tanzania. Int J Environ Res Public Health 2019 Jun 03;16(11) [FREE Full text] [doi: 10.3390/ijerph16111963] [Medline: 31163573]
11. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
12. Cormick G, Kim NA, Rodgers A, Gibbons L, Buekens PM, Belizán JM, et al. Interest of pregnant women in the use of SMS (short message service) text messages for the improvement of perinatal and postnatal care. Reprod Health 2012;9:9 [FREE Full text] [doi: 10.1186/1742-4755-9-9] [Medline: 22866753]
13. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
14. Saronga NJ, Burrows T, Collins CE, Ashman AM, Rollo ME. mHealth interventions targeting pregnancy intakes in low and lower-middle income countries: Systematic review. Matern Child Nutr 2019 Apr 06;15(2):e12777 [FREE Full text] [doi: 10.1111/mcn.12777] [Medline: 30609297]
15. Handayani PW, Meigasari DA, Pinem AA, Hidayanto AN, Ayuningtyas D. Critical success factors for mobile health implementation in Indonesia. Heliyon 2018 Nov;4(11):e00981 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e00981] [Medline: 30519665]
16. Nabila M. APJII: Penetrasi Pengguna Internet Indonesia Capai 143 Juta Orang. Dailysocial. 2018. URL: https://dailysocial.id/post/apjii-survei-internet-indonesia-2017 [accessed 2020-12-21]
17. White S, Thorseth AH, Dreibelbis R, Curtis V. The determinants of handwashing behaviour in domestic settings: An integrative systematic review. Int J Hyg Environ Health 2020 Jun;227:113512 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113512] [Medline: 32220763]
18. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
19. Dickinson KL, Patil SR, Pattanayak SK, Poulos C, Yang J. Nature's Call: Impacts of Sanitation Choices in Orissa, India. Economic Development and Cultural Change 2015 Oct;64(1):1-29. [doi: 10.1086/682958]
20. Orgill-Meyer J, Pattanayak SK. Improved sanitation increases long-term cognitive test scores. World Development 2020 Aug;132:104975. [doi: 10.1016/j.worlddev.2020.104975]
21. Orgill-Meyer J. Interaction of village and school latrines on educational outcomes in India Internet. Journal of Water, Sanitation and Hygiene for Development 2020 Jun 1;10(4):618-627. [doi: 10.2166/washdev.2020.049]
22. Spears D, Ghosh A, Cumming O. Open defecation and childhood stunting in India: an ecological analysis of new data from 112 districts. PLoS One 2013 Sep 16;8(9):e73784 [FREE Full text] [doi: 10.1371/journal.pone.0073784] [Medline: 24066070]
23. Seimetz E, Kumar S, Mosler H. Effects of an awareness raising campaign on intention and behavioural determinants for handwashing. Health Educ Res 2016 Apr 02;31(2):109-120. [doi: 10.1093/her/cyw002] [Medline: 26936481]
24. Tidwell JB, Gopalakrishnan A, Lovelady S, Sheth E, Unni A, Wright R, et al. Effect of Two Complementary Mass-Scale Media Interventions on Handwashing with Soap among Mothers. J Health Commun 2019 Mar 26;24(2):203-215. [doi: 10.1080/10810730.2019.1593554] [Medline: 30912707]
25. McDonald E, Cunningham T, Slavin N. Evaluating a handwashing with soap program in Australian remote Aboriginal communities: a pre and post intervention study design. BMC Public Health 2015 Nov 27;15(1):1188 [FREE Full text] [doi: 10.1186/s12889-015-2503-x] [Medline: 26614522]
26. Schmidt W, Aunger R, Coombes Y, Maina P, Matiko C, Biran A, et al. Determinants of handwashing practices in Kenya: the role of media exposure, poverty and infrastructure. Trop Med Int Health 2009 Dec;14(12):1534-1541 [FREE Full text] [doi: 10.1111/j.1365-3156.2009.02404.x] [Medline: 19793069]

Abbreviations

IMA: Interchurch Medical Assistance
NNCC: National Nutrition Communication Campaign
WASH: Water access, sanitation, and hygiene

Edited by T Sanchez; submitted 14.04.20; peer-reviewed by J Hartvigsen, K Fuji; comments to author 29.06.20; revised version received 17.08.20; accepted 01.12.20; published 14.01.21.

Please cite as:
Niedfeldt HJ, Beckstead E, Chahalis E, Jensen M, Reher B, Torres S, Rachmi CN, Jusril H, Hall C, West JH, Crookston BT
Use of Technology to Access Health Information/Services and Subsequent Association With WASH (Water Access, Sanitation, and Hygiene) Knowledge and Behaviors Among Women With Children Under 2 Years of Age in Indonesia: Cross-sectional Study
JMIR Public Health Surveill 2021;7(1):e19349
URL: http://publichealth.jmir.org/2021/1/e19349/
doi:10.2196/19349
PMID:33443485

Original Paper

Drivers of Acceptance of COVID-19 Proximity Tracing Apps in Switzerland: Panel Survey Analysis

Viktor von Wyl[1,2], PhD; Marc Höglinger[3], PhD; Chloé Sieber[1], MSc; Marco Kaufmann[1], PhD; André Moser[1], PhD; Miquel Serra-Burriel[1], PhD; Tala Ballouz[1], MD; Dominik Menges[1], MD; Anja Frei[1], PhD; Milo Alan Puhan[1], MD, PhD

[1]Epidemiology, Biostatistics & Prevention Institute, University of Zurich, Zürich, Switzerland
[2]Institute for Implementation Science in Health Care, University of Zurich, Zürich, Switzerland
[3]Winterthur Institute of Health Economics, Zurich University of Applied Sciences, Winterthur, Switzerland

Corresponding Author:
Viktor von Wyl, PhD
Epdemiology, Biostatistics & Prevention Institute
University of Zurich
Hirschengraben 84
Zürich, 8001
Switzerland
Phone: 41 446346380
Email: viktor.vonwyl@uzh.ch

Abstract

Background: Digital proximity tracing apps have been released to mitigate the transmission of SARS-CoV-2, the virus known to cause COVID-19. However, it remains unclear how the acceptance and uptake of these apps can be improved.

Objective: This study aimed to investigate the coverage of the SwissCovid app and the reasons for its nonuse in Switzerland during a period of increasing incidence of COVID-19 cases.

Methods: We collected data between September 28 and October 8, 2020, via a nationwide online panel survey (COVID-19 Social Monitor, N=1511). We examined sociodemographic and behavioral factors associated with app use by using multivariable logistic regression, whereas reasons for app nonuse were analyzed descriptively.

Results: Overall, 46.5% (703/1511) of the survey participants reported they used the SwissCovid app, which was an increase from 43.9% (662/1508) reported in the previous study wave conducted in July 2020. A higher monthly household income (ie, income >CHF 10,000 or >US $11,000 vs income ≤CHF 6000 or <US $6600 [reference]: odds ratio [OR] 1.92, 95% CI 1.40-2.64), more frequent internet use (ie, daily [reference] vs less than weekly: OR 0.37, 95% CI 0.16-0.85), better adherence to recommendations for wearing masks (ie, always or most of the time [reference] vs rarely or never: OR 0.28, 95% CI 0.15-0.52), and nonsmoker status (OR 1.32, 95% CI 1.01-1.71) were associated with an increased likelihood for app uptake. Citizenship status (ie, non-Swiss citizenship vs. Swiss [reference]: OR 0.61, 95% CI 0.43-0.87), and language region (French vs Swiss German [reference]: OR 0.61, 95% CI 0.46-0.80) were associated with a lower likelihood for app uptake. Further analysis in a randomly selected subsample (n=712) with more detailed information showed that higher levels of trust in government and health authorities were also associated with a higher likelihood for app uptake (ie, high vs low [reference] trust: OR 3.13, 95% CI 1.58-6.22). The most frequent reasons for app nonuse were lack of perceived benefit of using the app (297/808, 36.8%), followed by the lack of a compatible phone (184/808, 22.8%), and privacy concerns (181/808, 22.4%).

Conclusions: Eliminating technical hurdles and communicating the benefits of digital proximity tracing apps are crucial to promote further uptake and adherence of such apps and, ultimately, enhance their effectiveness to aid pandemic mitigation strategies.

(JMIR Public Health Surveill 2021;7(1):e25701) doi:10.2196/25701

KEYWORDS

COVID-19; SARS-CoV-2; digital proximity tracing; digital contact tracing; mHealth; tracing; compliance; acceptance; uptake; usability; communication

Introduction

Background

Safe and effective vaccines against SARS-CoV-2, the causative agent of COVID-19, are not largely available in most countries. Therefore, global and national health authorities continue to rely on nonpharmaceutical interventions in their fight against the ongoing COVID-19 pandemic. Cornerstones of pandemic mitigation measures include testing, tracing, isolation, and quarantine [1]. Digital proximity tracing (DPT) apps are expected to further enhance conventional mitigation measures, and classic, interview-based contact tracing in particular. These apps constitute a novel, still largely untested health technology that anonymously records the user's proximity contacts, that is, other app users who were within a prespecified radius for a certain amount of time [2]. In case the app user tests positive for COVID-19, they can notify their proximity contacts in an anonymous manner through such DPT apps.

The rationales for using DPT apps as pandemic mitigation tools are based on a modelling study, which suggests that DPT alone has the ability to stop the spread of the COVID-19 pandemic [3,4]. Classic contact tracing is labor- and time-consuming, and exposed contacts can sometimes only be reached and notified with substantial time lags [5]. By comparison, DPT can lead to faster notification and earlier self-quarantine of exposed contacts [3,6]. In addition, DPT has a wider reach than classic contact tracing, as it also includes exposed contacts that the infected person may not know by name (eg, chance encounters in a public space). However, the modelling studies further suggest that these expected effects of digital contact tracing depend on several assumptions. Specifically, a considerable proportion of the population must use the DPT app (eg, 60% and more if no other mitigation measures are implemented), the turnaround time of test results and digital notification of exposed contacts must be within 1-2 days and notified contacts should enter self-quarantine immediately [3,7].

How the Swiss DPT App (SwissCovid) Works

The Swiss DPT app, officially named "SwissCovid," follows the blueprint of decentralized, privacy-preserving proximity tracing (DP-3T). Detailed explanations of the DP-3T design can be found elsewhere [2,6]. The DP-3T app architecture has also become the basis for national DPT apps in countries such as Italy or Germany, and it has gained the support of Apple and Google, who provide application programming interfaces to support the app's functionality [8].

The SwissCovid app was publicly released on June 25, 2020 [9]. Similar to other DP-3T–inspired apps, smartphones with the SwissCovid app installed will send and receive Bluetooth Low Energy signals to and from other smartphones that also have the same app installed. Ephemeral, nonidentifiable keys are exchanged and stored locally on smartphones. As Bluetooth signals weaken with increasing distance between devices, signal attenuation can be employed to determine whether another phone or device was in close proximity (eg, <1.5 meters) and, if so, for how long. If any app user tests positive for COVID-19, this person will be issued an activation code (CovidCode) that should be entered into the SwissCovid app. By doing so, the user releases their ephemeral keys, which are then uploaded onto a central server system.

Smartphones with DP-3T–based apps regularly connect to this central server. The uploaded keys of users with known infection are downloaded by all smartphones using the SwissCovid app, and the smartphone user's locally stored encounter-history (ie, the list of exchanged keys) will be searched for matches with keys of infected users. If matches fulfilling the criteria for a close proximity encounter (<1.5 meters over at least 15 minutes) are found, the smartphone owner will be notified and advised to call an Infoline for further assistance. Users who receive such a notification are also advised to enter self-quarantine and undergo testing for COVID-19.

Thus, the effect of proximity tracing on pandemic containment is mediated by users being notified about possible exposure risks as soon as possible and entering quarantine to break further transmission chains (ie, by being "one step ahead"). However, emerging data from Switzerland indicate that procedural aspects (eg, speed of laboratory test results and delivery of CovidCodes) and user behavior (eg, the period before contacting the infoline after receiving an app notification of contact with another app user who has tested positive) have an influence on the performance of the DPT app notification cascade [10,11]. For example, frequent delays in issuing activation codes for app users who have tested positive for COVID-19 also delayed notification of exposed contacts (monitored in [12]).

Study Aims

DPT technologies have been developed and implemented with very limited real-life testing [1,13]. It currently remains unclear whether, and to what extent, assumptions stated by the modelling analyses are achievable under real-world conditions and whether these technologies can ultimately have a significant impact on the effectiveness of pandemic mitigation strategies [13,14].

Therefore, in this study, we aimed to investigate and synthesize to what extent some of the conditions for DPT functioning (namely, broad app uptake) were fulfilled during the first 3 months after the release of the SwissCovid app in Switzerland. Our analyses addressed 3 main research questions: (1) Which sociodemographic and health-related factors are associated with use of the SwissCovid app? (2) What are the most prominent concerns for nonuse of the SwissCovid app? (3) What is known about the adherence of app users with the recommended procedures in case of an app notification indicating proximity contact with another app user who has tested positive for COVID-19? To answer these questions, we analyzed data collected from a web-based nationwide, survey panel, complemented by publicly available data.

Methods

Data Source

This study was based on survey data collected from the Swiss COVID-19 Social Monitor project [15], a cohort study of participants randomly selected from an existing online panel population. A weighted sample from the panel, stratified based on age, gender, and language region, was used in order to make the sample representative of the Swiss population. Participants of this cohort receive an invitation every 2-6 weeks to complete a survey on various COVID-19 related topics. The survey was started on March 30, 2020; thus far, 10 study waves have been conducted, each with an average response from 1500-1700 persons from across Switzerland. All datasets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

The Swiss COVID-19 Social Monitor project collects information on sociodemographic features, comorbidities, and implementation of preventive measures related to COVID-19. In addition, 3 standardized questions were introduced to gather information about the use of the Swiss DPT app (SwissCovid; see Table S1 in Multimedia Appendix 1). The questions were jointly developed by study investigators, epidemiologists, and infectious disease experts. The standardized SwissCovid app-related questions were first introduced in Wave 8 and subsequently used in Waves 9 and 10.

The primary data source for these analyses was Wave 10 of the Swiss COVID-19 Social Monitor project (September 28 to October 8, 2020), which yielded responses from 1511 participants. Additional data on media use and trust in government, health authorities, or science, were collected for a randomly selected subsample in Wave 10 (n=712; Table S2 in Multimedia Appendix 1). Furthermore, data from 1299 participants were collected from Wave 8 (July 13-20, 2020) as well as Wave 10, which were used for analyzing within-person-changes in SwissCovid app use over time and reasons for app nonuse over time. Data from Waves 8, 9 (August 17-25, 2020), and 10 were used to evaluate user responses to app notifications.

Context of the Pandemic Situation

The observation period for this study started from app release on June 25, 2020, to approximately 3 months thereafter. By early October 2020, the app was downloaded 2.4 million times, and the number of active users was relatively stable at 1.6 million [12]. Active users were counted as the daily number of app dummy requests sent to the proximity tracing system, which tends to underestimate the real number of users [16]. Compared with Switzerland's population size of 8.6 million persons of all age groups (6.6 million in the age group between 18 and 79 years), the number of active app users corresponds to a population coverage of approximately 19% (24.2% among those aged 18-79 years).

In hindsight, the time period of this survey marked the starting point for large increases in the incidence of COVID-19 cases in Switzerland. A total of 8114 new COVID-19 cases (based on positive polymerase chain reaction tests) were reported during the study period, that is, from September 28, 2020, to October 8, 2020. By contrast, the number of new COVID-19 cases was considerably lower in the preceding 11-day period (ie, 3644 cases during September 17-27, 2020) [17].

Ethics Statement

For the COVID-19 Social Monitor project, the Ethics Committee of the Canton of Zurich confirmed that it does not fall under the Swiss Human Research Law (BASEC-Nr. Req-2020-00323). Therefore, informed consent from participants was not needed.

Measures

To study the uptake of the SwissCovid app, users and nonusers were compared by age (in 10-year categories), gender, partnership status, having children, citizenship, language region, education status, employment status, household income, smoking status, presence of self-reported comorbidities (eg, respiratory diseases, cardiovascular diseases, stroke, hypertension, diabetes, and cancer), application of preventive measures (eg, wearing masks and staying at home except for essential tasks), frequency of internet use, trust in government and health authorities, and trust in science. Individuals who reported they used the app permanently or who turned it off only occasionally were considered "app users" for the purpose of this study. Other individuals who reported not using the app (either with or without an intention to do so later) were considered "app nonusers."

Statistical Analyses

Factors Associated With App Uptake

Descriptive analyses were performed by summarizing continuous data as medians (interquartile ranges) and categorical data as percentages. Changes in app use status between Waves 8 and 10 were also analyzed descriptively among participants who contributed to both waves.

To investigate factors associated with app use, multivariable logistic regression models were constructed using the abovementioned measures as variables of interest. Age, gender, and comorbidity status were included as a priori fixed co-variables in all models; the remaining variables, including an a priori defined interaction term for age and gender, were added incrementally and retained if the Akaike Information Coefficient decreased by 2 points or more upon variable addition [18,19]. Further logistic regression analyses on the association between app use and media use and trust in government or science were performed for the subset of participants for whom this information was available. Results from regression analyses are reported as odds ratio (OR) with 95% confidence intervals.

Investigation of Reasons for App Nonuse

Reasons for nonuse of the SwissCovid app were further explored descriptively represented as n (%) based on the answer options provided, as well as an open answer field for describing other reasons. The analysis was limited to one primary reason for each participant. Sociodemographic and other characteristics as listed above were compared descriptively across the 3 most frequent reasons for app nonuse, as well as a fourth group subsuming all other reasons. All analyses were performed using Stata version 13 (Stata Corp).

Results

Sample Characteristics

The Wave 10 survey yielded 1511 responses; participant characteristics are shown in Table 1. The median age of survey participants was 48 years, and 48.8% (738/1511) were female. Almost two-thirds (975/1511, 64.5%) of the participants lived in the German language region, 22.1% (334/1511) lived in the French language region, and 13.4% (202/1511) lived in the

Italian language regions. Furthermore, 46.5% (703/1511) of the participants reported to have the app installed, of which 7.7% (116/1511) occasionally switched it off. By comparison, app installation coverage was 43.9% (662/1508) in Wave 8 (data not shown). Among the 1299 respondents participating in both Waves 8 and 10, only 75 of 733 (10.2%) app nonusers from Wave 8 had the app installed by Wave 10 (data not shown). However, 5.3% (30/566) of the app users from Wave 8 had uninstalled the app by Wave 10.

Table 1. Study populations of users and nonusers of the SwissCovid app, deployed in Switzerland as a mitigation measure for the COVID-19 pandemic.

Characteristic	Value		
	Social Monitor Project (N=1511)	App nonuse (n=808)	App use (n=703)
Age, median (IQR)	48 (34, 59)	49 (35, 58)	46 (34, 59)
Gender, female, n (%)	738 (48.8)	389 (48.1)	349 (49.6)
Partnership status, n (%)			
No partner	440 (29.1)	246 (30.4)	194 (27.6)
Living with partner	951 (62.9)	490 (60.6)	461 (65.6)
Not living with partner	120 (7.9)	72 (8.9)	48 (6.8)
Has children, yes, n (%)	163 (10.8)	92 (11.4)	71 (10.1)
Citizenship status, n (%)			
Swiss	1220 (80.7)	624 (77.2)	596 (84.8)
Swiss and other	129 (8.5)	83 (10.3)	46 (6.5)
Non-Swiss	162 (10.7)	101 (12.5)	61 (8.7)
Language region, n (%)			
German	975 (64.5)	494 (61.1)	481 (68.4)
French	334 (22.1)	200 (24.8)	134 (19.1)
Ticino	202 (13.4)	114 (14.1)	88 (12.5)
Education, n (%)			
Only mandatory schooling	93 (6.2)	60 (7.4)	33 (4.7)
Completed professional education	728 (48.2)	406 (50.2)	322 (45.8)
University or university of applied sciences	690 (45.7)	342 (42.3)	348 (49.5)
Currently employed, n (%)	1066 (70.5)	563 (69.7)	503 (71.6)
Monthly household income, n (%)			
≤CHF 6000 (US $6600)	397 (26.3)	246 (30.4)	151 (21.5)
CHF 6000-10,000 (US $6600-11,000)	491 (32.5)	261 (32.3)	230 (32.7)
>CHF 10,000 (US $11,000)	343 (22.7)	146 (18.1)	197 (28)
No answer	280 (18.5)	155 (19.2)	125 (17.8)
Smoker, yes, n (%)	313 (20.7)	188 (23.3)	125 (17.8)
Self-reported chronic illness[a], n (%)	378 (25)	197 (24.4)	181 (25.7)
Use of protective masks, n (%)			
Always or most of the time	962 (63.7)	494 (61.1)	468 (66.6)
Sometimes	484 (32)	264 (32.7)	220 (31.3)
Rarely or never	65 (4.3)	50 (6.2)	15 (2.1)
Staying at home except for essential tasks, n (%)			
Always or most of the time	409 (27.1)	224 (27.7)	185 (26.3)
Sometimes	622 (41.2)	316 (39.1)	306 (43.5)
Rarely or never	480 (31.8)	268 (33.2)	212 (30.2)
Frequency of internet use, n (%)			
Once daily or several times a day	1329 (88)	685 (84.8)	644 (91.6)
Once weekly or several days per week	150 (9.9)	99 (12.3)	51 (7.3)
Never or less than once weekly	32 (2.1)	24 (3)	8 (1.1)

Characteristic	Value		
	Social Monitor Project (N=1511)	App nonuse (n=808)	App use (n=703)
Trust in government[b], n (%)			
Little	60/712 (8.4)	47/375 (12.5)	13/337 (3.9)
Somewhat	163/712 (22.9)	102/375 (27.2)	61/337 (18.1)
Large	489/712 (68.7)	226/375 (60.3)	263/337 (78)
Trust in science[b], n (%)			
Little	58/710 (8.2)	42/374 (11.2)	16/336 (4.8)
Somewhat	207/710 (29.2)	129/374 (34.5)	78/336 (23.2)
Large	445/710 (62.7)	203/374 (54.3)	242/336 (72)
SwissCovid app use, n (%)			
App user	587 (38.8)	N/A[c]	N/A
App user, occasionally switching off the app	116 (7.7)	N/A	N/A
Intends to use the app	53 (3.5)	N/A	N/A
Has uninstalled app	66 (4.4)	N/A	N/A
Not using the app	689 (45.6)	N/A	N/A

[a]Presence of chronic illness was defined based on self-reporting of at least one of the following conditions: asthma, chronic obstructive pulmonary disease, diabetes, hypertension, cardiovascular disease, stroke, and cancer.

[b]Data only available in a randomly selected split-sample, including 47.1% (712/1511) of the full study population.

[c]N/A: not applicable.

Factors Associated With App Uptake

Multivariable logistic regression analyses revealed that several factors were associated with app uptake (see Table 2). Analysis of the full study sample showed that citizenship status (Swiss and second citizenship: OR 0.58, 95% CI 0.40-0.86; non–Swiss citizenship: OR 0.61, 95% CI 0.43-0.87 vs Swiss citizenship only), and language region (French-speaking region: OR 0.61, 95% CI 0.46-0.80; Italian-speaking region: OR 0.78, 95% CI 0.57-1.08 vs German-speaking region) were associated with lower app uptake.

By contrast, a higher monthly household income (OR 1.92, 95% CI 1.40-2.64 for an income >CHF 10,000 [US $11,000] vs income ≤CHF 6000 [US $6600]), more frequent internet use (daily [reference] vs less than weekly: OR 0.37, 95% CI 0.16-0.85), better adherence to mask-wearing recommendations (always or most of the time [reference] vs rarely or never OR 0.28, 95% CI 0.15-0.52), and nonsmoker status (OR 1.32, 95% CI 1.01-1.71) were associated with increased app uptake.

The same model was also applied to the random subsample (see Table 2), which provided additional information on trust in government and science (n=712). Of note, ORs of variables included in both multivariable models (ie, full and subsample) were not altered substantially, but CIs became wider due to the lower sample size. Furthermore, increasing levels of trust in government and health authorities were also associated with a higher likelihood of app uptake (OR 3.13, 95% CI 1.58-6.22 for high vs low [reference] trust), whereas the inclusion of trust in science did not improve the multivariable model fit.

Table 2. Results of a multivariable logistic regression analysis investigating factors associated with the use of the SwissCovid app

Characteristic	Value, odds ratio (95% CIs)		
	Univariable; full sample (N=1511)	Multivariable; full sample (N=1511)	Multivariable; random sub-sample interviewed on trust in government and science (n=712)
Age (per 10 years)	1 (0.99; 1.01)	0.99 (0.92, 1.06)	1.09 (0.98, 1.22)
Female Gender (vs male)	1.06 (0.87, 1.30)	1.10 (0.89, 1.36)	0.94 (0.68, 1.30)
Partnership status			
No partner	ref[a]	N/A[b]	N/A
Living with partner	1.19 (0.95, 1.50)	N/A	N/A
Not living with partner	0.85 (0.56, 1.27)	N/A	N/A
Has children (vs not)	0.87 (0.63, 1.21)	N/A	N/A
Citizenship status			
Swiss	ref	ref	ref
Swiss and other	0.58 (0.40, 0.85)	0.58 (0.40, 0.86)	0.52 (0.28, 0.96)
Non-Swiss	0.63 (0.45, 0.89)	0.61 (0.43, 0.87)	0.68 (0.39, 1.20)
Language region			
German	ref	ref	ref
French	0.69 (0.53, 0.89)	0.61 (0.46, 0.80)	0.56 (0.37, 0.84)
Ticino	0.79 (0.58, 1.08)	0.78 (0.57, 1.08)	0.90 (0.54, 1.51)
Education			
Only mandatory schooling	ref	ref	ref
Completed professional education	1.44 (0.92, 2.26)	1.32 (0.83, 2.12)	1.23 (0.57, 2.63)
University or university of applied sciences	1.85 (1.18, 2.90)	1.50 (0.94, 2.42)	1.58 (0.73, 3.45)
Currently employed (vs unemployed)	0.91 (0.73, 1.14)	N/A	N/A
Monthly household income			
≤CHF 6000 (US $6600)	ref	ref	ref
CHF 6000-10,000 (US $6600-11,000)	1.44 (1.10, 1.88)	1.29 (0.97, 1.71)	1.14 (0.75, 1.74)
>CHF 10,000 (US $11,000)	2.20 (1.64, 2.95)	1.92 (1.40, 2.64)	1.53 (0.94, 2.48)
No answer	1.31 (0.96, 1.79)	1.18 (0.85, 1.63)	1.06 (0.66, 1.71)
Nonsmoker (vs smoker)	1.40 (1.09, 1.81)	1.32 (1.01, 1.71)	1.51 (1.02, 2.25)
Self-reported chronic illness[c] (vs none)	1.08 (0.85, 1.36)	1.11 (0.87, 1.43)	0.88 (0.61, 1.27)
Use of protective masks			
Always or most of the time	ref	ref	ref
Sometimes	0.88 (0.71, 1.10)	0.75 (0.60, 0.96)	0.77 (0.54, 1.10)
Rarely or never	0.32 (0.18, 0.57)	0.28 (0.15, 0.52)	0.32 (0.12, 0.86)
Staying at home except for essential tasks			
Always or most of the time	ref	N/A	N/A
Sometimes	1.17 (0.91, 1.51)	N/A	N/A
Rarely or never	0.96 (0.73, 1.25)	N/A	N/A
Frequency of internet use			
Once daily or several times a day	ref	ref	ref
Once weekly or several days per week	0.55 (0.38, 0.78)	0.55 (0.38, 0.80)	0.59 (0.35, 1.00)
Never or less than once weekly	0.35 (0.16, 0.79)	0.37 (0.16, 0.85)	0.32 (0.11, 0.95)

Characteristic	Value, odds ratio (95% CIs)		
	Univariable; full sample (N=1511)	Multivariable; full sample (N=1511)	Multivariable; random subsample interviewed on trust in government and science (n=712)
Trust in government or health authorities[d]			
Little	ref	N/A	ref
Somewhat	2.16 (1.08, 4.32)	N/A	1.71 (0.82, 3.58)
Large	4.21 (2.22, 7.97)	N/A	3.13 (1.58, 6.22)
Trust in science[d]			
Little	ref	N/A	N/A
Somewhat	1.59 (0.84, 3.01)	N/A	N/A
Large	3.13 (1.71, 5.73)	N/A	N/A

[a]ref: reference value.

[b]N/A: data not applicable or not included because it did not improve model fit

[c]Presence of chronic illness was defined based on self-reporting of at least one of the following conditions: asthma, chronic obstructive pulmonary disease, diabetes, hypertension, cardiovascular disease, stroke, and cancer.

[d]Data only available in a randomly selected split-sample, including 47% (712/1511) of the full study population.

Reasons for App Nonuse

The responses of participants who reported not having used the app (808/1511) were analyzed further with respect to the reasons for app nonuse (Table 3). This group included both users who stated that they intended to use the app and those who did not intend to use it. Overall, the most important reasons for not installing the app were a perceived lack of usefulness of the app (297/808, 36.8%), followed by not having a suitable smartphone or operating system (184/808, 22.8%), and concerns about privacy (181/808, 22.4%). Other reasons (amounting to 18%) included lack of knowledge about the app, doubts about technological reliability, and concerns about excessive battery usage, among other reasons.

When compared to responses from Wave 8, the proportion of app nonusers (846/1508, 56.1%) who reported a perceived lack of app usefulness (228/846, 27%) was considerably lower. Moreover, the differences between waves for the other reasons of app nonuse (ie, not having the right phone: 221/846, 26.1%; privacy concerns: 202/846, 23.9%; and other reasons: 195/846, 23%; data not shown) were less pronounced.

As shown in Table 3, the distribution of reasons for app nonuse also varied with participants' intentions for using the app later (ie, maybe, no, or already uninstalled the app). While the lack of perceived benefits was the dominant reason for not installing the app (262/689, 38%) and having uninstalled (20/66, 30.3%) it, 34% (18/53) of the participants who intended to install the app at a later time point reported not having a compatible smartphone. Furthermore, it is noteworthy that excessive battery consumption also appeared to be an important reason for uninstalling the app (11/66, 16.7%).

The descriptive comparison of sociodemographic and other characteristics across the 3 major reasons for app nonuse (and a fourth category subsuming all other reasons; see Table 4) suggests that some reasons may be more prevalent in specific subgroups. The subpopulation citing problems with installing the app ("not the right phone") was the oldest (median age, 57.5 years), had the highest burden of chronic comorbidities (61/184, 33.2%), and tended to have high trust in government (large trust category: 69/89, 77.5%) and science (large trust category: 61/89, 68.5%) compared with the other subgroups. By contrast, those reporting privacy concerns for app nonuse were younger (median age: 44 years), more frequently living in the French-speaking part of Switzerland (65/181, 35.9%), and generally had less trust in the government (large trust category: 35/80, 43.8%) or science (large trust category: 30/80, 37.5%). No specific patterns were observed for the demographics of subpopulations reporting the remaining 2 reasons (ie, "not useful" and "other reasons").

Table 3. Reasons for nonuse of the SwissCovid app.

Reason	Value, n (%)			
	May install app later (n=53)	App not installed (n=689)	Uninstalled app (n=66)	All (n=808)
Perceived as not useful	15 (28.3)	262 (38)	20 (30.3)	297 (36.8)
Not the right phone	18 (34)	158 (22.9)	8 (12.1)	184 (22.8)
Concerned about privacy	8 (15.1)	164 (23.8)	9 (13.6)	181 (22.4)
Don't know the app	2 (3.8)	25 (3.6)	0 (0)	27 (3.3)
Technical doubts about reliability, maturity	1 (1.9)	20 (2.9)	4 (6.1)	25 (3.1)
Concerned about battery use	1 (1.9)	8 (1.2)	11 (16.7)	20 (2.5)
Don't believe in seriousness of Corona; lack of trust in government	0 (0)	9 (1.3)	1 (1.5)	10 (1.2)
Inertia, not had the time yet	5 (9.4)	2 (0.3)	0 (0)	7 (0.9)
Opposed out of principle, no specific reason	0 (0)	7 (1)	0 (0)	7 (0.9)
Don't want Bluetooth permanently on	0 (0)	3 (0.4)	2 (3)	5 (0.6)
Worried about consequences/quarantine	0 (0)	2 (0.3)	2 (3)	4 (0.5)
Currently outside of Switzerland	1 (1.9)	3 (0.4)	0 (0)	4 (0.5)
Would have to turn off app at work	0 (0)	2 (0.3)	1 (1.5)	3 (0.4)
Would feel stressed/scared by app use	0 (0)	2 (0.3)	0 (0)	2 (0.2)
Already protecting themselves, rarely leave the house	0 (0)	1 (0.1)	0 (0)	1 (0.1)

Table 4. Sociodemographic characteristics of SwissCovid app nonusers, stratified by the reason for app nonuse (3 most frequent reasons and "other").

Characteristic	Value			
	Not the right phone (n=184)	Privacy concerns (n=181)	Not useful (n=297)	Other reason (n=146)
Age, median (IQR)	57.5 (44.5, 67)	44 (35, 54)	46 (31, 57)	44 (31, 57)
Gender, female, n (%)	95 (51.6)	102 (56.4)	120 (40.4)	72 (49.3)
Partnership status, n (%)				
No partner	45 (24.5)	58 (32)	100 (33.7)	43 (29.5)
Living with partner	123 (66.8)	106 (58.6)	172 (57.9)	89 (61)
Not living with partner	16 (8.7)	17 (9.4)	25 (8.4)	14 (9.6)
Has children	15 (8.2)	21 (11.6)	27 (9.1)	29 (19.9)
Citizenship status, n (%)				
Swiss	152 (82.6)	132 (72.9)	232 (78.1)	108 (74)
Swiss and other	10 (5.4)	21 (11.6)	35 (11.8)	17 (11.6)
Non-Swiss	22 (12)	28 (15.5)	30 (10.1)	21 (14.4)
Language region, n (%)				
German	115 (62.5)	100 (55.2)	188 (63.3)	91 (62.3)
French	41 (22.3)	65 (35.9)	64 (21.5)	30 (20.5)
Ticino	28 (15.2)	16 (8.8)	45 (15.2)	25 (17.1)
Education, n (%)				
Only mandatory schooling	16 (8.7)	16 (8.8)	22 (7.4)	6 (4.1)
Completed professional education	90 (48.9)	83 (45.9)	157 (52.9)	76 (52.1)
University or university of applied sciences	78 (42.4)	82 (45.3)	118 (39.7)	64 (43.8)
Currently employed, n (%)	97 (52.7)	140 (77.3)	221 (74.4)	105 (71.9)
Monthly household income, n (%)				
≤CHF 6000 (US $6600)	65 (35.3)	52 (28.7)	87 (29.3)	42 (28.8)
CHF 6000-10,000 (US $6600-11,000)	62 (33.7)	48 (26.5)	104 (35)	47 (32.2)
>CHF 10,000 (US $11,000)	25 (13.6)	30 (16.6)	61 (20.5)	30 (20.5)
No answer	32 (17.4)	51 (28.2)	45 (15.2)	27 (18.5)
Smoker, n (%)	36 (19.6)	47 (26)	78 (26.3)	27 (18.5)
Self-reported chronic illness[a], n (%)	61 (33.2)	43 (23.8)	66 (22.2)	27 (18.5)
Use of protective masks, n (%)				
Always or most of the time	127 (69)	110 (60.8)	172 (57.9)	85 (58.2)
Sometimes	50 (27.2)	62 (34.3)	106 (35.7)	46 (31.5)
Rarely or never	7 (3.8)	9 (5)	19 (6.4)	15 (10.3)
Staying at home except for essential tasks, n (%)				
Always or most of the time	54 (29.3)	48 (26.5)	83 (27.9)	39 (26.7)
Sometimes	90 (48.9)	69 (38.1)	108 (36.4)	49 (33.6)
Rarely or never	40 (21.7)	64 (35.4)	106 (35.7)	58 (39.7)
Frequency of internet use, n (%)				
Once daily or several times a day	142 (77.2)	160 (88.4)	257 (86.5)	126 (86.3)
Once weekly or several days per week	32 (17.4)	14 (7.7)	36 (12.1)	17 (11.6)
Never or less than once weekly	10 (5.4)	7 (3.9)	4 (1.3)	3 (2.1)
Trust in government[b], n (%)				

Characteristic	Value			
	Not the right phone (n=184)	Privacy concerns (n=181)	Not useful (n=297)	Other reason (n=146)
Little	5 (5.6)	16 (20)	12 (8.6)	14 (21.2)
Somewhat	15 (16.9)	29 (36.3)	43 (30.7)	15 (22.7)
Large	69 (77.5)	35 (43.8)	85 (60.7)	37 (56.1)
Trust in science[b], n (%)				
Little	7 (7.9)	13 (16.3)	12 (8.6)	10 (15.2)
Somewhat	21 (23.6)	37 (46.3)	47 (33.8)	24 (36.4)
Large	61 (68.5)	30 (37.5)	80 (57.6)	32 (48.5)

[a]Presence of chronic illness was defined based on self-reporting of at least one of the following conditions: asthma, chronic obstructive pulmonary disease, diabetes, hypertension, cardiovascular disease, stroke, and cancer.

[b]Data only available in a randomly selected split-sample, including 50% of the full study population (n=712). New denominators for respective reasons of app nonuse were not right phone (n=89), privacy concerns (n=80), not useful (n=140), and other (n=66).

SwissCovid App Notifications and User Response

In the 3 survey waves (Waves 8, 9, and 10), a total of 15 participants reported having received an app notification: 2 users in Wave 8 (July), 6 users in Wave 9 (August), and 7 users in Wave 10 (October). Overall, 8 of these 15 (53.3%) users reported to have called the recommended infoline, whereas the remaining 6 users reported not to have undertaken any steps, and 1 user undertook other steps, which were left unspecified.

Since Wave 10, participants were also asked whether they had undergone COVID-19 testing in the past 4 weeks and, if so, what their test results were. Of the 5 users who called the infoline at Wave 10, 2 users reported to have undergone COVID-19 testing, and 1 of them reported to have tested positive for COVID-19.

Discussion

Principal Findings

By analyzing information on the use of the SwissCovid app from a longitudinal, web-based, panel survey, we evaluated factors related to the use of the DPT app in Switzerland.

Our data suggested that, 3 months after app release, 46.5% of the survey respondents had downloaded the app (of whom 38.8% had the SwissCovid app permanently activated). This proportion is an overestimation of actual app coverage in the general population and is most likely caused by the above-average affinity for such technologies of online panel participants. Moreover, social desirability might have led to some over-reporting of app use, despite this being an anonymous online survey [20]. In early October 2020, the official number of active app users was estimated at 1.6 million [12], which implies that around 1 in 4 (24.2%) adults residing in Switzerland were actively using the app. A recent modelling study suggests that this proportion of app uptake may, in fact, be sufficient to reduce the number of new infections to “manageable levels” [21].

We also deduced several population characteristics that may influence the uptake of the SwissCovid app. For example, younger age, higher income, or a nonsmoker status were associated with a higher app uptake. In contrast, characteristics such as foreign (non-Swiss) nationality or living in the French- or Italian-speaking regions of Switzerland were associated with a lower app uptake. Furthermore, app uptake was associated with the level of trust placed in the government and health authorities. Following recommended preventive measures and wearing masks, in particular, were also associated with a higher likelihood to use the app, which could imply higher levels of awareness, worry related to the COVID-19 pandemic, or increased health consciousness.

We further investigated participants’ stated reasons for nonuse of the SwissCovid app, which were dominated by technical aspects (ie, not having a suitable smartphone or operating system), privacy concerns, and perceived lack of usefulness. Ignorance or lack of information about the app did not seem to be a relevant reason, as only 3% of the participants cited this as a reason (ie, “don’t know the app”). Privacy concerns as a reason for app nonuse were associated with a lack of trust in the government and health authorities, as well as with a migration background. Participants whose app use was hindered owing to technical reasons seemed to be more trustful in the government but tended to be older. Therefore, streamlining installation processes and establishing compatibility with older phone devices may be worthwhile in order to increase app uptake among this subgroup.

Comparison With Previous Work

To our knowledge, this is the first study since the SwissCovid app release to systematically investigate DPT app uptake and the reasons for app nonuse in Switzerland. One survey was conducted since the app release in late-June 2020 in Switzerland, comprising 1000 Swiss individuals [22]; however, the data have not been published in detail. This previous study found that 43% of the Swiss population were using or considering using the Swiss proximity tracing app, with higher percentages observed among younger respondents. Our study results show similar proportions of app users.

Overall, our findings also correspond well with other international population surveys of DPT app use, most of which were performed before [23-29], and some after app release [30]. For example, several studies confirmed the finding that higher education status [25], and younger age [23] were associated with a higher willingness to use DPT apps. Of note, these characteristics could reflect the profile of typical early adopters of technology [31]. On the other hand, the observed sociodemographic patterns could also reflect generally higher health- and digital literacy among DPT app users. This alternative explanation raises concerns about the existence of a digital divide [32], in which individual who may benefit the most from the preventive effect of DPT are the least likely to use the app. Nevertheless, as suggested by a separate analysis of the COVID-19 Social Monitor project data, the majority of elderly persons adhered well to other preventive measures such as social distancing [33]. Furthermore, it can also be argued that transmission prevention among younger persons, who are disproportionately affected by the COVID-19 incidence in Switzerland, may also yield a protective effect for older adults.

This study adds to the scarce literature on motivations, as well as technical and nontechnical barriers for app use in settings where apps have already been deployed. Previous studies have utilized models based on psychology such as the health belief model (HBM) [29] and implementation science such as the normalization process theory (NPT) [10] to analyze adoption and nontechnical implementation challenges for DPT apps. Both these models emphasize the importance of a perceived benefit of preventive interventions, with HBM focusing on individuals and NPT focusing on a systems perspective [34]. The lack of perceived benefits was also a major reason for app nonuse in our survey, cited by 37% of all nonusers. Therefore, according to HBM and NPT, communicating usefulness of the app to individual and the society may be key to achieve greater app adoption; this can be achieved by testimonials of users who have had a positive experience with the DPT app. Moreover, optimizing economic incentives or removing existing disincentives for DPT use may further improve the benefit-risk balance [35,36]. In Switzerland, users who received an app notification were eligible for a free COVID-19 test, but quarantine was neither mandatory nor subject to salary compensations (which is currently being reconsidered).

Our observations of technical problems and persistent privacy concerns as reasons for nonuse or uninstallations of the SwissCovid app are consistent with those of a study in Australia [30], which reported similar user complaints. Furthermore, (a lack of) government trust emerged as a strong influencing factor for app usage in our and other surveys [23]. For example, approximately 11% of the respondents of the Australian survey cited government mistrust as a reason for not using the app [30]. In Switzerland, there was an early, rather strong consensus that the DPT app must be issued and managed by the government [37]. Nevertheless, the prevalent and persistent privacy concerns and trust issues remain to be challenging. For example, although the SwissCovid app implements privacy by design, the fact that the app relies on application programming interfaces provided by Google and Apple is sometimes still criticized. One solution to address this challenge could be to establish an independent oversight committee for the management of DPT operations [38]. Such committees would demonstrate transparency and increase public trust that governments will hold their promises, for example, to maintain voluntariness, prevent mission creep, or cease the use of DPT after the end of the pandemic.

Finally, we found evidence that external factors and the overall pandemic context have an effect on benefit and threat perceptions (as postulated by HBM [29]). In Switzerland, COVID-19 cases increased rapidly during the second half of October 2020, and the number of active app users also increased by 200,000 [12]. Similarly, the dominant reasons for app nonuse seem to have evolved in our study. Taken together, these observations suggest that public knowledge, perceptions, and app uptake respond dynamically to the pandemic situation. In line with these suggestions, first reports from vaccine development trials [39] raise hopes for a nearing availability of effective COVID-19 vaccines, which may also impact the public discourse on DPT apps. However, initial vaccination campaigns will likely focus on elderly subpopulations, who are at the greatest risk for a more severe disease course [40]. Given current limitations in vaccine production capacity, herd immunity will remain unachievable in the near future. Therefore, DPT apps will likely continue to play an important role in pandemic mitigation efforts, particularly among younger subpopulations wherein COVID-19 cases are often asymptomatic [40] and DPT app use is comparatively high.

Strengths and Limitations

Overall, our analysis contributes to the literature by being among the first studies to longitudinally investigate DPT app use patterns in the context of a changing pandemic. A key strength of our study is the availability of data from different survey waves, which allowed us to verify the robustness of our findings. Furthermore, our sample of 1500 participants is based on a random sample and is therefore likely to be quite representative with respect to age, gender, and language region for the Swiss population. However, we cannot fully exclude potential biases such as over-reporting or social desirability bias regarding app use. In addition, the fact that the Social Monitor project sample was drawn from an online panel population might have led to an overestimation of the app use among the general population.

Conclusions

To summarize, our study findings provide a clearer understanding of the motivations, barriers, and other factors associated with the uptake of DPT apps. Our data points toward complex interactions between motivations, trust, and incentives. Our study also reveals significant research gaps; for example, regarding how to effectively persuade persons with privacy concerns or how to create equitable incentives for app use. Similar studies are needed to evaluate the contribution of DPT on pandemic mitigation efforts, as well as generic, robust research methods to study privacy-preserving health technologies. From a practical perspective, our data suggest that DPT sponsors should scale-up communication efforts to not only build trust and mitigate privacy fears but also reduce technical challenges, as well as simplify onboarding procedures in order to reach a broader population, including persons with low digital or health literacy.

Acknowledgments

We thank the participants of the COVID-19 Social Monitor project for their important contribution to this study. The COVID-19 Social Monitor project has received funding from the Federal Office of Public Health and from Health Promotion Switzerland. The funders have no influence on the design, conduct, analyses, and publication of this study.

Authors' Contributions

VVW conceived and designed the work, conducted statistical analyses, and drafted the manuscript. CS, MSB, DM, TB, MH, AM, MK, AF, and MAP provided inputs on the analytic strategy. CS performed parts of the statistical analyses. MH, AM, and MAP designed the COVID-19 Social Monitor project and collected the data. MH and AM prepared the datasets for this project. All authors contributed to the interpretation of the data and critically revised the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Standardized questions on the use of SwissCovid app based on the COVID-19 Social Monitor project and sociodemographic characteristics of app users who were asked detailed questions about trust in health authorities or science.

[DOCX File , 19 KB - publichealth_v7i1e25701_app1.docx]

References

1. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss Med Wkly 2020 Mar 09;150:w20225 [FREE Full text] [doi: 10.4414/smw.2020.20225] [Medline: 32191813]
2. von Wyl V, Bonhoeffer S, Bugnion E, Puhan MA, Salathé M, Stadler T, et al. A research agenda for digital proximity tracing apps. Swiss Med Wkly 2020 Jul 13;150:w20324 [FREE Full text] [doi: 10.4414/smw.2020.20324] [Medline: 32672340]
3. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491) [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
4. Anglemyer A, Moore TH, Parker L, Chambers T, Grady A, Chiu K, et al. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database of Systematic Reviews 2020 Aug 18(8):CD013699. [doi: 10.1002/14651858.cd013699]
5. SARS-CoV-2 contact tracing strategy: epidemiologic and strategic considerations. Swiss National COVID-19 Science Task Force: Policy Briefs. 2020 Apr 26. URL: https://sciencetaskforce.ch/en/policy-brief/contact-tracing-strategy-2/ [accessed 2020-12-24]
6. Digital Proximity Tracing. Swiss National COVID-19 Science Task Force: Policy Briefs (2020). 2020 May 15. URL: https://sciencetaskforce.ch/en/policy-brief/digital-proximity-tracing-2/ [accessed 2020-12-24]
7. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
8. Apple / Google: Privacy-Preserving Contact Tracing. 2020. URL: https://www.apple.com/covid19/contacttracing [accessed 2020-12-24]
9. Coronavirus: Federal government to assume test costs, SwissCovid app to start on 25 June. Swiss Federal Council. 2020 Jun 24. URL: https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-79584.html [accessed 2020-12-26]
10. von Wyl V. Challenges for non-technical implementation of digital proximity tracing: early experiences from Switzerland. medRxiv. Preprint posted online on October 27, 2020. [doi: 10.1101/2020.10.22.20218057]
11. Salathé M, Althaus C, Anderegg N, Antonioli D, Ballouz T, Bugnon E, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. Swiss Med Wkly 2020 Dec 14;150:w20457 [FREE Full text] [doi: 10.4414/smw.2020.20457] [Medline: 33327003]
12. SwissCovid App Monitoring. Swiss Federal Statistical Office. 2020 Jun 25. URL: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.html [accessed 2020-12-26]
13. Servick K. COVID-19 contact tracing apps are coming to a phone near you. How will we know whether they work? Science.: American Association for the Advancement of Science; 2020 May 21. URL: https://www.sciencemag.org/news/2020/05/countries-around-world-are-rolling-out-contact-tracing-apps-contain-coronavirus-how [accessed 2020-12-24]
14. Zastrow M. Coronavirus contact-tracing apps: can they slow the spread of COVID-19? Nature 2020 May 19. [doi: 10.1038/d41586-020-01514-2] [Medline: 32433633]

15. Moser A, Carlander M, Wieser S, Hämmig O, Puhan MA, Höglinger M. The COVID-19 Social Monitor longitudinal online panel: Real-time monitoring of social and public health consequences of the COVID-19 emergency in Switzerland. PLoS One 2020;15(11):e0242129 [FREE Full text] [doi: 10.1371/journal.pone.0242129] [Medline: 33175906]
16. Rolf Weitkunat. Calculation methods for estimating the number of active SwissCovid apps. Swiss Federal Statistical Office. 2020 Jul 23. URL: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.assetdetail.13667538.html [accessed 2020-12-26]
17. Coronavirus: Situation in Switzerland. Swiss Federal Office of Public Health FOPH. URL: https://www.bag.admin.ch/bag/de/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html [accessed 2020-07-24]
18. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974 Dec;19(6):716-723. [doi: 10.1109/tac.1974.1100705]
19. Faraway JJ. Practical regression ANOVA using R. Bath: University of Bath; Jul 2002.
20. Höglinger M, Jann B. More is not always better: an experimental individual-level validation of the randomized response technique and the crosswise model. PLoS One 2018;13(8):e0201770 [FREE Full text] [doi: 10.1371/journal.pone.0201770] [Medline: 30106973]
21. López JAM, Arregui-GarcÍa B, Bentkowski P, Bioglio L, Pinotti F, Boëlle PY, et al. Anatomy of digital contact tracing: role of age, transmission setting, adoption and case detection. medrxiv 2020.
22. Mehr als die Hälfte der Schweizer Bevölkerung will SwissCovid-App nicht installieren Share Repräsentative Comparis-Befragung zur Nutzung der Covid-19-Tracing-App (Article in German). Comparis. 2020 Jul 09. URL: https://www.comparis.ch/comparis/press/medienmitteilungen/artikel/2020/digital/tracing-app/wird-kaum-installiert [accessed 2020-07-24]
23. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
24. Hargittai E, Redmiles EM, Vitak J, Zimmer M. Americans' willingness to adopt a COVID-19 tracking app. First Monday 2020 Oct 06. [doi: 10.5210/fm.v25i11.11095]
25. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact-tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
26. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
27. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2020 Oct 16 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
28. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
29. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
30. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian Government's COVIDSafe app: cross-sectional survey study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
31. Berwick DM. Disseminating innovations in health care. JAMA 2003 Apr 16;289(15):1969-1975. [doi: 10.1001/jama.289.15.1969] [Medline: 12697800]
32. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
33. Haag C, Höglinger M, Moser A, Hämmig O, Puhan M, von Wyl V. Social mixing and risk exposures for SARS-CoV-2 infections in elderly persons. Swiss Med Wkly 2020 Nov 30;150:w20416 [FREE Full text] [doi: 10.4414/smw.2020.20416] [Medline: 33277915]
34. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8(12):63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
35. Bonardi J, Brülhart M, Danthine JP, Saxena A, Thöni C, Thoenig M, et al. How to make digital proximity tracing work: the view from economics. 2020 Jun 24. URL: https://e4s.center/document/how-to-make-digital-proximity-tracing-work-the-view-from-economics/ [accessed 2020-12-26]
36. Loi M. How to fairly incentivise digital contact tracing. J Med Ethics 2020 Jul 09 [FREE Full text] [doi: 10.1136/medethics-2020-106388] [Medline: 32647047]

37. Hargittai E, Thouvenin F. Weite Teile der Bevölkerung sind bereit, eine Tracking-App zu nutzen wenn diese von Bund und Kantonen herausgegeben wird (Article in German). Neue Züricher Zeitung. 2020 May 02. URL: https://www.nzz.ch/schweiz/tracking-app-chancen-stehen-gut-ld.1554352?reduced=true [accessed 2020-12-26]
38. Blasimme A, Vayena E. What's next for COVID-19 apps? Governance and oversight. Science 2020 Nov 13;370(6518):760-762. [doi: 10.1126/science.abd9006] [Medline: 33184192]
39. Jackson L, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, mRNA-1273 Study Group. An mRNA vaccine against SARS-CoV-2 - preliminary Rreport. N Engl J Med 2020 Nov 12;383(20):1920-1931 [FREE Full text] [doi: 10.1056/NEJMoa2022483] [Medline: 32663912]
40. Moore S. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. medRxiv. Preprint posted online on September 24, 2020. [FREE Full text] [doi: 10.1101/2020.09.22.20194183]

Abbreviations

DPT: digital proximity tracing
DP-3T: decentralized, privacy-preserving proximity tracing
HBM: health belief model
NPT: normalization process theory
OR: odds ratio

Edited by T Sanchez; submitted 12.11.20; peer-reviewed by R Lucas, V Stara, K Liu; comments to author 27.11.20; revised version received 03.12.20; accepted 03.12.20; published 06.01.21.

Please cite as:
von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, Ballouz T, Menges D, Frei A, Puhan MA
Drivers of Acceptance of COVID-19 Proximity Tracing Apps in Switzerland: Panel Survey Analysis
JMIR Public Health Surveill 2021;7(1):e25701
URL: http://publichealth.jmir.org/2021/1/e25701/
doi:10.2196/25701
PMID:33326411

Original Paper

Mapping Research Trends of Universal Health Coverage From 1990 to 2019: Bibliometric Analysis

Mahboubeh Khaton Ghanbari[1], PhD; Masoud Behzadifar[2], PhD; Leila Doshmangir[3], PhD; Mariano Martini[4], PhD; Ahad Bakhtiari[5], PhD; Mahtab Alikhani[6], PhD; Nicola Luigi Bragazzi[4], MPH, MD, PhD

[1]Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran

[2]Social Determinants of Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran

[3]Tabriz Health Services Management Research Center, Iranian Center of Excellence in Health Management, Tabriz University of Medical Sciences, Tabriz, Iran

[4]Department of Health Sciences, School of Public Health, University of Genoa, Genoa, Italy

[5]Department of Health Management and Economics, School of Public Health, Tehran University of Medical Science, Tehran, Iran

[6]Department of Health Services Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Corresponding Author:
Masoud Behzadifar, PhD
Social Determinants of Health Research Center
Lorestan University of Medical Sciences
Anooshirvan Rezaei Square
Khorramabad, 6813833946
Iran
Phone: 98 9128934237
Email: masoudbehzadifar@gmail.com

Abstract

Background: Universal health coverage (UHC) is one of many ambitious, health-related, sustainable development goals. Sharing various experiences of achieving UHC, in terms of challenges, pitfalls, and future prospects, can help policy and decision-makers reduce the likelihood of committing errors. As such, scholarly articles and technical reports are of paramount importance in shedding light on the determinants that make it possible to achieve UHC.

Objective: The purpose of this study is to conduct a comprehensive analysis of UHC-related scientific literature from 1990 to 2019.

Methods: We carried out a bibliometric analysis of papers related to UHC published from January 1990 to September 2019 and indexed in Scopus via VOSviewer (version 1.6.13; CWTS). Relevant information was extracted: the number of papers published, the 20 authors with the highest number of publications in the field of UHC, the 20 journals with the highest number of publications related to UHC, the 20 most active funding sources for UHC-related research, the 20 institutes and research centers that have produced the highest number of UHC-related research papers, the 20 countries that contributed the most to the research field of UHC, the 20 most cited papers, and the latest available impact factors of journals in 2018 that included the UHC-related items under investigation.

Results: In our analysis, 7224 articles were included. The publication trend was increasing, showing high interest in the scientific community. Most researchers were from the United States, the United Kingdom, and Canada, with Thailand being a notable exception. The Lancet accounted for 3.95% of published UHC-related research. Among the top 20 funding sources, the World Health Organization (WHO), the Bill and Melinda Gates Foundation, and the National Institutes of Health (NIH) accounted for 1.41%, 1.34%, and 1.02% of published UHC-related research, respectively. The highest number of citations was found for articles published in The Lancet, the American Journal of Psychiatry, and the Journal of the American Medical Association (JAMA). The top keywords were "health insurance," "insurance," "healthcare policy," "healthcare delivery," "economics," "priority," "healthcare cost," "organization and management," "health services accessibility," "reform," "public health," and "health policy."

Conclusions: The findings of our study showed an increasing scholarly interest in UHC and related issues. However, most research concentrated in middle- and high-income regions and countries. Therefore, research in low-income countries should be promoted and supported, as this could enable a better understanding of the determinants of the barriers and obstacles to UHC achievement and improve global health.

(JMIR Public Health Surveill 2021;7(1):e24569) doi:10.2196/24569

KEYWORDS

bibliometrics; scientometrics; universal health coverage; universal health; health coverage; developing countries; low-income countries

Introduction

Universal health coverage (UHC) was one of the ambitious, health-related "sustainable development goals" (SDGs) set by the United Nations (UN) General Assembly in 2015, and is one of the top priorities of their 2030 agenda. UHC represents the hope for better health for the world's poorest [1-3]. The World Health Organization (WHO) has defined UHC as a policy for "ensuring that all people can use the promotive, preventive, curative, rehabilitative and palliative health services they need, of sufficient quality to be effective, while also ensuring that the use of these services does not expose the user to financial hardship" [4].

At least half of the world's population does not have access to full coverage for a package of essential health services [5]. Health expenses lead more than 100 million people worldwide to extreme poverty every year, often forcing people to make intolerably difficult choices between purchasing food for their children and families, paying for child education, or paying for vital health services [2,6].

Countries differ in the way they address UHC provision based on a wide range of factors, such as political, economic, social, epidemiological, and technical considerations [7,8]. The path to UHC involves important policy choices and inevitable trade-offs [9]. The extent of the impact of a successful UHC implementation is referred to as the "Third Global Health Transition" [10]. Sharing various experiences of achieving UHC, in terms of challenges, pitfalls, and future prospects, can help policy and decision-makers benefit from global good practices and reduce the likelihood of committing errors and wasting resources better allocated elsewhere. As such, scholarly articles and technical reports are of paramount importance in shedding light on the determinants that make UHC achievement possible [11,12].

Nearly all of the Organization for Economic Co-operation and Development (OECD) countries and emerging economies, such as Brazil, China, Colombia, Costa Rica, India, Indonesia, and Russia, have achieved UHC [13]. These countries' experiences can be a major source of evidence of why UHC is desirable and how it should be achieved. Evidence shows a strong relationship between life expectancy at birth and UHC indicators, reflecting the 3 core dimensions of universal health coverage [14]. In moving to UHC, some countries such as Ghana, Indonesia, and Vietnam have increased their UHC indices over time, 1.43%, 1.85%, and 2.26%, respectively, mostly by improving both financial protection and service coverage [15,16].

In recent years, researchers have been using scientometrics, a branch of information science and a subfield of bibliometrics, to quantitatively investigate emerging research patterns in the scientific literature [17]. In addition, scientometrics enables an assessment of trends in article citations and how these indicators and measurements can impact policy and management. Using scholarly databases and visualization technology allows researchers to gain a good understanding of the publication trends related to a given topic [18,19].

To the best of our knowledge, there is a dearth of information concerning research patterns in the field of health care management and, specifically, UHC. Therefore, the purpose of this study is to conduct a comprehensive analysis of UHC-related scientific literature from 1990 to 2019.

Methods

Ethics Approval and Consent to Participate

This study was waived from ethical approval because it did not include data on animals or human subjects, and it was based on publicly available data.

Data Sources

This quantitative study was based on medical informatics, data and text mining, and scientometrics techniques [20]. Independently, 2 authors searched Scopus from January 1, 1990, to September 24, 2019. Disagreements between them were resolved through discussion until consensus was reached.

Inclusion and Exclusion Criteria

We limited our search to only scholarly items dealing with UHC, using "universal health coverage" as the keyword. The search was performed without language restrictions. All records relevant to the field of UHC were deemed eligible and, as such, retained in our investigation.

Data Extraction

Data were downloaded in comma-separated values (CSV) format. Independently, 2 authors extracted relevant data, namely, (1) the number of documents published within the study period, (2) the 20 authors with the highest number of publications in the field of UHC, (3) the 20 journals with the highest number of publications related to UHC, (4) the 20 most active funding sources for UHC-related research, (5) the 20 institutes and research centers that have produced the highest number of UHC-related research papers, (6) the 20 countries that contributed the most to the research field of UHC, (7) the 20 most highly cited papers, and (8) the latest available impact factor of journals in 2018 that included the UHC-related items under investigation. Any disagreements between the 2 authors were resolved through discussion until consensus was reached.

Data Analysis

Ad hoc visualization software was used to visualize UHC-related research hotspots, patterns, directions of research development, and other relevant trends, using networks and graphs. All data were imported and loaded into VOSviewer (version 1.6.13; CWTS). For visualization publication density worldwide (ie, publication trends among countries), the open-source tool GunnMap was used [21].

Results

After searching Scopus, a pool of 7224 records was included in our analysis. The increasing publication trend related to UHC from January 1990 to September 2019 is shown in Table 1.

The 20 authors with the highest number of publications in the field of UHC are listed in Table 2. Of the 20 authors, 4 are from the United States, 4 are from the United Kingdom, and 3 are from Thailand.

The network distribution of authors publishing in the field of UHC is shown in Figure 1. The 20 journals with the highest number of publications related to UHC are listed in Table 3. *The Lancet* accounted for 3.95% of published UHC-related research.

Table 1. Number of publications related to universal health coverage per year, as indexed in Scopus.

Year	Number of publications
1990	25
1991	21
1992	38
1993	50
1994	103
1995	70
1996	53
1997	70
1998	77
1999	73
2000	115
2001	83
2002	87
2003	138
2004	129
2005	136
2006	188
2007	266
2008	254
2009	308
2010	251
2011	312
2012	393
2013	398
2014	449
2015	581
2016	628
2017	668
2018	784
2019	525

Table 2. Authors with the highest number of manuscripts related to universal health coverage.

Rank	Author's name	Country	Number of publications	Citations	Percentage (n/7224)	H-index
1	Tangcharoensathien V	Thailand	47	3117	0.64	29
2	Atun R	United States	35	9121	0.48	48
3	Teerawattananon Y	Singapore	31	3865	0.42	27
4	Chalkidou K	United Kingdom	23	2223	0.31	22
5	McIntyre D	South Africa	23	2197	0.31	27
6	Norheim OF	Norway	23	19675	0.31	42
7	Ridde V	Canada	23	2455	0.31	24
8	Hanson K	United Kingdom	21	5268	0.28	39
9	McKee M	United Kingdom	21	51327	0.28	96
10	Mills A	United Kingdom	21	9487	0.28	56
11	Ataguba JE	South Africa	20	893	0.27	15
12	Shibuya K	Japan	20	58828	0.27	66
13	Bello AK	Canada	19	7133	0.26	32
14	Kruk ME	United States	19	4671	0.26	38
15	Woolhandler S	United States	19	10241	0.26	47
16	Limwattananon S	Thailand	18	10241	0.24	47
17	Patcharanarumol W	Thailand	18	714	0.24	12
18	Prinja S	India	18	1087	0.24	19
19	Reich MR	United States	18	3580	0.24	32
20	Bhutta ZA	Pakistan	17	69758	0.23	114

Figure 1. The distribution of authors publishing research in the field of universal health coverage.

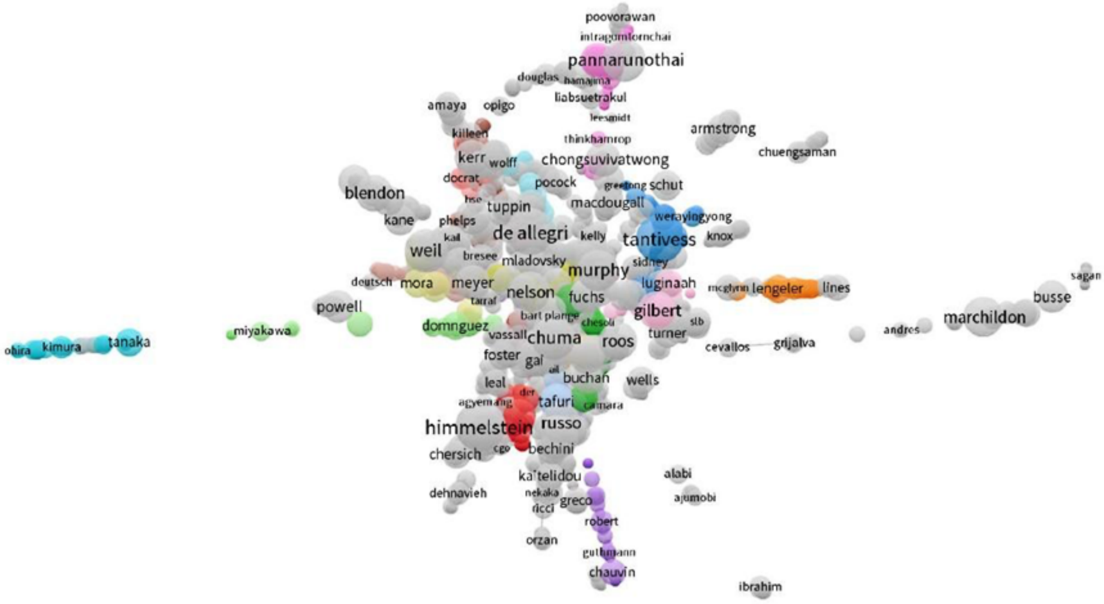


Table 3. Journals with the highest number of articles related to universal health coverage.

Rank	Journal	Number of publications	Percentage (n/7224)	Impact factor (2018)	Quartile in category (2018)	H-index
1	Lancet	286	3.95	59.102	Q1	700
2	Health Affairs	152	2.09	5.711	Q1	156
3	Plos One	152	2.09	2.776	Q1	268
4	BMC Health Services Research	131	1.8	1.932	Q1	90
5	Bulletin of The World Health Organization	124	1.71	6.818	Q1	148
6	Modern Healthcare	120	1.65	—[a]	Q4	9
7	International Journal for Equity in Health	115	1.58	2.473	Q1	46
8	Health Policy and Planning	105	1.44	2.717	Q1	80
9	Health Policy	95	1.31	2.075	Q1	79
10	Social Science and Medicine	93	1.28	3.087	Q1	213
11	BMC Public Health	82	1.13	2.567	Q1	117
12	Vaccine	81	1.11	4.760	Q1	164
13	Malaria Journal	77	1.06	2.798	Q1	87
14	New England Journal of Medicine	75	1.03	70.670	Q1	933
15	BMJ Global Health	63	0.86	4.28	Q1	21
16	Journal of Health Politics Policy and Law	54	0.74	1.839	Q2	48
17	American Journal of Public Health	52	0.71	0.774	Q1	236
18	Health Systems and Reform	51	0.7	—	—	—
19	International Journal of Health Planning and Management	50	0.68	1.450	Q2	37
20	Global Health Action	50	0.68	1.817	Q1	33

[a] —not available.

Table 4 shows the 20 most active funding sources for UHC-related research. Among them, the WHO, the Bill and Melinda Gates Foundation, and the National Institutes of Health (NIH) accounted for 1.41%, 1.34%, and 1.02% of published UHC-related research, respectively.

Table 5 lists the 20 institutes and research centers that have produced the highest number of UHC-related research papers.

Table 6 shows the countries that contributed the most to the research field of UHC. Among them, the United States, the United Kingdom, and Canada contributed 2426, 919, and 545 papers, respectively.

Figure 2 shows the density distribution of UHC-related publications among different countries and regions around the world.

The 20 most highly cited papers are listed in Table 7. The highest number of citations was found for papers published in *The Lancet*, the *American Journal of Psychiatry*, and the *Journal of the American Medical Association* (JAMA).

In Figure 3, the network of words, themes, and topics associated with UHC is shown. Among them, the top keywords were "health insurance," "insurance," "healthcare policy," "healthcare delivery," "economics," "priority," "healthcare cost," "organization and management," "health services accessibility," "reform," "public health," and "health policy."

Table 4. Most active funding sources for universal health coverage (UHC)-related research.

Rank	Name of Institute	Number of publications
1	World Health Organization	102
2	London School of Hygiene & Tropical Medicine	97
3	Harvard School of Public Health	74
4	University of Toronto	65
5	Harvard Medical School	47
6	University of Cape Town	45
7	Johns Hopkins Bloomberg School of Public Health	39
8	Imperial College London	34
9	Centers for Disease Control and Prevention	33
10	Thailand Ministry of Public Health	30
11	University of California, San Francisco	26
12	Johns Hopkins University	24
13	University of Oxford	22
14	University of Washington, Seattle	22
15	University of Witwatersrand	21
16	Harvard University	21
17	Columbia University in the City of New York	20
18	The World Bank	19
19	UCL	18
20	University of Melbourne	17

Table 5. Highest producing institutes and research centers for universal health coverage research.

Institute	Number of publications	Percentage of total
Organisation Mondiale de la Santé	388	5.35
London School of Hygiene & Tropical Medicine	269	3.71
Harvard School of Public Health	194	2.67
University of Toronto	164	2.26
Harvard Medical School	147	2.02
University of Cape Town	112	1.54
Johns Hopkins Bloomberg School of Public Health	104	1.43
Imperial College London	104	1.43
Centers for Disease Control and Prevention	102	1.4
Thailand Ministry of Public Health	98	1.35
University of California, San Francisco	93	1.28
Johns Hopkins University	89	1.22
University of Oxford	87	1.2
University of Washington, Seattle	86	1.18
University of Witwatersrand	79	1.09
Harvard University	78	1.07
Columbia University in the City of New York	75	1.03
The World Bank, USA	73	1
UCL	69	0.95
University of Melbourne	68	0.93

Table 6. Countries and regions that contributed the most to the research field of universal health coverage (UHC) during 1990-2019.

Country	Number of UHC-related research papers contributed
United States	2426
United Kingdom	919
Canada	545
Switzerland	469
India	395
Australia	370
South Africa	299
Thailand	285
Brazil	219
China	215
France	205
Japan	181
Italy	176
Netherlands	173
Germany	161
Spain	158
Belgium	149
Mexico	131
Taiwan	129
Kenya	120

Figure 2. Density of publications related to the research field of universal health coverage worldwide.

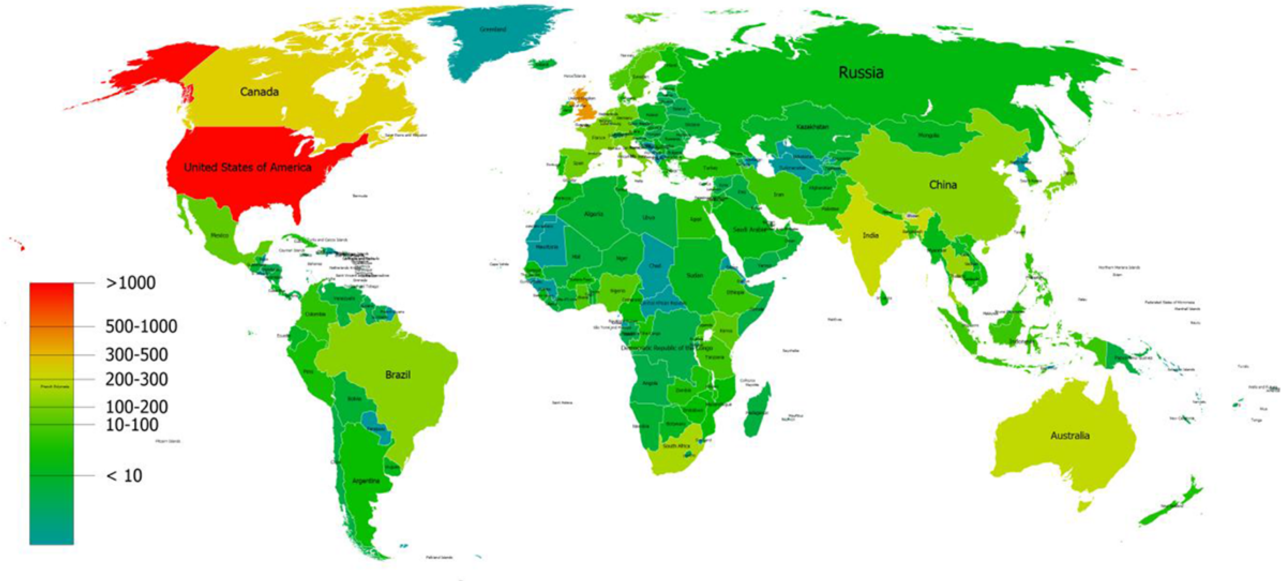
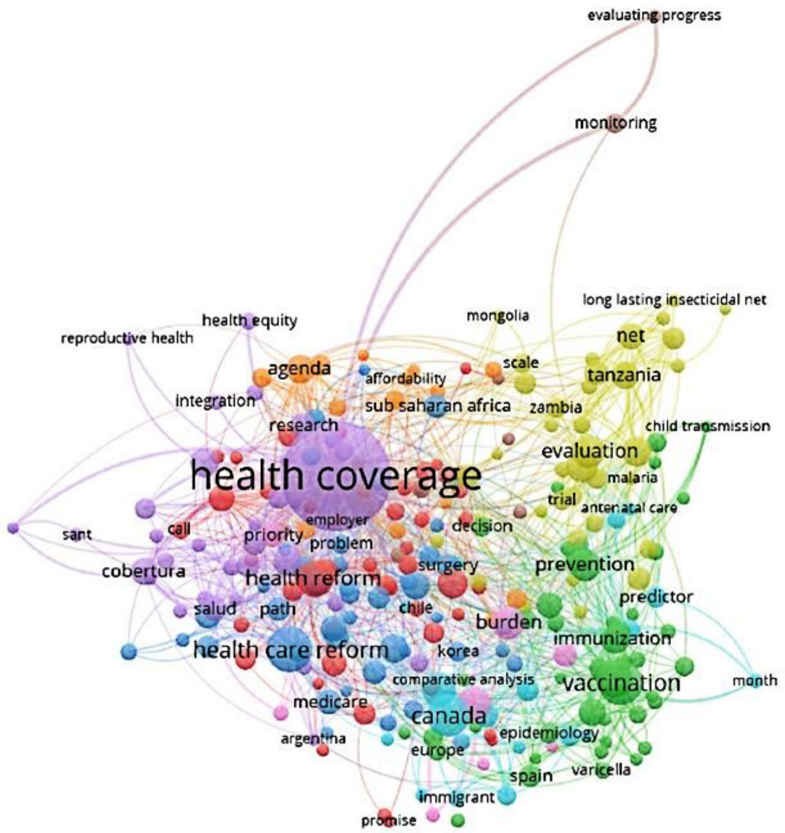


Table 7. Most cited papers related to universal health coverage.

No.	Title	Year	Journal	Number of citations
1	Evidence-based, cost-effective interventions: How many newborn babies can we save?	2005	Lancet	933
2	Social consequences of psychiatric disorders, I: Educational attainment	1995	American Journal of Psychiatry	729
3	Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development	2015	Lancet	716
4	Socioeconomic Inequalities in Health: No Easy Solution	1993	JAMA	696
5	Hepatitis B virus infection: Epidemiology and vaccination	2006	Epidemiologic Reviews	615
6	Persistence of use of lipid-lowering medications: A cross-national study	1998	Journal of the American Medical Association	550
7	Early appraisal of China's huge and complex health-care reforms	2012	Lancet	541
8	Applying an equity lens to child health and mortality: More of the same is not enough	2003	Lancet	485
9	Taiwan's new national health insurance program: Genesis and experience so far	2003	Health Affairs	458
10	Varicella disease after introduction of varicella vaccine in the United States, 1995-2000	2002	Journal of the American Medical Association	427
11	Does universal health insurance make health care unaffordable? Lessons from Taiwan	2003	Health Affairs	406
12	Maternal and child health in Brazil: Progress and challenges	2011	Lancet	402
13	Establishment of a universal size standard strain for use with the pulsenet standardized pulsed-field gel electrophoresis protocols: Converting the national databases to the new size standard	2005	Journal of Clinical Microbiology	400
14	Help-seeking and access to mental health care in a university student population	2007	Medical Care	369
15	Policy statement: Recommendations for the prevention of pneumococcal infections, including the use of pneumococcal conjugate vaccine (Prevnar), pneumococcal polysaccharide vaccine, and antibiotic prophylaxis	2008	Pediatrics	348
16	Explaining income-related inequalities in doctor utilisation in Europe	2004	Health Economics	345
17	Access to care, health status, and health disparities in the United States and Canada: Results of a Cross-National Population-Based Survey	2006	American Journal of Public Health	325
18	The history and principles of managed competition	1993	Health Affairs	317
19	Socioeconomic Disparities in Preventive Care Persist Despite Universal Coverage: Breast and Cervical Cancer Screening in Ontario and the United States	1994	JAMA	310
20	Policy relevant determinants of health: An international perspective	2002	Health Policy	307

Figure 3. Network of the most used keywords related to universal health coverage.



Discussion

Principal Findings

This study quantitatively assessed the publication trend related to UHC over the past 19 years. UHC-related publications have been on the rise in recent years, and this seems to be the major focus of researchers, given the important role that UHC can play in improving equity in access to health services and provisions. UHC can enable important achievements in the health sector worldwide. The growth of health-related scientific publications in the field of policy and management, and especially UHC, reflects the global interest and participation of different stakeholders, including researchers, in identifying the different dimensions and determinants that can make it possible to achieve UHC.

Undoubtedly, relying on scholarly publications can improve the performance of the health sector to achieve UHC-related goals. The rigor of the scientific method, if properly followed, can lead to fundamental changes in all areas of life, including health. Bibliometrics-based literature reviews can play an important role in examining the process of scientific publications and orienting researchers in this field [22].

From 1990 to 2019, scholarly publications in the field of UHC have been gradually increasing, especially after 2015 when policy and decision-makers have given particular emphasis to achieving UHC as one of the SDGs. Political commitment and support on this issue has contributed to the prioritization of UHC and put it on the policy and research agenda [23].

This investigation shows that authors from the United States, the United Kingdom, Canada, and Thailand produced the highest number of publications related to UHC. Scientists from the United States, the United Kingdom, and Canada have done research on the possible ways to achieve UHC goals in collaboration with various stakeholders, including health care policy and decision-makers. Thailand is one of the countries working hard to improve its health sector by making profound reforms. Since 2002, despite economic-financial problems and political instability, proper support for UHC has provided Thai citizens with a good level of health services and provisions. Therefore, researchers in this country have tried to disseminate their experiences and practices in the field of UHC to other countries around the world [24].

Usually, researchers aim to have their scientific findings published in prestigious journals so that their papers can have the highest exposure in terms of impact and receive adequate attention and citations from other researchers [25]. *The Lancet*, which has a high impact factor and plays an important role in influencing and shaping future scientific research, has published the highest number of articles related to UHC. Also, journals in the fields of health care policymaking, decision-making, and management have attracted authors' interest in submitting papers. UHC is a major topic because of its impact on all aspects of health [26].

The WHO, the London School of Hygiene and Tropical Medicine, and the Harvard School of Public Health were among the institutions and research centers that played a major role in supporting UHC-related research. The WHO's institutional nature makes it naturally interested in topics such as UHC, as it strives to provide the best evidence for a given health-related issue. The London School of Hygiene and Tropical Medicine is also one of the most prestigious institutions that, in recent decades, has promoted UHC-related studies, especially in lower-income countries, to improve health in these countries and achieve UHC goals. It strives to empower researchers in the field of health and provide high-quality public health education, as does Harvard School of Public Health.

It is important to note that, in the last decades, these two institutions have become prominent in the fields of health policy and management, indicating that they play an important role in developing UHC-related issues.

Limitations

Despite strengths such as methodological rigor, transparency, and reproducibility, this study has some limitations that should be properly recognized. Its major limitation is the use of a single bibliographic database (Scopus). As such, results should be replicated utilizing other major scholarly databases like PubMed/MEDLINE or Web of Science.

Conclusion

The findings of our study showed an increasing scholarly interest in UHC and related issues. However, most researchers were from the United States, the United Kingdom, and Canada, with Thailand being a notable exception. Research in low-income countries should be promoted and supported, as this could enable a better understanding of the determinants of the barriers and obstacles to UHC achievement and improve global health.

Authors' Contributions

MB, LD, and NLB conceived of the premise for this paper. MB, MM, AB, and MKG designed the research. MB, LD, AB, NLB, and MA conceived of the methodology. MB, LD, NLB, and AB performed the data analysis, graphics, and data interpretation. MB, LD, MM, AB, MKG, and NLB wrote and submitted the manuscript.

Conflicts of Interest

None declared.

References

1. Kruk ME, Gage AD, Joseph NT, Danaei G, García-Saisó S, Salomon JA. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. The Lancet 2018 Nov;392(10160):2203-2212. [doi: 10.1016/s0140-6736(18)31668-4]
2. Evans DB, Hsu J, Boerma T. Universal health coverage and universal access. Bull. World Health Organ 2013 Aug 01;91(8):546-546A. [doi: 10.2471/blt.13.125450]
3. Hogan DR, Stevens GA, Hosseinpoor AR, Boerma T. Monitoring universal health coverage within the Sustainable Development Goals: development and baseline data for an index of essential health services. The Lancet Global Health 2018 Feb;6(2):e152-e168. [doi: 10.1016/s2214-109x(17)30472-2]
4. Pandey KR. From health for all to universal health coverage: Alma Ata is still relevant. Global Health 2018 Jul 3;14(1):62. [doi: 10.1186/s12992-018-0381-6]
5. Giovanella L, Mendoza-Ruiz A, Pilar A, Rosa M, Martins G, Santos I. Universal health system and universal health coverage: assumptions and strategie. Cien Saude Colet 2018;23(6):1763-1776. [doi: 10.1093/oso/9780190662455.003.0011]
6. Garg S. Universal health coverage in India: Newer innovations and the role of public health. Indian J Public Health 2018;62(3):167-170. [doi: 10.4103/ijph.ijph_221_18]
7. Gupta V, Kerry VB, Goosby E, Yates R. Politics and Universal Health Coverage — The Post-2015 Global Health Agenda. N Engl J Med 2015 Sep 24;373(13):1189-1192. [doi: 10.1056/nejmp1508807]
8. Greer SL, Méndez CA. Universal Health Coverage: A Political Struggle and Governance Challenge. Am J Public Health 2015 Nov;105(S5):S637-S639. [doi: 10.2105/ajph.2015.302733]
9. Assan A, Takian A, Aikins M, Akbarisari A. Challenges to achieving universal health coverage through community-based health planning and services delivery approach: a qualitative study in Ghana. BMJ Open 2019 Feb 22;9(2):e024845. [doi: 10.1136/bmjopen-2018-024845]
10. Yuan B, Balabanova D, Gao J, Tang S, Guo Y. Strengthening public health services to achieve universal health coverage in China. BMJ 2019 Jun 21;365:l2358. [doi: 10.1136/bmj.l2358]
11. Binagwaho A, Adhanom Ghebreyesus T. Primary healthcare is cornerstone of universal health coverage. BMJ 2019 Jun 03;365:l2391. [doi: 10.1136/bmj.l2391]
12. Sanogo NA, Fantaye AW, Yaya S. Universal Health Coverage and Facilitation of Equitable Access to Care in Africa. Front. Public Health 2019 Apr 26;7:102. [doi: 10.3389/fpubh.2019.00102]
13. Limwattananon S, Tangcharoensathien V, Tisayathicom K, Boonyapaisarncharoen T, Prakongsai P. Why has the Universal Coverage Scheme in Thailand achieved a pro-poor public subsidy for health care? BMC Public Health 2012;12(Suppl 1):S6. [doi: 10.1186/1471-2458-12-s1-s6]

14. Tangcharoensathien V, Mills A, Palu T. Accelerating health equity: the key role of universal health coverage in the Sustainable Development Goals. BMC Med 2015 Apr 29;13(1):101. [doi: 10.1186/s12916-015-0342-3]
15. Tangcharoensathien V, Tisayaticom K, Suphanchaimat R, Vongmongkol V, Viriyathorn S, Limwattananon S. Financial risk protection of Thailand's universal health coverage: results from series of national household surveys between 1996 and 2015. Int J Equity Health 2020 Jan 13;19(1):163. [doi: 10.21203/rs.2.20691/v1]
16. Wagstaff A, Flores G, Hsu J, Smitz M, Chepynoga K, Buisman LR, et al. Progress on catastrophic health spending in 133 countries: a retrospective observational study. The Lancet Global Health 2018 Feb;6(2):e169-e179. [doi: 10.1016/s2214-109x(17)30429-1]
17. Wang M, Li W, Tao Y, Zhao L. Emerging trends and knowledge structure of epilepsy during pregnancy research for 2000-2018: a bibliometric analysis. PeerJ 2019;7:e7115. [doi: 10.7717/peerj.7115]
18. Zhao F, Du F, Shi D, Zhou W, Jiang Y, Ma L. Mapping research trends of retinal vein occlusion from 2009 to 2018: a bibliometric analysis. PeerJ 2019;7:e7603. [doi: 10.7717/peerj.7603]
19. Zou X, Yue WL, Vu HL. Visualization and analysis of mapping knowledge domain of road safety studies. Accident Analysis & Prevention 2018 Sep;118:131-145. [doi: 10.1016/j.aap.2018.06.010]
20. Wallin JA. Bibliometric Methods: Pitfalls and Possibilities. Basic Clin Pharmacol Toxicol 2005 Nov;97(5):261-275. [doi: 10.1111/j.1742-7843.2005.pto_139.x]
21. 2019 (30 September 2019). GunnMap 2. URL: http://lert.co.nz/map/ [accessed 2019-09-30]
22. Burton C, Elliott A, Cochran A, Love T. Do healthcare services behave as complex systems? Analysis of patterns of attendance and implications for service delivery. BMC Med 2018 Sep 7;16(1):138. [doi: 10.1186/s12916-018-1132-5]
23. Gonani A, Muula A. Point of View: The importance of Leadership towards universal health coverage in Low Income Countries. Mal. Med. J 2015 Apr 24;27(1):34-37. [doi: 10.4314/mmj.v27i1.9]
24. Sumriddetchkajorn K, Shimazaki K, Ono T, Kusaba T, Sato K, Kobayashi N. Universal health coverage and primary care, Thailand. Bull. World Health Organ 2019 Apr 01;97(6):415-422. [doi: 10.2471/blt.18.223693]
25. Sandesh N, Wahrekar S. Choosing the scientific journal for publishing research work: perceptions of medical and dental researchers. Medicine and Pharmacy Reports 2017 Apr 26;90(2):196-202. [doi: 10.15386/cjmed-704]
26. Merigó JM, Núñez A. Influential journals in health research: a bibliometric study. Global Health 2016 Aug 22;12(1):46. [doi: 10.1186/s12992-016-0186-4]

Abbreviations

SDGs: sustainable development goals
UHC: universal health coverage
UN: United Nations
WHO: World Health Organization

Edited by T Sanchez; submitted 24.09.20; peer-reviewed by HA Gorji, S Vatankhah, M Saran; comments to author 29.09.20; revised version received 15.11.20; accepted 02.12.20; published 11.01.21.

Please cite as:
Ghanbari MK, Behzadifar M, Doshmangir L, Martini M, Bakhtiari A, Alikhani M, Bragazzi NL
Mapping Research Trends of Universal Health Coverage From 1990 to 2019: Bibliometric Analysis
JMIR Public Health Surveill 2021;7(1):e24569
URL: http://publichealth.jmir.org/2021/1/e24569/
doi:10.2196/24569
PMID:33427687

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org