JMIR Public Health and Surveillance

A multidisciplinary journal that focuses on public health and technology, public health informatics, mass media campaigns, surveillance, participatory epidemiology, and innovation in public health practice and research

Volume 7 (2021), Issue 2 ISSN: 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Characterizing the COVID-19 Infodemic on Chinese Social Media: Exploratory Study (e26090)
Shuai Zhang, Wenjing Pian, Feicheng Ma, Zhenni Ni, Yunmei Liu. 4

The Influence of Media Coverage and Governmental Policies on Google Queries Related to COVID-19 Cutaneous Symptoms: Infodemiology Study (e25651)
Solene Huynh Dagher, Guillaume Lamé, Thomas Hubiche, Khaled Ezzedine, Tu Duong. 24

General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis (e24429)
Katja Reuter, Melissa Wilson, Meghan Moran, NamQuyen Le, Praveen Angyan, Anuja Majmundar, Elsi Kaiser, Jennifer Unger. 35

Key Population Hotspots in Nigeria for Targeted HIV Program Planning: Mapping, Validation, and Reconciliation (e25623)
Julia Lo, Samuel Nwafor, Amee Schwitters, Andrew Mitchell, Victor Sebastian, Kristen Stafford, Idoteyin Ezirim, Man Charurat, Anne McIntyre.
4 8

Development of a Web-Based System for Exploring Cancer Risk With Long-term Use of Drugs: Logistic Regression Approach (e21401)
Hsuan-Chia Yang, Md Islam, Phung Nguyen, Ching-Huan Wang, Tahmina Poly, Chih-Wei Huang, Yu-Chuan Li. 73

Conducting Virtual, Synchronous Focus Groups Among Black Sexual Minority Men: Qualitative Study (e22980)
Derek Dangerfield II, Charleen Wylie, Janeane Anderson. 88

Fostering Engagement With Health and Housing Innovation: Development of Participant Personas in a Social Housing Cohort (e25037)
Andrew Williams, Tamaryn Menneer, Mansi Sidana, Tim Walker, Kath Maguire, Markus Mueller, Cheryl Paterson, Michael Leyshon, Catherine Leyshon, Emma Seymour, Zoë Howard, Emma Bland, Karyn Morrissey, Timothy Taylor. 98

The Global Infectious Diseases Epidemic Information Monitoring System: Development and Usability Study of an Effective Tool for Travel Health Management in China (e24204)
Dayong Gu, Jianan He, Jie Sun, Xin Shi, Ying Ye, Zishuai Zhang, Xiangjun Wang, Qun Su, Wenjin Yu, Xiaopeng Yuan, Ruiling Dong. 114

Expert Opinions on the Most Promising Treatments and Vaccine Candidates for COVID-19: Global Cross-sectional Survey of Virus Researchers in the Early Months of the Pandemic (e22483)
Bernardo Cabral, Luiza Braga, Fabio Mota. 129

XSL•FO
RenderX

Evaluating the Effectiveness of a Web-Based Program (POP4Teens) to Prevent Prescription Opioid Misuse Among Adolescents: Randomized Controlled Trial (e18487)
Lisa Marsch, Sarah Moore, Michael Grabinski, Sarah Bessen, Jacob Borodovsky, Emily Scherer. 158

Clinical Characterization of Patients With COVID-19 in Primary Care in Catalonia: Retrospective Observational Study (e25452)
Miguel Mayer, Josep Vidal-Alaball, Anna Puigdellívol-Sánchez, Francesc Marín Gomez, Angela Leis, Jacobo Mendioroz Peña. 184

Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach (e24585)
Tiago de Melo, Carlos Figueiredo. 208

Promoting Mask Use on TikTok: Descriptive, Cross-sectional Study (e26392)
Corey Basch, Joseph Fera, Isabela Pierce, Charles Basch. 227

Evaluating Apple Inc Mobility Trend Data Related to the COVID-19 Outbreak in Japan: Statistical Analysis (e20335)
Junko Kurita, Yoshiyuki Sugishita, Tamie Sugawara, Yasushi Ohkusa. 234

Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation (e24445)
Mark Wong, Dinesh Gunasekeran, Simon Nusinovici, Charumathi Sabanayagam, Khung Yeo, Ching-Yu Cheng, Yih-Chung Tham. 241

Public Knowledge, Attitudes, and Practices Related to COVID-19 in Iran: Questionnaire Study (e21415)
Mohsen Abbasi-Kangevari, Ali-Asghar Kolahi, Seyyed-Hadi Ghamari, Hossein Hassanian-Moghaddam. 253

Distance Management of Spinal Disorders During the COVID-19 Pandemic and Beyond: Evidence-Based Patient and Clinician Guides From the Global Spine Care Initiative (e25484)
Scott Haldeman, Margareta Nordin, Patricia Tavares, Rajani Mullerpatan, Deborah Kopansky-Giles, Vincent Setlhare, Roger Chou, Eric Hurwitz, Caroline Treanor, Jan Hartvigsen, Michael Schneider, Ralph Gay, Jean Moss, Joan Haldeman, David Gryfe, Adam Wilkey, Richard Brown, Geoff Outerbridge, Stefan Eberspaecher, Linda Carroll, Reginald Engelbrecht, Kait Graham, Nathan Cashion, Stefanie Ince, Erin Moon. 267

Short Paper

Instagram Posts Related to Backwoods Cigarillo Blunts: Content Analysis (e22946)
Stephanie Kim, Alia Mourali, Jon-Patrick Allem, Jennifer Unger, Tess Boley Cruz, Sabrina Smiley. 18

Protocol

Influenza and Respiratory Virus Surveillance, Vaccine Uptake, and Effectiveness at a Time of Cocirculating COVID-19: Protocol for the English Primary Care Sentinel System for 2020-2021 (e24341)
Simon de Lusignan, Jamie Lopez Bernal, Rachel Byford, Gayatri Amirthalingam, Filipa Ferreira, Oluwafunmi Akinyemi, Nick Andrews, Helen Campbell, Gavin Dabrera, Alexandra Deeks, Alex Elliot, Else Krajenbrink, Harshana Liyanage, Dylan McGagh, Cecilia Okusi, Vaishnavi Parimalanathan, Mary Ramsay, Gillian Smith, Manasa Tripathy, John Williams, William Victor, Maria Zambon, Gary Howsam, Brian Nicholson, Victoria Tzortziou Brown, Christopher Butler, Mark Joy, FD Hobbs. 59

XSL•FO
RenderX

Corrigenda and Addenda

Correction: General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis (e28131)
Katja Reuter, Melissa Wilson, Meghan Moran, NamQuyen Le, Praveen Angyan, Anuja Majmundar, Elsi Kaiser, Jennifer Unger. 86

Review

Methods for Social Media Monitoring Related to Vaccination: Systematic Scoping Review (e17149)
Emilie Karafillakis, Sam Martin, Clarissa Simas, Kate Olsson, Judit Takacs, Sara Dada, Heidi Larson. 141

Viewpoints

COVID-19 and Black, Asian, and Minority Ethnic Communities: A Complex Relationship Without Just Cause (e22581)
Peter Phiri, Gayathri Delanerolle, Ayaat Al-Sudani, Shanaya Rathod. 176

Collaborating in the Time of COVID-19: The Scope and Scale of Innovative Responses to a Global Pandemic (e25935)
Theresa Bernardo, Kurtis Sobkowich, Russell Forrest, Luke Stewart, Marcelo D'Agostino, Enrique Perez Gutierrez, Daniel Gillis. 194

XSL•FO
RenderX

Original Paper

Characterizing the COVID-19 Infodemic on Chinese Social Media: Exploratory Study

Shuai Zhang[1], PhD; Wenjing Pian[2], PhD; Feicheng Ma[1], MSc; Zhenni Ni[1], BM; Yunmei Liu[1], PhD

[1]School of Information Management, Wuhan University, Wuhan, China

[2]School of Economics and Management, Fuzhou University, Fuzhou, China

Corresponding Author:
Feicheng Ma, MSc
School of Information Management
Wuhan University
No 299 Bayi Road, Wuchang District
Wuhan,
China
Phone: 86 13507119710
Email: fchma@whu.edu.cn

Abstract

Background: The COVID-19 infodemic has been disseminating rapidly on social media and posing a significant threat to people's health and governance systems.

Objective: This study aimed to investigate and analyze posts related to COVID-19 misinformation on major Chinese social media platforms in order to characterize the COVID-19 infodemic.

Methods: We collected posts related to COVID-19 misinformation published on major Chinese social media platforms from January 20 to May 28, 2020, by using PythonToolkit. We used content analysis to identify the quantity and source of prevalent posts and topic modeling to cluster themes related to the COVID-19 infodemic. Furthermore, we explored the quantity, sources, and theme characteristics of the COVID-19 infodemic over time.

Results: The daily number of social media posts related to the COVID-19 infodemic was positively correlated with the daily number of newly confirmed (r=0.672, P<.01) and newly suspected (r=0.497, P<.01) COVID-19 cases. The COVID-19 infodemic showed a characteristic of gradual progress, which can be divided into 5 stages: incubation, outbreak, stalemate, control, and recovery. The sources of the COVID-19 infodemic can be divided into 5 types: chat platforms (1100/2745, 40.07%), video-sharing platforms (642/2745, 23.39%), news-sharing platforms (607/2745, 22.11%), health care platforms (239/2745, 8.71%), and Q&A platforms (157/2745, 5.72%), which slightly differed at each stage. The themes related to the COVID-19 infodemic were clustered into 8 categories: "conspiracy theories" (648/2745, 23.61%), "government response" (544/2745, 19.82%), "prevention action" (411/2745, 14.97%), "new cases" (365/2745, 13.30%), "transmission routes" (244/2745, 8.89%), "origin and nomenclature" (228/2745, 8.30%), "vaccines and medicines" (154/2745, 5.61%), and "symptoms and detection" (151/2745, 5.50%), which were prominently diverse at different stages. Additionally, the COVID-19 infodemic showed the characteristic of repeated fluctuations.

Conclusions: Our study found that the COVID-19 infodemic on Chinese social media was characterized by gradual progress, videoization, and repeated fluctuations. Furthermore, our findings suggest that the COVID-19 infodemic is paralleled to the propagation of the COVID-19 epidemic. We have tracked the COVID-19 infodemic across Chinese social media, providing critical new insights into the characteristics of the infodemic and pointing out opportunities for preventing and controlling the COVID-19 infodemic.

(JMIR Public Health Surveill 2021;7(2):e26090) doi:10.2196/26090

KEYWORDS

COVID-19; infodemic; infodemiology; epidemic; misinformation; spread characteristics; social media; China; exploratory; dissemination

XSL•FO
RenderX

Introduction

Background

As the COVID-19 pandemic continued to develop, we experienced the parallel rise of the COVID-19 infodemic [1,2]. This infodemic is a phenomenon of overabundance of information caused by COVID-19 misinformation, which has rapidly propagated on social media and attracted widespread attention from the government and health agencies during the ongoing pandemic [3,4]. The infodemic has made the pandemic worse, harmed more people, and jeopardized the global health system's reach and sustainability [5,6]. Thus, the World Health Organization (WHO) has called it a disease accompanying the COVID-19 epidemic [7].

The term "infodemic" is derived from a combination of the root words "information" and "epidemic" and was coined by Eysenbach in 2002 [8], when a SARS outbreak had emerged in the world. It was not until the WHO Director-General reintroduced the term "infodemic" at the Munich Security Conference on February 15, 2020, that it had begun to be used more widely, summarizing the challenges posed by COVID-19 misinformation to our society [9]. In this study, the term "infodemic" refers to an information abundance phenomenon wherein the lack of reliable, trustworthy, and accurate information associated with the COVID-19 epidemic has enabled COVID-19 misinformation to disseminate rapidly across a variety of social media platforms [10]. Thus, the COVID-19 infodemic is also called the COVID-19 misinformation epidemic [11].

Misinformation refers to a claim that is not supported by scientific evidence and expert opinion [12]. This definition explains that misinformation can act as an umbrella concept to explain different types of incorrect information, such as false information, fake news, misleading information, rumors, and anecdotal information, regardless of the degree of facticity and deception [13]. Research linking misinformation to epidemic diseases is emerging [14]. There have been multiple instances where misinformation has been correlated with negative public health outcomes, including the spread of Zika virus [15] and vaccine-preventable infectious diseases in many countries worldwide [16]. Another salient example is the COVID-19 pandemic. For instance, Nsoesie and Oladeji [17] investigated the impact of misinformation on public health during the COVID-19 pandemic. They found that COVID-19 misinformation prevented people from demonstrating effective health behaviors and weakened the public's trust in the health care system. Therefore, dealing with COVID-19 misinformation requires urgent attention.

The increasing global access of social media via mobile phones has led to an exponential increase in the generation of misinformation as well as the number of possible ways to obtain it, thus resulting in an infodemic. Infodemics have co-occurred with epidemics such as Ebola and Zika virus in the past [18,19]. However, the COVID-19 infodemic is significantly different from the earlier ones. It has been reported as "the first true social-media infodemic" [20]. It is also the first infodemic to have been disseminated widely through social media and has significantly impacted public health [21]. By the beginning of 2020, more than 3.8 billion people used social media [22]. Moreover, social media is one of the most popular media for information dissemination and distribution, with 20%-87% usage surging during the crisis [23]. Recently, Oxford's Reuters Institute investigated the dissemination of misinformation and found that a majority (88%) of the misinformation about COVID-19 originated from social media [24]. In Italy, approximately 46,000 posts posted per day on social media in March 2020 were linked to COVID-19 misinformation [25].

In China, the COVID-19 infodemic was more serious [26]. Two-thirds of the Chinese population used social media, and approximately 87% of all users encountered relevant misinformation during the COVID-19 crisis [27]. Examples of such misinformation spread on Chinese social media include that compound Chinese medicine and Banlangen could cure COVID-19; consuming methanol, ethanol, or bleach could protect or cure COVID-19; pneumonia vaccines could protect against SARS-CoV-2; eating garlic could kill the virus; and 5G mobile network has spread COVID-19 [28]. Moreover, China was the first country to experience the COVID-19 infodemic [18]. In December 2019, the first case of COVID-19 was reported in China [29]. In subsequent weeks, the rapid spread of novel coronavirus caused increasing discussion among social media users. Countless unproven stories, advice, and therapies related to COVID-19 were prevalent and skyrocketed on Chinese social media platforms [30].

The COVID-19 infodemic is immensely concerning because all social media users can be affected by it, which poses a severe threat to public health [31]. A study showed that 5800 people were admitted to the hospital as a result of the COVID-19 misinformation disseminated on social media [32]. More seriously, the misinformation that consumption of neat alcohol can cure COVID-19 led to hundreds of deaths due to poisoning [33]. Moreover, the infodemic on social media can also lead to inappropriate actions by users and endanger the government and health agencies' efforts to manage COVID-19, inducing panic and xenophobia [2,34].

Given the negative impact of the COVID-19 infodemic on social media, especially on Chinese social media, the government and health agencies need to assess the COVID-19 infodemic on Chinese social media. Therefore, in this study, we aimed to analyze the quantity, source, and theme characteristics of the COVID-19 infodemic by collecting posts related to COVID-19 misinformation on published on Chinese social media from January 20 to May 28, 2020. Specifically, we used content analysis to analyze the quantity and source of the COVID-19 infodemic. Then, we used topic modeling to analyze various themes of the infodemic. Finally, we explored the quantity, source, and theme characteristics of the COVID-19 infodemic over time.

Prior Works

Previous studies have investigated the distribution and themes of infodemics on social media in other countries. For example, Oyeyemi et al [35] used the Twitter search engine to collect posts about the Ebola virus from September 1 to 7, 2014. They found that 58.9% of the posts were identified as misinformation.

XSL•FO
RenderX

Moreover, the study indicated that misinformation was rampant on social media and had a greater impact on users than did correct information. Similarly, Tran and Lee [36] investigated the propagation of the Ebola infodemic and found that misinformation was more widespread on social media than correct information. Glowacki et al [37] further collected posts about the Zika virus on the live Twitter chat initiated by the Centers for Disease Control and Prevention. They applied topic modeling and derived the following 10 topics relevant to the Zika epidemic: "virology of Zika," "spread," "consequences for infants," "promotion of the chat," "prevention and travel precautions," "education and testing for the virus," "consequences for pregnant women trying to conceive," "insect repellant," "sexual transmission," and "symptoms."

With the world's commitment to the fight against COVID-19, there has been active research in many areas, including social media and quantitative analyses. For example, Kouzy et al [11] assessed the source characteristics of the COVID-19 infodemic being spread on Twitter. They used descriptive statistics to analyze Twitter accounts and post characteristics and found that 66% of misinformation posts regarding the COVID-19 epidemic was posted by unverified individual or group accounts, and 19.2% were posted by verified Twitter users' accounts. Moreover, they indicated that the COVID-19 infodemic is being propagated at an alarming rate on social media. Another study by the COVID-19 Infodemic Observatory found that robots generated approximately 42% of the social media posts related to the pandemic, of which 40% were considered unreliable [38]. Similarly, the Bruno Kessler Foundation analyzed 112 million social media posts about COVID-19 information [26]. The results showed that 40% of this information was from unreliable sources [22]. At the same time, Moon et al [39] collected 200 of the most viewed Korean-language YouTube videos on COVID-19 published from January 1 to April 30, 2020. They found that 37.14% of the videos contained misinformation, and independent videos generated by the user showed the highest proportion of misinformation at 68.09%, whereas all government-generated videos were regarded as useful. Additionally, Naeem et al [23] selected 1225 pieces of misinformation about COVID-19 published in the English language on various social media platforms from January 1 to April 30, 2020, and coded the data using an open coding scheme. They concluded that the theme characteristics of the COVID-19 infodemic include "false claims," "half-backed conspiracy theories," "pseudoscientific therapies," "regarding the diagnosis," "treatment," "prevention," "origin," and "spread of the virus."

Objectives

An increasing number of studies have begun to highlight the COVID-19 infodemic on social media. However, attempts to characterize the spread of the COVID-19 infodemic on social media, especially on Chinese social media platforms, are currently lacking. Hence, in this study, we used content analysis and topic modeling to analyze the COVID-19 infodemic across Chinese social media platforms to gain new insights into the quantity, source, and theme characteristics of the infodemic over time and propose measures to contain the dissemination of misinformation during the COVID-19 infodemic.

Methods

Data Collection

The database for this study was obtained from Qingbo Big Data Agency [40], which covers data from almost all major Chinese social media platforms, such as WeChat, Weibo, and TikTok. The posts collected included microblogs, messages, or short articles shared on these social media platforms. Our search strategy to retrieve post data comprised of the following keywords in Chinese: "coronavirus," "2019-nCoV," "COVID-19," "corona," "new pneumonia," and "new crown." We used PythonToolkit to crawl the data searched using the abovementioned keywords from January 20 to May 28, 2020. The data collection process was as follows. First, we searched the Qingbo Big Data Agency to obtain the results page. Second, the web link crawler was initiated, and the title and URL fields of all web pages were collected. Third, these fields were stored in the url_list dataset of the MongoDB database. Fourth, the web page details crawler was launched, the post published time, source, and text fields of the details page were collected. Finally, these fields were stored in the info_list dataset of the MongoDB database. After data collection was completed, datasets url_list and info_list from the MongoDB database were exported. It should be noted that for video-sharing platforms, the textual description of the video was captured as the post data. Data collection began on January 20, 2020, when the Chinese State Council officially announced the COVID-19 epidemic as a public health emergency [31]. Data collection ended on May 28, 2020, when the National Health Commission of the People's Republic of China issued that the number of new confirmed cases and new suspected cases of COVID-19 in China was zero for the first time. This data collection period could reflect the overall spread of the COVID-19 infodemic on Chinese social media.

All data regarding COVID-19 posts were retrieved, and 723,216 posts were extracted in total. To improve the representativeness of data, we removed incomplete data from the fields and deleted text longer than 400 Chinese characters [41], thus obtaining data from a total of 143,197 posts. Because most of these posts were reposts, we only retained 19,188 of the original post data. We verified the authenticity of post data using the following 2 steps. First, we conduct fact-checking according to the authority organization, such as the National Health Commission of the People's Republic of China, the Chinese Center for Disease Control and Prevention, and the Cyberspace Administration of China. We only retained those posts that were judged to be fake and obtained data from 1729 posts. Next, two independent researchers reviewed and evaluated the remaining posts. One of them is a doctoral student in Library and Information Science, and the other has a bachelor's degree in Medicine. Discrepancies between the 2 researchers were resolved through mutual discussion. The Cohen kappa coefficient was used to analyze the interreviewer reliability for coding. Cohen's Kappa value for the 2 researchers was 0.79, suggesting substantial agreement between them [42]. Ultimately, we obtained 2745 posts related to COVID-19 misinformation as the final analysis sample for this study, which was the largest dataset the study team could obtain with the available resources. The post data was organized

XSL•FO
RenderX

and stored chronologically, and the title, URL, post date, source, and text were recorded. Table 1 details the data format of the posts collected for the analysis.

Table 1. Data format of COVID-19 misinformation posts (partial) on Chinese social media.

Title	URL	Post date	Source	Text
Reposting well-known! One article to understand the new coronavirus…	https://mp.weixin.qq.com/s?src=11&timestamp=1598007513&ver…	2020-01-20	WeChat	…Wuhan virus is the long-standing SARS Coronavirus…
Highly concerned! Wuhan pneumonia continues to spread, 2 cases in Beijing and 1 case in Shenzhen, the public should…	https://mp.weixin.qq.com/s?src=11&timestamp=1598005483&ver…	2020-01-20	WeChat	…WeChat users, who claim to be medical staff, said: "there are several cases in our hospital, which have been strictly isolated. 80% of the cases are said to be SARS case"…
Reposted from Weibo by Cui Tiange, a North American bioinformatics researcher: the new crown virus…	https://m.weibo.cn/status/4463141235003931?sudaref…	2020-01-21	Weibo	…The "mysterious disease" in Wuhan has been confirmed as a new type of SARS virus, or the similarity between Wuhan virus and SARS is as high as 90%…
Weibo_#Academician Zhong Nanshan's team recommends saltwater gargle antivirus#…	https://weibo.com/5044281310/IqH405BUW?type=comment….	2020-01-22	Weibo	…Academician Zhong Nanshan suggests that saltwater gargle prevent new coronavirus…
Six latest facts about Wuhan pneumonia…	https://zhuanlan.zhihu.com/p/103781132…	2020-01-22	Zhihu	…Wuhan virus is a new type of SARS virus. SARS has not disappeared and has been parasitic in bats…
Burst! A patient with "Wuhan pneumonia" fled from Peking Union Medical College Hospital…	http://news.sina.com.cn/c/2020-01-22/doc-iihnzhha4099491…	2020-01-22	Sina	…A patient identified as "Wuhan pneumonia" escaped from Peking Union Medical College Hospital and lost contact…

Data Processing

We used Python (version 3.8.5) and SPSS software (version 25.0; IBM Corp) to perform all data processing and analyses. Time segmentation adapted from the practice of Zhao et al [18] was used to divide the period into 19 time segments (T_1: January 20-26, 2020; T_2: January 27 to February 2, 2020; T_3: February 3-9, 2020; T_4: February 10-16, 2020; T_5: February 17-23, 2020; T_6: February 24 to Mar 1, 2020; T_7: March 2-8, 2020; T_8: March 9-15, 2020; T_9: March 16-22, 2020; T_{10}: March 23-29, 2020; T_{11}: March 30 to April 5, 2020; T_{12}: April 6-12, 2020; T_{13}: April 13-19, 2020; T_{14}: April 20-26, 2020; T_{15}: April 27 to May 3, 2020; T_{16}: May 4-10, 2020; T_{17}: May 11-17, 2020; T_{18}: May 18-24, 2020; and T_{19}: May 25-28, 2020). Among these segments, the last time segment is 4 days long, and the other time segments are 7 days long each, with the total period spanning 130 days.

Based on the classification of social media websites by the CNNIC (China Internet Network Information Center) [43], the sources of posts were categorized into 5 types: chat platforms, video-sharing platforms, news-sharing platforms, health care platforms, and Q&A platforms. The chat platforms included WeChat, Weibo, and QQ. The video-sharing platforms included TikTok, Kuaishou, and Pear Video. The news-sharing platforms included Toutiao, Sina, and Tencent. The health care platforms included DXY.cn, Haodf.com, and Chunyu Yisheng. The Q&A platforms include Zhihu, Douban, and Jianshu (see a full list of Chinese social media types and major social media sites in Multimedia Appendix 1).

The "jieba" package in Python was used to segment post text. We limited the parts of speech of the post text to 9 categories ("n," "nr," "ns," "nt," "eng," "v," "vn," "vs," and "d"). We adapted the method described by Medford et al [29] to merge synonyms into a unified form (eg, "disinfectant powder" and "disinfectant water" into "disinfectants" and "suspense of business" and "termination of business" into "close down"). The Gensim package in Python was used to perform latent Dirichlet allocation (LDA) model. A post contains only one dominant topic. We used different numbers of topics to iteratively train multiple LDA models to maximize the topic coherence score. After more than 10 tests, the results with the highest coherence score in the use of the LDA model with 8 topics were selected. Each topic contains 15 words adhering to convention and is manually tagged with a theme.

Data Analysis

We explored characteristics of the COVID-19 infodemic on Chinese social media from the perspective of quantity, source, and theme. From the perspective of quantity, we counted the daily number of posts and obtained the number of newly confirmed cases and suspected cases each day from the official website of the Chinese Center for Disease Control and Prevention. We performed Pearson correlation analysis to explore the relationship between the daily number of posts with the number of newly confirmed cases and suspected cases per day. Moreover, we calculated the maximum, minimum, upper quartile, lower quartile, and median number of posts in each time segment, and we visualized them to intuitively evaluate the characteristics of post propagation. From the perspective of source and theme, we calculated the sources and themes of posts

XSL•FO
RenderX

based on the number of occurrences. Additionally, we visualized the number of sources of posts in each time segment to analyze the source characteristics of the COVID-19 infodemic. We then created a visualization of the time segment of themes of posts to assess the change in themes over time.

Results

Quantity Characteristics

Figure 1 shows the daily number of posts related to the COVID-19 misinformation on Chinese social media that was published from January 20 to May 28, 2020. The maximum number of posts published in a day was 105, whereas the minimum number was 3 (mean 21.12, SD 17.35). Pearson correlation analysis shows that the daily number of posts related to the COVID-19 infodemic was positively correlated with the daily number of newly confirmed (r=0.672, P<.01) and newly suspected (r=0.497, P<.01) COVID-19 cases in China. In other words, the more posts related to the COVID-19 misinformation that were published per day, the greater was the severity of the COVID-19 epidemic, and vice versa.

Figure 1. Daily number of posts related to COVID-19 misinformation on Chinese social media platforms. Different colored lines indicate the number of posts published.

We used a box plot to describe the spread of social media posts according to different time segments (Figure 2). We found that the posts presented a spread characteristic indicating gradual progress. That is, the number of posts first increases slowly with the time segment, then concentrates on the burst, and then moderates gradually as the time segments continue to advance. Furthermore, the COVID-19 infodemic on Chinese social media can be divided into 5 periods (see Table 2). During the incubation period (Stage A: T_1-T_2), the number of posts showed slow growth, with the mean and median values of approximately 20 per day. Then, the number of posts rapidly increased during the outbreak period (Stage B: T3-T4), and the mean and median values soared to approximately 50 per day. During the stalemate period (Stage C: T_5-T_8), the number of posts remained at a high level, and the mean and median values were approximately 30 per day. During the control period (Stage D: T_9-T_{15}), the number of posts dropped significantly, with mean and median values of approximately 14 per day. Finally, the number of posts has decreased sluggishly in the recovery period (Stage E: T_{16}-T_{19}), and the mean and median values remained at approximately 7 per day.

XSL•FO
RenderX

Figure 2. Box plot of the number of social media posts in each time segment.

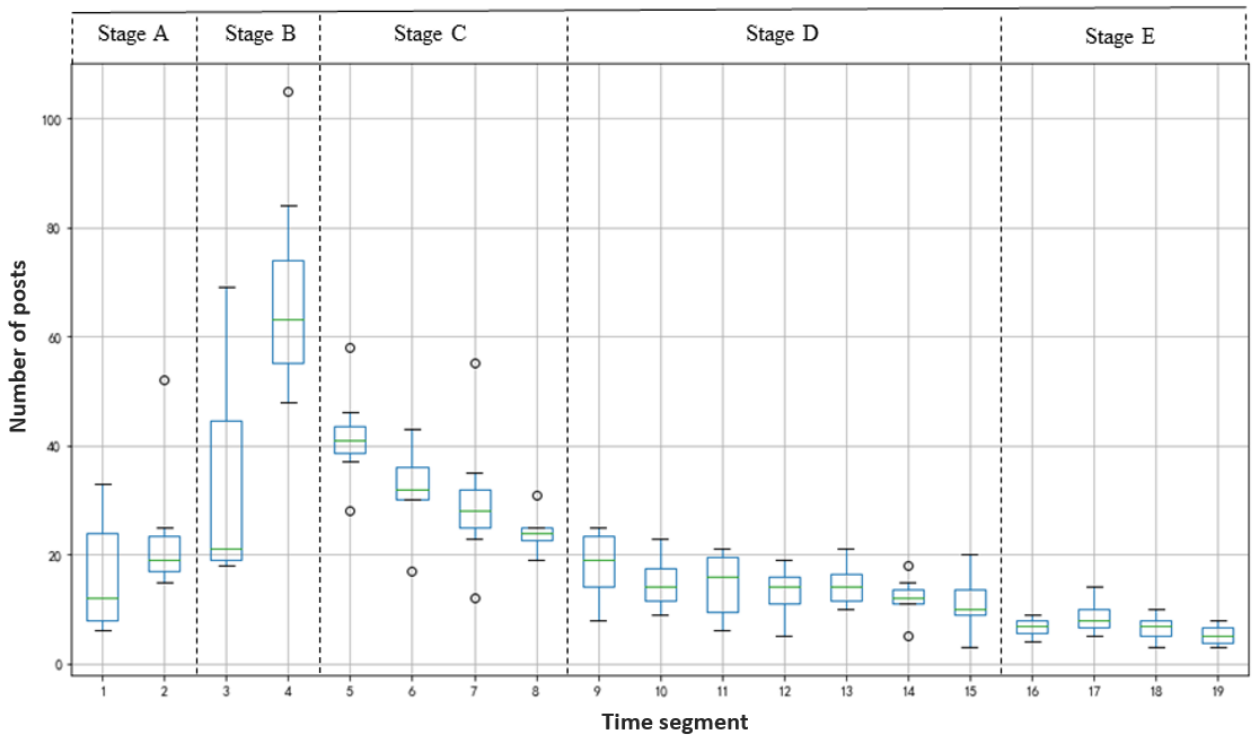


Table 2. Periods of the COVID-19 infodemic based on data from relevant Chinese social media posts.

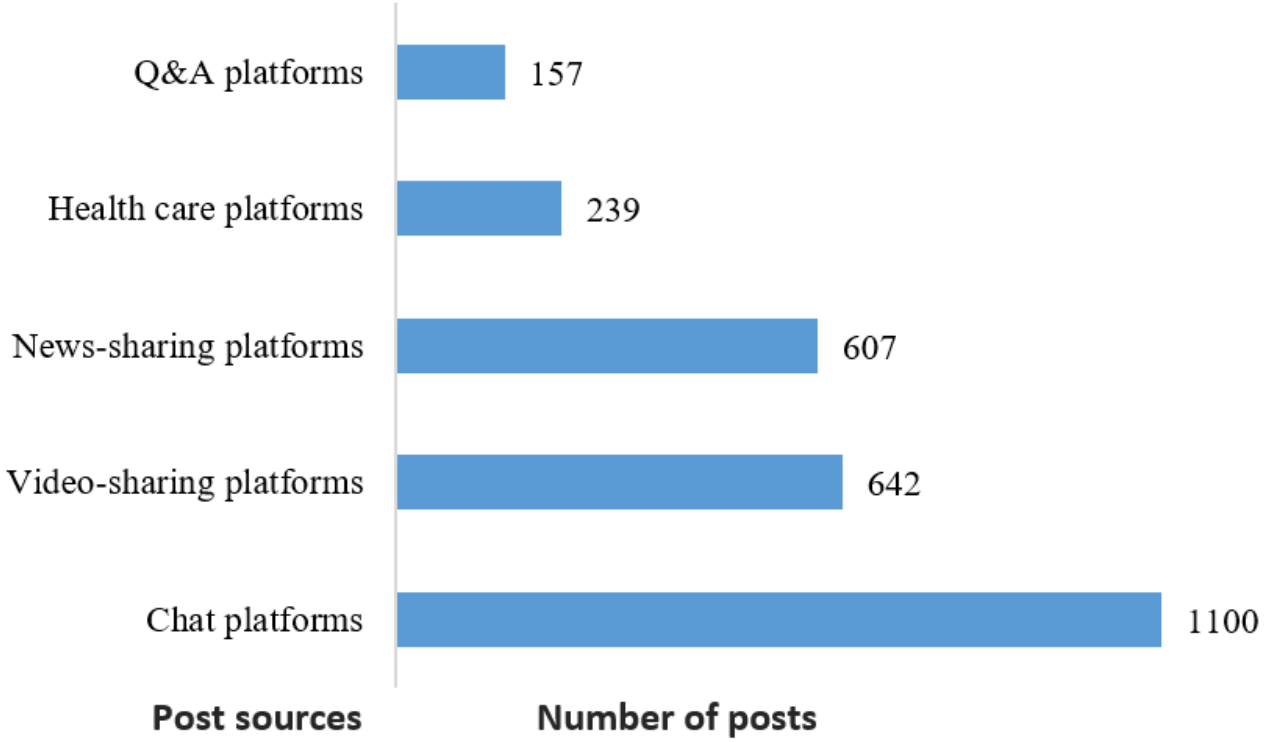
Post metric	Incubation period	Outbreak period	Stalemate period	Control period	Recovery period
Time segment	T_1-T_2	T_3-T_4	T_5-T_8	T_9-T_{15}	T_{16}-T_{19}
Mean (SD) (days)	20.14 (11.72)	50.64 (25.89)	31.89 (10.56)	14.02 (5.18)	6.69 (2.55)
Range	6-52	18-105	12-58	3-25	3-14
Median (IQR) (days)	18 (12.75-24.25)	54 (23-63.75)	30 (24.75-39.25)	14 (10-18)	7 (5-8)

Source Characteristics

Of the posts related to the COVID-19 misinformation that were classified (Figure 3), chat platforms (1100/ 2745, 40.07%) represented the largest source of the COVID - 19 infodemic, followed by video-sharing platforms (642/ 2745, 23.39%) and news-sharing platforms (607/2745, 22.11%). The proportions of health care platforms (239/2745, 8.71%) and Q&A platforms (157/2745, 5.72%) were relatively small.

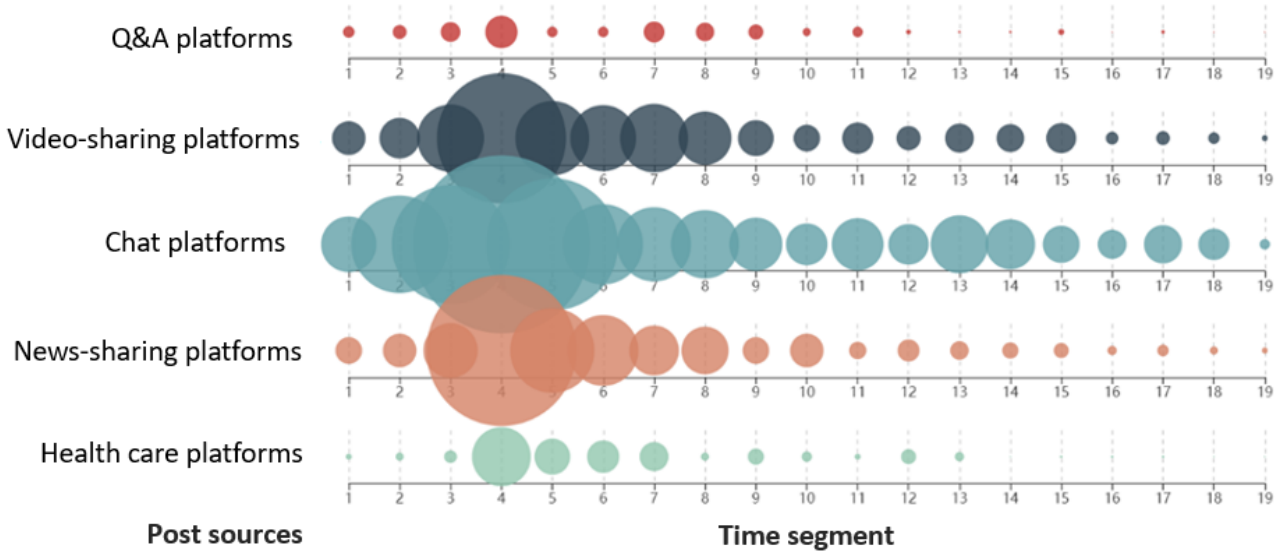
XSL•FO
RenderX

Figure 3. Sources of posts about COVID - 19 misinformation on various Chinese social media platforms.



We visualized the number of sources of posts in each time segment (Figure 4). Chat, video-sharing, and news-sharing platforms were the main sources for the spread of posts during the incubation period (T_1-T_2). Then, the posts began to spread toward the health care and Q&A platforms during the outbreak period (T_3-T_4). Thereafter, the posts were broadly spread on all social media platforms and were maintained at a high level during the stalemate period (T_5-T_8). During the control period (T_9-T_{15}), the spread of the posts on chat and video-sharing platforms alternately increased and decreased, whereas the spread of posts on news-sharing, health care, and Q&A platforms evidently declined. Finally, the spread of posts on chat platforms also gradually decreased during the recovery period (T_{16}-T_{19}), and the spread on other social media platforms dropped sharply and remained at a low level.

Figure 4. Number of sources of social media posts in each time segment. Different colored dots represent different sources, and their sizes represent the proportion of sources.



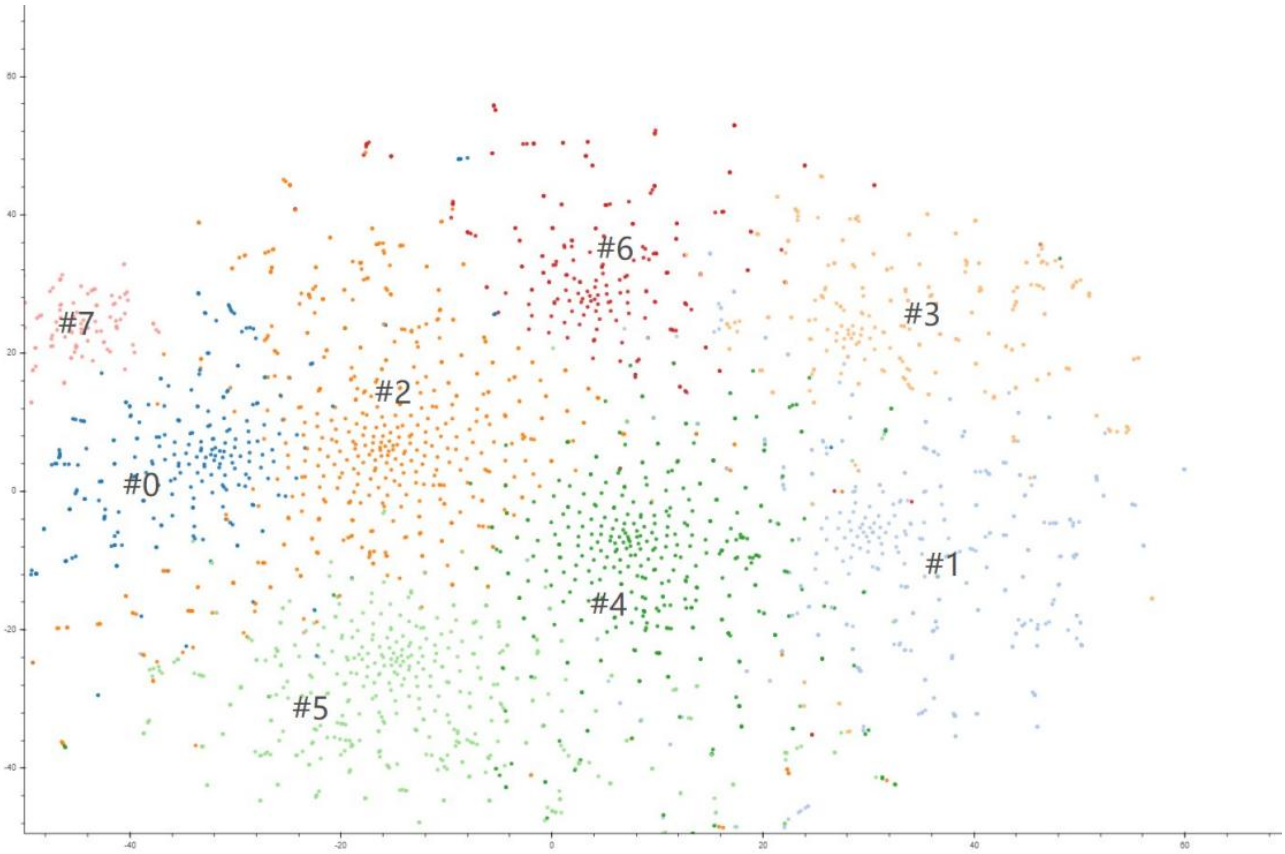
Theme Characteristics

Topic modeling identified 8 different themes, which are illustrated in Figure 5. The 15 keywords that contributed to each theme with their potential theme labels are shown in Table 3. Based on LDA analysis, we obtained a specific theme for each post. The popularity of each theme was determined based on the proportion of posts in each theme considering the overall post data. The most common primary theme was “conspiracy theories” (648/2745, 23.61%), which included topics such as

XSL•FO
RenderX

"Academician Zhong Nanshan did not wear a mask for rounds," "Academician Li Lanjuan helped her son sell medicines," "Dr. Li Wenliang danced before his death," and "Wuhan Huoshenshan was designed by the Japanese." The second most common theme was "government response" (544/2745, 19.82%), which included the following topics: "The city would be closed down at 2:00 PM on January 25, 2020, in Xinyang, Henan province;" "Wuhan gas stations would be closed;" and "Jingzhou, Hubei Province, would suspend issuing permits for leaving Hubei Province." Thereafter, the themes discussed were "prevention action" (411/2745, 14.97%) and "new cases" (365/2745, 13.30%), which included topics such as "Wearing multi-layer masks can prevent the virus," "Smoking vinegar can prevent the virus," "Six promoters of Wuhan Zhongbai Supermarket were confirmed with novel coronavirus pneumonia," and "More than 20,000 new confirmed close contacts in Qingdao." The other common themes included "transmission route" (244/2745, 8.89%) as well as "origin and nomenclature" (228/2745, 8.30%). These themes included the following topics: "Catkins can transmit COVID-19," "COVID-19 is a biological weapon," and "COVID-19 was made by the laboratory." Other themes included "vaccines and medicines" (154/2745, 5.61%) as well as "symptoms and detection" (151/2745, 5.50%), which included topics such as "CT image is used as the latest standard for judging the diagnosis of COVID-19," "Hold your breath for 10 seconds to test whether you are infected with the virus," "The first COVID-19 vaccine was successfully developed and injected," and "Hydroxychloroquine and chloroquine are specific drugs for COVID-19."

Figure 5. Visualization of themes identified by latent Dirichlet allocation.



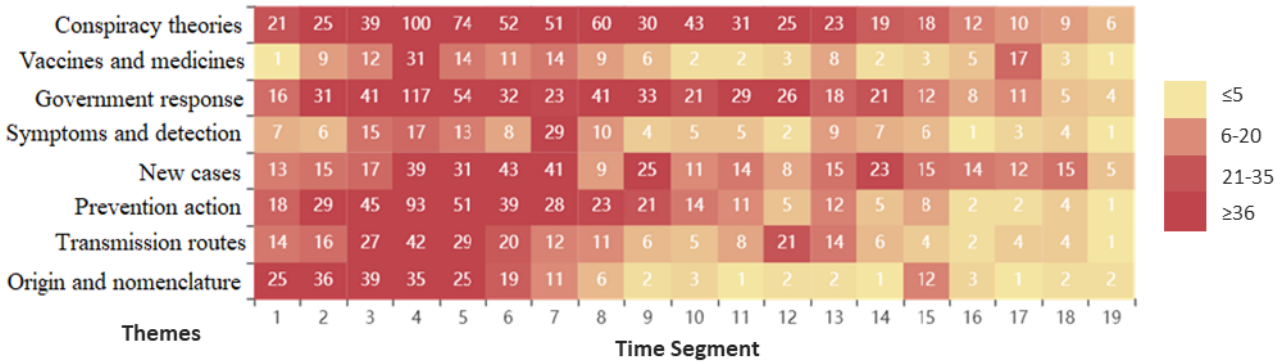
XSL•FO
RenderX

Table 3. Theme labels and keywords contributing to the topic model.

Theme labels	Keywords contributing to topic model
Origin and nomenclature (#0)	COVID-19, SARS, Corona, SARI, host animals, bat, pangolin, variation, pestilence, influenza, the natural world, man-made, biological weapon, laboratory, patient zero
Transmission routes (#3)	5G, seafood, aerosol, catkin, mosquito, paper money, tap water, aquatic product, public toilet, sweater, air conditioner, pet dog, freshwater fish, salmon, subway ticket
Prevention action (#4)	prevention, face mask, disinfectant, alcohol, N95, chlorine, liquor, onion, garlic, vinegar, tea, smoke, strawberries, eyedrops, balm
New cases (#1)	infection, case, confirmed, suspected, patient, isolation, hospital, community, airport, hotel, school, nursing home, student, old people, infant
Symptoms and detection (#7)	detection, test positive, cough, fever, outpatient, computed tomography, lung, blood type, plasma, antibody, diagnostic kit, self-test, suffocation, asymptomatic, expectoration
Government response (#5)	lockdown, road closure, close down, health code, living material, trip, network, transportation, traffic control, home quarantine, traffic permitting, work resumption, school opens, customs office, inbound
Vaccines and medicines (#6)	vaccine, chloroquine, remdesivir, azithromycin, Shuanghuanglian oral liquid, Lianhua Qingwen capsule, Banlangen, oseltamivir, azithromycin, aspirin, Angong Niuhuang Wan, traditional Chinese medicine, Bacillus Calmette-Guerin, toxic strain, Chinese fevervine herb
Conspiracy theories (#2)	Zhong Nanshan, Li Lanjuan, Li Wenliang, Leishenshan, Huoshenshan, Donald John Trump, modular hospital, doctors, nurses, online course, blood donation, suicide, escape, medical corps, cleaner, Red Cross Society

The "Pyecharts" package in Python was used to draw a heat map of themes according to the time segments (Figure 6). We found that different hot themes were discussed at each stage of the COVID-19 infodemic. The theme "origin and nomenclature" was discussed from the start of the incubation period (T_1-T_2). The themes "government response," "new cases," and "transmission routes" were debated on social media during the outbreak period (T_3-T_4). The discussion of "conspiracy theories" and "symptoms and detection" increased significantly in the stalemate period (T_5-T_8). During the control period (T_9-T_{15}), the discussion of "prevention action" was concentrated. Subsequently, the theme "vaccines and medicines" was the focus of discussion on social media during the recovery period (T_{16}-T_{19}).

Figure 6. Heat map of themes related to the COVID-19 infodemic according to time segments. Data within the figure represent the number of posts per theme in each time segment. Individual values in the matrix are represented in different background colors according to the number of posts (range) on a particular theme in that time segment.



We further found that the COVID-19 infodemic presented a spread characteristic of repeated fluctuations across time segments. As shown in Figure 6, each theme is repeated in the time segment, and the theme discussion rate gradually decreases. For example, the theme "government response" not only appeared in the time segment T_2-T_6, but it was also spread in the time segment T_8-T_9, T_{11}-T_{12}, T_{14}, and T_{17}. Moreover, we determined the number of repeated posts for each theme in the time segment (see Table 4) and calculated that the total ratio of repeated posts to be 0.2849 (782/2745), which means that 28.49% of the posts were posted repeatedly in various time segment. This once again verified the spread characteristic of the COVID-19 infodemic that fluctuates repeatedly across time segments. Additionally, the repetition percentage of the themes "conspiracy theories" (198/648, 30.6%), "new cases" (110/365, 30.1%), and "prevention action" (121/411, 29.4%) were particularly high, followed by the themes "government response" (157/544, 28.9%), "origin and nomenclature" (63/228, 27.6%), and "transmission routes" (64/244, 26.2%). The repetition percentage of the themes "vaccines and medicines" (37/154, 24%) and "symptoms and detection" (32/151, 21.2%), however, were relatively low. Differences in repetition among the themes were analyzed by analysis of variance and post hoc analysis, which revealed significant differences in the repetition of themes (F=2.402, *P*=.02). The post hoc tests showed that the theme of "conspiracy theories" was more significant than the

XSL•FO
RenderX

theme "symptoms and detection" (P<.01) and the theme "vaccines and medicines" (P=.04). However, no significant differences were observed between the themes "symptoms and detection" and "vaccines and medicines" (P=.29).

Table 4. Percentage of repeated posts categorized by themes.

Theme categories	Number of posts	Number of repeated posts	Repeated posts (%)
Conspiracy theories	648	198	30.56
Vaccines and medicines	154	37	24.03
Government response	544	157	28.86
Symptoms and detection	151	32	21.19
New cases	365	110	30.14
Prevention action	411	121	29.44
Transmission routes	244	64	26.23
Origin and nomenclature	228	63	27.63
Total	2745	782	28.49

Discussion

Principal Findings

To our knowledge, this study is the first of its kind to analyze posts related to the COVID-19 infodemic on Chinese social media platforms. Previous studies about the COVID-19 infodemic on social media have been mainly qualitative in nature [1,7]. In this study, we analyzed 2745 posts about the COVID-19 infodemic published on Chinese social media platforms between January 20, 2020, and May 28, 2020, which had more than 100 million views cumulatively. We analyzed various characteristics of the COVID-19 infodemic on Chinese social media from the perspective of quantity, source, and theme, to provide decision support for government and health agencies. Below, we discuss 5 key findings of our study that are noteworthy.

First, it was interesting to find that the daily number of posts related to the COVID-19 misinformation on Chinese social media was positively correlated with the daily number of newly confirmed (r=0.672, P<.01) and newly suspected (r=0.497, P<.01) COVID-19 cases in China. This finding indicated that the COVID-19 infodemic paralleled the propagation of the COVID-19 outbreak in China. Our finding is similar to previous studies on posts related to the H7N9 outbreak on Weibo, which showed a positive correlation between the daily number of posts published and the daily number of deaths due to H7N9 infection [44].

Second, we found that the COVID-19 infodemic was characterized by gradual progress, which can be divided into 5 stages. During the incubation period (T_1-T_2), since COVID-19 cases were only reported in Wuhan, the COVID-19 infodemic showed slow growth. Subsequently, the COVID-19 infodemic increased rapidly during the outbreak period (T_3-T_4), as the COVID-19 began to spread across China, causing a mass of public discussion on social media. Thereafter, as the number of COVID-19 cases continued to increase, the COVID-19 infodemic maintained a high level in the stalemate period (T_5-T_8). During the control period (T_9-T_{15}), because of the remarkable decrease in the number of COVID-19 cases, the COVID-19 infodemic also significantly declined. Finally, during the recovery period (T_{16}-T_{19}), the COVID-19 infodemic generally decreased, as the number of COVID-19 cases dropped constantly.

Third, our study found that the COVID-19 infodemic was characterized by videoization. Sources of the COVID-19 infodemic can be divided into 5 types (ie, chat, video-sharing, news-sharing, health care, and Q&A platforms). Among these, video-sharing platforms (23.38%) emerged as the second-largest source after chat platforms. The dissemination mode of "seeing is believing" was subduing public awareness of the COVID-19 epidemic. Moreover, it may be a new spread characteristic for the infodemic. Additionally, we found that the COVID-19 infodemic was more prevalent on chat, video-sharing, and news-sharing platforms than on health care and Q&A platforms. One possible explanation for this difference is that on chat, video-sharing, and news-sharing platforms, users tend to post personal experiences more centrally, which may often be inaccurate, whereas more professional expertise may likely be shared on health care and Q&A platforms.

Fourth, we found that the themes of the COVID-19 infodemic changed with different spread characteristics across stages. Users posted a large number of posts about "origin and nomenclature" in the incubation period (T_1-T_2) and gradually changed to themes such as "government response," "new cases," and "transmission routes" in the outbreak period (T_3-T_4). Subsequently, the themes changed to "conspiracy theories" and "symptoms and detection" in the stalemate period (T_5-T_8), and then progressively concentrated on the themes "prevention action" in the control period (T_9-T_{15}). Finally, in the recover period (T_{16}-T_{19}), the theme changed to "vaccines and medicines." This phenomenon is in line with the characteristic that public opinion online would result in a change in themes in a given period [45,46].

Fifth, our study found that the COVID-19 infodemic showed the characteristic of repeated fluctuations. It indicated that the governance of the COVID-19 infodemic on social media is a "protracted-war." Prior study has also pointed out that the effect of refuting misinformation usually lasts for less than a week

XSL•FO
RenderX

[47,48]. Moreover, we found that the repetition rate of the COVID-19 infodemic themes also differed according to the time segments. The theme "conspiracy theories" was significantly more thrive than the themes "symptoms and detection" and "vaccines and medicines." One possible explanation is that the theme "conspiracy theories" comprised more uncertain knowledge than the themes "symptoms and detection" and "vaccines and medicines." Therefore, users are more inclined to repeat posts of the theme "conspiracy theories."

With regard to the practical implications to curb the COVID-19 infodemic on Chinese social media, our findings suggest that the government and health agencies should manage the infodemic in a stage-wise manner and take more efforts to disseminate accurate and professional information via social media to ameliorate the spread of falsehoods. For instance, expert-approved or peer-reviewed videos are expected to provide credible health information. Furthermore, government and health agencies must pay close attention to the spread of the infodemic on video-sharing platforms. Third, they should coordinate with social media companies to establish long-term systems for the prevention and control of the infodemic. For example, social media platforms can curb the repeated dissemination of COVID-19 misinformation by setting alert labels for repeated misinformation and regularly pushing corrective information to users. Additionally, social media may offer novel opportunities for the government and health agencies to assess and predict the trend of epidemic outbreaks.

Limitations

There are some limitations to this study. First, we targeted posts on Chinese social media; thus, our conclusions may not be applied to social media platforms in other countries, such as Twitter. Second, we collected and analyzed only a relevant subset of all posts about the COVID-19 infodemic, which inevitably introduces some selection bias. Third, as the COVID-19 infodemic continues to disseminate, we should extend the time and expand the data volume to provide the government and health agencies with a more comprehensive prevention and control response. Additionally, our analyses of the repetition of infodemic are still inadequate, and we will further explore this interesting phenomenon in a future study.

Conclusions

Our study found that the COVID-19 infodemic on Chinese social media was characterized by gradual progress, videoization, and repeated fluctuations. Our findings suggest that the COVID-19 infodemic paralleled the propagation of the COVID-19 epidemic. These findings can help the government and health agencies collaborate with major social media companies to develop targeted measures to prevent and control the COVID-19 infodemic on Chinese social media. Moreover, social media offers a novel opportunity for the government and health agencies to surveil epidemic outbreaks.

Acknowledgments

SZ acknowledges financial support from the National Natural Science Foundation of China (No. 71420107026).

Authors' Contributions

SZ and FM conceptualized the study design. SZ and NN collected and analyzed the data. SZ, FM, and WP interpreted the results and wrote the manuscript. SZ, FM, WP, and YL revised the manuscript. All authors have read and approved the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Different types of Chinese social media and major social media platforms.
[PDF File (Adobe PDF File), 163 KB - publichealth_v7i2e26090_app1.pdf]

References

1. Buchanan M. Managing the infodemic. Nat Phys 2020 Sep 03;16(9):894-894. [doi: 10.1038/s41567-020-01039-5]
2. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
3. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced who technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
4. 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. [doi: 10.2196/18444] [Medline: 32250960]
5. Martinez-Juarez LA, Sedas AC, Orcutt M, Bhopal R. Governments and international institutions should urgently attend to the unjust disparities that COVID-19 is exposing and causing. EClinicalMedicine 2020 Jun;23:100376 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100376] [Medline: 32632411]

XSL•FO
RenderX

6. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
7. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
8. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec 18;113(9):763-765 [FREE Full text] [doi: 10.1016/S0002-9343(02)01473-0]
9. Munich Security Conference. World Health Organization. 2020 Feb 15. URL: https://www.who.int/director-general/speeches/detail/munich-security-conference [accessed 2020-09-27]
10. An ad hoc WHO technical consultation managing the COVID-19 infodemic: call for action. World Health Organization. 2020 Apr 7-8. URL: https://apps.who.int/iris/bitstream/handle/10665/334287/9789240010314-eng.pdf [accessed 2020-09-27]
11. 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]
12. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
13. Li Y, Cheung CMK, Shen X, Lee MKO. Health misinformation on social media: a literature review. Health misinformation on social media; 2019 Presented at: 23rd Pacific Asia Conference on Information Systems; July 8-12, 2019; Xi'an, China p. 8-12 URL: https://scholars.cityu.edu.hk/en/publications/health-misinformation-on-social-media(991aef31-8d00-43b9-b8fc-7321e35c2f86).html
14. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
15. Carey JM, Chi V, Flynn DJ, Nyhan B, Zeitzoff T. The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in Brazil. Sci Adv 2020 Jan;6(5):eaaw7449 [FREE Full text] [doi: 10.1126/sciadv.aaw7449] [Medline: 32064329]
16. Larson HJ. The biggest pandemic risk? Viral misinformation. Nature 2018 Oct;562(7727):309-310. [doi: 10.1038/d41586-018-07034-4] [Medline: 30327527]
17. Nsoesie EO, Oladeji O. Identifying patterns to prevent the spread of misinformation during epidemics. BMJ 2020 Apr;349:g6178. [doi: 10.37016/mr-2020-014]
18. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
19. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
20. The coronavirus is the first true social-media “infodemic”. MIT Technology Review. 2020 Feb 12. URL: https://www.technologyreview.com/2020/02/12/844851/the-coronavirus-is-the-first-true-social-media-infodemic/ [accessed 2020-09-30]
21. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
22. Digital 2020: 3.8 billion people use social media. We Are Social. 2020 Jan 30. URL: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [accessed 2020-09-30]
23. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2020 Jul:e [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]
24. Xu Q, Shen Z, Shah N, Cuomo R, Cai M, Brown M, et al. Characterizing Weibo social media posts from wuhan, china during the early stages of the COVID-19 pandemic: qualitative content analysis. JMIR Public Health Surveill 2020 Dec 07;6(4):e24125 [FREE Full text] [doi: 10.2196/24125] [Medline: 33175693]
25. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 7. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2020-10-02]
26. Hollowood E, Mostrous A. Fake news in the time of C-19. 2020 Mar 23. URL: https://members.tortoisemedia.com/2020/03/23/the-infodemic-fake-news-coronavirus/content.html [accessed 2020-10-02]
27. Jiang S. Infodemic: study on the spread of and response to rumors about COVID-19 [in Chinese]. Sudies on Science Popularization 2020 Feb;15(1):70-78 [FREE Full text] [doi: 10.19293/j.cnki.1673-8357.2020.01.011]
28. Coronavirus disease (COVID-19) advice for the public [in Chinese]. Chinese Center for Disease Control and Prevention. URL: http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_2275/index_17.html [accessed 2020-10-02]
29. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "Infodemic": leveraging high-volume Twitter data to understand public sentiment for the COVID-19 outbreak. medRxiv 2020:e [FREE Full text] [doi: 10.1101/2020.04.03.20052936]
30. Ngai CSB, Singh RG, Lu W, Koon AC. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
31. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]

XSL•FO
RenderX

32. 'Hundreds dead' because of Covid-19 misinformation. BBC. London; 2020 Aug 13. URL: http://m.theindependentbd.com/post/251574 [accessed 2020-10-07]
33. Coronavirus: Hundreds dead in Iran from drinking methanol amid fake reports it cures disease. The Independent. 2020 Mar 27. URL: https://www.independent.co.uk/news/world/middle-east/iran-coronavirus-methanol-drink-cure-deaths-fake-a9429956.html [accessed 2020-10-07]
34. Caulfield T. Pseudoscience and COVID-19 - we've had enough already. Nature 2020 Apr 27:e. [doi: 10.1038/d41586-020-01266-z] [Medline: 32341556]
35. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
36. Tran T, Lee K. Understanding citizen reactions and Ebola-related information propagation on social media. New York: IEEE; 2016 Aug Presented at: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 18-21, 2016; San Francisco, CA. [doi: 10.1109/ASONAM.2016.7752221]
37. Glowacki EM, Lazard AJ, Wilcox GB, Mackert M, Bernhardt JM. Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. Am J Infect Control 2016 Dec 01;44(12):1709-1711. [doi: 10.1016/j.ajic.2016.05.025] [Medline: 27544795]
38. Sharma D, Pathak A, Chaurasia RN, Joshi D, Singh RK, Mishra VN. Fighting infodemic: Need for robust health journalism in India. Diabetes Metab Syndr 2020;14(5):1445-1447 [FREE Full text] [doi: 10.1016/j.dsx.2020.07.039] [Medline: 32755849]
39. Moon H, Lee GH. Evaluation of Korean-language COVID-19-related medical information on YouTube: cross-sectional infodemiology study. J Med Internet Res 2020 Aug 12;22(8):e20775 [FREE Full text] [doi: 10.2196/20775] [Medline: 32730221]
40. Qingbo Big Data Agency [in Chinese]. URL: http://www.gsdata.cn/ [accessed 2020-10-01]
41. Zhu B, Zheng X, Liu H, Li J, Wang P. Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos Solitons Fractals 2020 Nov;140:110123 [FREE Full text] [doi: 10.1016/j.chaos.2020.110123] [Medline: 32834635]
42. Olmos M, Antelo M, Vazquez H, Smecuol E, Mauriño E, Bai J. Systematic review and meta-analysis of observational studies on the prevalence of fractures in coeliac disease. Dig Liver Dis 2008 Jan;40(1):46-53. [doi: 10.1016/j.dld.2007.09.006] [Medline: 18006396]
43. The 45th China Statistical Report on Internet Development (Report in Chinese). China Internet Network Information Center. 2020. URL: http://cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/202004/P020200428399188064169.pdf [accessed 2020-10-10]
44. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
45. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
46. 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]
47. Rapp DN. The consequences of reading inaccurate information. Curr Dir Psychol Sci 2016 Aug 10;25(4):281-285. [doi: 10.1177/0963721416649347]
48. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol Sci 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: 10.1177/0956797617714579] [Medline: 28895452]

Abbreviations

LDA: latent Dirichlet allocation
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.11.20; peer-reviewed by X Wang, L Chen; comments to author 06.12.20; revised version received 13.12.20; accepted 15.01.21; published 05.02.21.

Please cite as:
Zhang S, Pian W, Ma F, Ni Z, Liu Y
Characterizing the COVID-19 Infodemic on Chinese Social Media: Exploratory Study
JMIR Public Health Surveill 2021;7(2):e26090
URL: http://publichealth.jmir.org/2021/2/e26090/
doi:10.2196/26090
PMID:33460391

XSL•FO
RenderX

©Shuai Zhang, Wenjing Pian, Feicheng Ma, Zhenni Ni, Yunmei Liu. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 05.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Instagram Posts Related to Backwoods Cigarillo Blunts: Content Analysis

Stephanie Kim[1*], MPH; Alia Mourali[1*], BS; Jon-Patrick Allem[1*], PhD, MA; Jennifer B Unger[1*], PhD; Tess Boley Cruz[1*], PhD, MPH, CHES; Sabrina L Smiley[1*], PhD, MPH, MCHES

Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Sabrina L Smiley, PhD, MPH, MCHES
Department of Preventive Medicine
Keck School of Medicine
University of Southern California
2001 North Soto Street
Los Angeles, CA, 90032
United States
Phone: 1 3234427367
Fax: 1 3234428201
Email: slsmiley@usc.edu

Abstract

Background: Instagram, one of the most popular social media platforms among youth, offers a unique opportunity to examine blunts—partially or fully hollowed-out large cigars, little cigars, and cigarillos that are filled with marijuana. Cigarillo brands like Backwoods (Imperial Tobacco Group Brands LLC) have product features that facilitate blunt making, including a variety of brand-specific flavors that enhance the smoking experience (eg, honey, dark stout). Backwoods has an active online presence with a user-friendly website.

Objective: This study examined the extent to which Backwoods cigarillo–related posts on Instagram showed blunt making. Instagram offers a unique opportunity to examine blunt making as Instagram accounts will contain images reflective of behavior occurring without the prime of a researcher.

Methods: Data consisted of publicly available Instagram posts with the hashtag #backwoods collected from August 30 to September 12, 2018. Inclusion criteria for this study included an Instagram post with the hashtag "#backwoods". Rules were established to content analyze posts. Categories included *Type of post* (ie, photo, video, or both); *Blunt-related hashtags* (ie, the corresponding post caption contained one or more hashtags like #blunts, #cannabis, and #weed that were identified in previous social media research); *Rolling blunts* (ie, the post contained an image of one or more individuals rolling a Backwoods cigarillo visibly containing marijuana); and *Smoking blunts* (ie, the post contained an image of one or more individuals blowing smoke or holding a lit blunt). We coded images for *Product flavor* reference, where a code of 1 showed a Backwoods cigarillo pack with a brand-specific flavor (eg, honey, dark stout, Russian crème) visible in the blunt-related image, and a code of 0 indicated that it was not visible anywhere in the image.

Results: Among all posts (N=1206), 871 (72.2%) were coded as *Blunt-related hashtags*. A total of 125 (10.4%) images were coded as *Smoking blunts*, and 25 (2.1%) were coded as *Rolling blunts* (ie, Backwoods cigarillo explicitly used to roll blunts). Among blunt images, 434 of 836 (51.9%) were coded as *Product flavor* (ie, a Backwoods pack with a brand-specific flavor was visible).

Conclusions: Most Backwoods cigarillo–related Instagram images were blunt-related, and these blunt-related images showed Backwoods packages indicating flavor preference. Continued monitoring and surveillance of blunt-related posts on Instagram is needed to inform policies and interventions that reduce the risk that youth may experiment with blunts. Specific policies could include restrictions on product features (eg, flavors, perforated lines, attractive resealable foil pouches, sale as singles) that facilitate blunt making.

(JMIR Public Health Surveill 2021;7(2):e22946) doi:10.2196/22946

XSL•FO
RenderX

KEYWORDS

Instagram; blunts; Backwoods cigarillos; smoking

Introduction

Blunts are partially or fully hollowed-out cigars, including little cigars or cigarillos, that are refilled with marijuana. Blunts are an increasingly popular way to smoke marijuana [1,2] and are associated with escalation in use of nicotine and marijuana among adolescents and adults [3-5]. Among US adolescents who have ever used a cigarillo or little cigar, 40% used them to make blunts [6]. Blunt users are exposed to nicotine through the tobacco wrap [7,8], and they are exposed to greater carbon monoxide compared to non–blunt users who smoke marijuana wrapped in a cigarette rolling paper that does not contain nicotine [9,10].

Cigarillos contain just as much nicotine and carcinogens as cigarette smoke, if not more [11,12], leading to increased health risks [13]. Youth and young adults in the United States have among the highest prevalence of cigarillo use [14]. Cigarillos also face fewer federal restrictions (eg, allowed in flavors, sold as singles and in packs of two) than cigarettes [14], but are similar to cigarettes in size, shape, and combustible use [12,14]. Cigarillos are widely available [15-18], and popular brands like Swisher Sweets (Swisher International, Inc) and Backwoods (Imperial Tobacco Group Brands LLC) have features that facilitate blunt making, including perforated lines or tobacco wrappers that are easy to unroll and fill with marijuana, smell-proof resealable foil pouches to conceal marijuana, and availability in a variety of flavors that enhance the smoking experience (eg, honey bourbon, sweet aromatic). Backwoods, in particular, uses advertisement claims that are misleading, like “always true” [19].

Prior research has investigated Backwoods-related posts on Instagram and found marijuana was a common theme [20]. However, it was not determined whether Backwoods cigarillos were being used to complement marijuana use (ie, dual use) or explicitly used to roll blunts. As such, this study is an initial step to examine the extent to which Backwoods cigarillo–related posts on Instagram showed blunt making. Instagram is an image-based platform that has been used to study health-related attitudes and behaviors as well as promotional material from companies [20,21]. Instagram offers a unique opportunity to examine blunt making, as Instagram accounts will contain images reflective of behavior occurring without the prime of a researcher [20,21]. Instagram also offers multicontextual content (images and text) that has provided useful insights about user experiences with tobacco products [20,21]. This is important because blunt use has adverse health effects, and these product features might be expanding the population of tobacco users from marijuana users who otherwise would not use any tobacco.

Methods

Data consisted of publicly available Instagram posts with the hashtag #backwoods collected from August 30 to September 12, 2018. Netlytic, an Instagram-approved vendor that accessed the public application programming interface of Instagram, was used to collect data. A total of 12,306 posts included the hashtag #backwoods during the study period. Similar to prior Instagram studies [20,21], we numbered each observation and then randomly drew observations using a random number generator until 10% of the sample was culled from the initial corpus. Multiple posts from the same users in either the overall sample frame or in the randomly selected posts were not treated as independent observations. Backwoods (Imperial Tobacco Group Brands LLC) was not the source of any of the posts. Rules were established to content analyze 1206 posts.

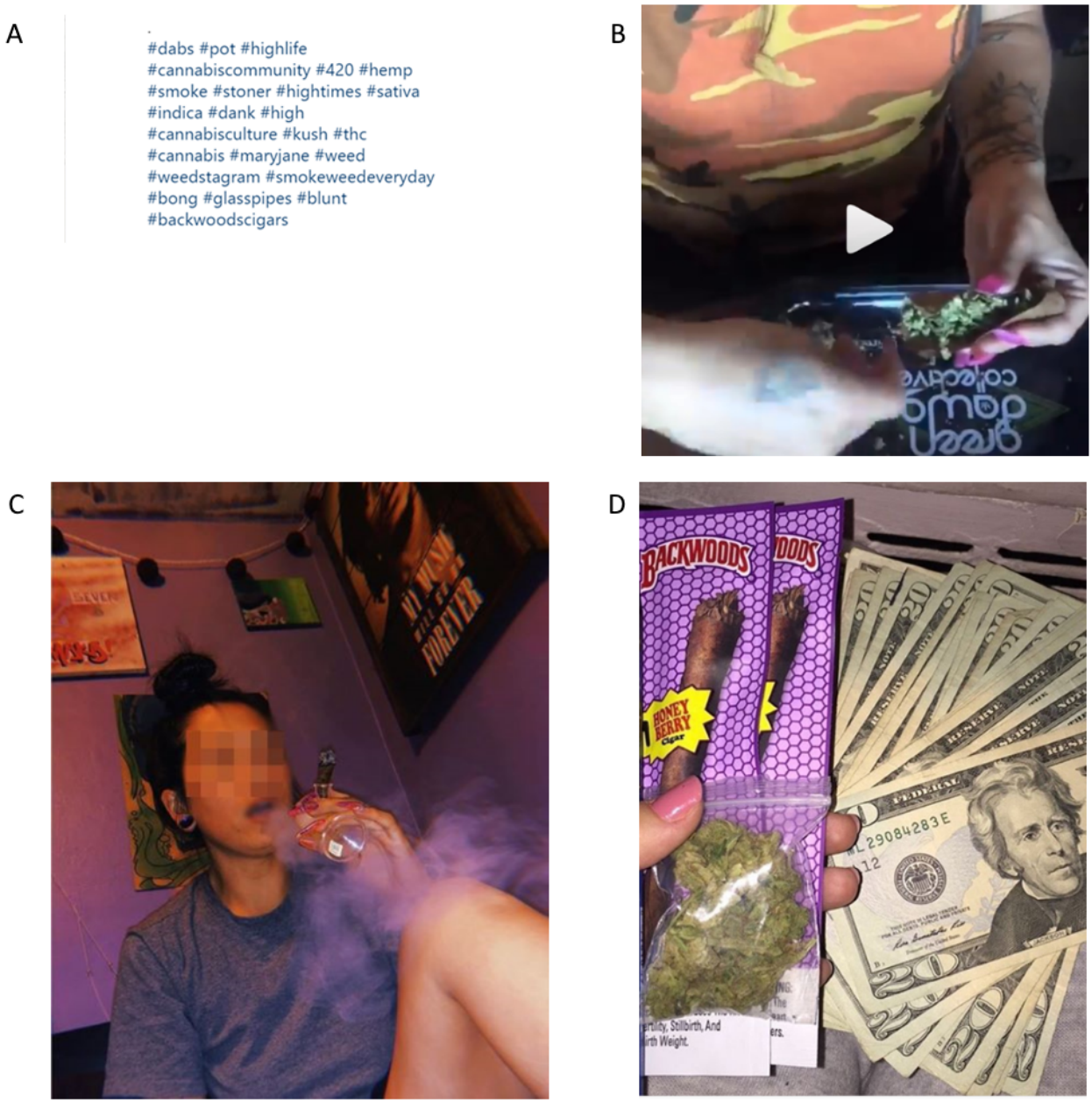
The first and second authors generated a codebook based on prior research [20-22] and reviewed a subsample (N=200) of the posts to identify prominent themes. The unit of analysis was the individual Instagram post (ie, the image and corresponding caption), and the coding strategy assessed themes found in the posts. The coding strategy included (1) *Type of post* (ie, photo, video, or both); (2) *Blunt-related hashtags* (ie, the corresponding post caption contained one or more hashtags like #blunts, #cannabis, and #weed that were identified in previous social media research [23,24]); (3) *Rolling blunts* (ie, the post contained an image of one or more individuals rolling a Backwoods cigar visibly containing marijuana); and (4) *Smoking blunts* (ie, the post contained an image of one or more individuals blowing smoke or holding a lit blunt). Similar to previous research using Instagram data [21], we coded images for (5) *Product flavor* reference, where a code of 1 showed a Backwoods package with a brand-specific flavor (eg, honey, dark stout, Russian crème) visible in the image (eg, next to a Backwoods cigarillo that contained marijuana), and a code of 0 indicated that a Backwoods pack with a brand-specific flavor was not visible anywhere in the image. Two investigators (SK and AM) independently coded all posts, and percentage agreement was substantial at 97.0% (*Type of post*; 1170/1206), 90.0% (*Blunt-related hashtags*; 784/871), 100% (*Rolling blunts*; 25/25), 99.2% (*Smoking blunts*; 124/125), and 100% (*Product flavor*; 434/434). Discrepancies were resolved via in-person discussion. We report the percentages of posts for each theme.

Results

Among all posts (N=1206), 913 (75.7%) were photos, 268 (22.2%) were videos, and 25 (2.1%) included both photos and videos. A total of 774 (64.2%) were posts from individual Instagram users, and 432 (35.8%) were posts from online tobacco retailers. A total of 871 (72.2%) were coded as *Blunt-related hashtags* (Figure 1A). A total of 125 (10.4%) images were coded as *Smoking blunts* (Figure 1C), and 25 (2.1%) were coded as *Rolling blunts* (ie, Backwoods cigarillo explicitly used to roll blunts; Figure 1B). Among blunt images, 434 of 836 (51.9%) were coded as *Product flavor* (ie, a Backwoods pack with a brand-specific flavor, such as honey, dark stout, or Russian crème, was visible; Figure 1D).

XSL•FO
RenderX

Figure 1. Images representative of select themes: (A) Blunt-related hashtags; (B) Rolling blunts; (C) Smoking blunts; (D) Product flavor.



Discussion

Findings from this study suggest that blunt users perceive Backwoods cigarillos as vehicles for marijuana consumption and that they are comfortable sharing images of this behavior on Instagram. This may be the start of a growing trend as Imperial Brands, the parent company of Imperial Tobacco Group Brands LLC (America's third-largest tobacco company [25]), recently signed a US $123 million investment deal with Auxly Cannabis Group Inc of Vancouver, Canada [26]. It appears that marijuana may play a substantial role in the tobacco market of the future, especially given that Imperial's most popular machine-made cigar brand is Backwoods [27].

When considering Backwoods online presence, it should be noted that the age verification on the brand's website currently reads, "To roll with us, you've got to be 21 or over" [19]. This is concerning because Backwoods is clearly aligning itself with blunt making with a website greeting that suggests consumers should consider their tobacco products as good choices for rolling blunts. Findings complement previous research [5,18,19] that cigarillo product features (eg, Backwoods brand name, "natural" leaf wrappers, flavors) facilitate blunt making, including on Instagram.

Findings suggest that images on Instagram capture the social context in which individuals increasingly display blunt making. Instagram is one of the most popular social media platforms among youth; thus, they could view these public blunt-making posts. Research is needed to examine how youth might be turning to Instagram to learn about blunt making. This includes assessing whether posts were somehow more targeted to youth. Research is also needed to compare the reactions and responses of followers to the posts from individual Instagram users and online tobacco retailers across categories (ie, *Rolling blunts*,

XSL•FO
RenderX

Smoking blunts). Additionally, blunt making facilitation and use on Instagram should be considered when designing smoking prevention programs for youth.

Findings from this study should be considered with several limitations in mind, including the sole focus on the cigarillo brand Backwoods and related images on Instagram. Findings may not generalize to other companies or social media platforms (eg, Twitter, Facebook, Tumblr). The images analyzed in this study were collected from a 2-week time period and may not generalize to other time periods. Future research should examine longer timeframes, different social media platforms, and additional brands to fully characterize the blunt making and use experience.

This study demonstrated that more than half of Backwoods cigarillo–related Instagram images were blunt-related, and over half of these blunt-related images showed Backwoods packages indicating flavor preference. This study also found that consumers of Backwoods cigarillos were using Instagram to promote smoking blunts and the blunt-making process (eg, rolling). Findings inform the Food and Drug Administration's regulation of cigar products that are covered under the 2016 Final Deeming Rule [28]. Specific regulations could include restrictions on product features (eg, flavors, perforated lines, attractive resealable foil pouches, sale as singles) that facilitate blunt making.

The University of Southern California Institutional Review Board approved all study procedures.

Acknowledgments

The authors would like to thank Christopher Rogers for his support to data management.

Research reported in this publication was supported by grant number U54CA180905 from the National Cancer Institute (NIH-NCI) and the Food and Drug Administration Center for Tobacco Products (FDA-CTP). At the time of the study, SLS was supported through an award from the National Cancer Institute Research Supplement to Promote Diversity in Health-Related Research. The NIH-NCI or FDA-CTP had no role in study design, data collection, data analysis, interpretation of data, writing the report, or the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH-NCI or FDA-CTP.

Authors' Contributions

SK and SLS conceived of the study. JBU and TBC received funding for the study. SK and AM coded the images and analyzed the data. J-PA collected the data. SK and SLS drafted the initial manuscript. J-PA, JBU, and TBC revised the manuscript for important intellectual content. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Antognoli E, Cavallo D, Trapl E, Step M, Koopman Gonzalez S, Perez R, et al. Understanding nicotine dependence and addiction among young adults who smoke cigarillos: a qualitative study. Nicotine Tob Res 2018 Feb 07;20(3):377-382. [doi: 10.1093/ntr/ntx008] [Medline: 28184877]
2. Giovenco DP, Miller Lo EJ, Lewis MJ, Delnevo CD. "They're Pretty Much Made for Blunts": Product Features That Facilitate Marijuana Use Among Young Adult Cigarillo Users in the United States. Nicotine Tob Res 2017 Nov 01;19(11):1359-1364 [FREE Full text] [doi: 10.1093/ntr/ntw182] [Medline: 27613943]
3. Timberlake DS. A comparison of drug use and dependence between blunt smokers and other cannabis users. Subst Use Misuse 2009;44(3):401-415. [doi: 10.1080/10826080802347651] [Medline: 19212929]
4. Badiani A, Boden JM, De Pirro S, Fergusson DM, Horwood LJ, Harold GT. Tobacco smoking and cannabis use in a longitudinal birth cohort: evidence of reciprocal causal relationships. Drug Alcohol Depend 2015 May 01;150:69-76. [doi: 10.1016/j.drugalcdep.2015.02.015] [Medline: 25759089]
5. Fairman BJ. Cannabis problem experiences among users of the tobacco-cannabis combination known as blunts. Drug Alcohol Depend 2015 May 01;150:77-84 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.02.014] [Medline: 25746234]
6. Kong G, Bold KW, Simon P, Camenga DR, Cavallo DA, Krishnan-Sarin S. Reasons for Cigarillo Initiation and Cigarillo Manipulation Methods among Adolescents. Tob Regul Sci 2017 Apr;3(2 Suppl 1):S48-S58 [FREE Full text] [doi: 10.18001/TRS.3.2(Suppl1).6] [Medline: 29085867]
7. Peters EN, Schauer GL, Rosenberry ZR, Pickworth WB. Does marijuana "blunt" smoking contribute to nicotine exposure?: Preliminary product testing of nicotine content in wrappers of cigars commonly used for blunt smoking. Drug Alcohol Depend 2016 Nov 01;168:119-122. [doi: 10.1016/j.drugalcdep.2016.09.007] [Medline: 27639129]
8. Djordjevic MV, Doran KA. Nicotine content and delivery across tobacco products. Handb Exp Pharmacol 2009(192):61-82. [doi: 10.1007/978-3-540-69248-5_3] [Medline: 19184646]

XSL•FO
RenderX

9. Cooper ZD, Haney M. Comparison of subjective, pharmacokinetic, and physiological effects of marijuana smoked as joints and blunts. Drug Alcohol Depend 2009 Aug 01;103(3):107-113 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.01.023] [Medline: 19443132]
10. Meier E, Hatsukami DK. A review of the additive health risk of cannabis and tobacco co-use. Drug Alcohol Depend 2016 Sep 01;166:6-12. [doi: 10.1016/j.drugalcdep.2016.07.013] [Medline: 27476751]
11. National Cancer Institute. Cigars: Health Effects and Trends Tobacco Control Monograph 9. Washington, DC: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Institutes of Health, National Cancer Institute; 1998.
12. Goel R, Trushin N, Reilly SM, Bitzer Z, Muscat J, Foulds J, et al. A Survey of Nicotine Yields in Small Cigar Smoke: Influence of Cigar Design and Smoking Regimens. Nicotine Tob Res 2018 Sep 04;20(10):1250-1257. [doi: 10.1093/ntr/ntx220] [Medline: 29059441]
13. Nonnemaker J, Rostron B, Hall P, MacMonegle A, Apelberg B. Mortality and Economic Costs From Regular Cigar Use in the United States, 2010. Am J Public Health 2014 Sep;104(9):e86-e91. [doi: 10.2105/ajph.2014.301991] [Medline: 25033140]
14. Delnevo CD, Hrywna M, Giovenco DP, Miller Lo EJ, O'Connor RJ. Close, but no cigar: certain cigars are pseudo-cigarettes designed to evade regulation. Tob Control 2016 May 24;26(3):349-354. [doi: 10.1136/tobaccocontrol-2016-052935] [Medline: 27220622]
15. Delnevo CD, Giovenco DP, Miller Lo EJ. Changes in the Mass-merchandise Cigar Market since the Tobacco Control Act. Tob Regul Sci 2017 Apr 01;3(2):S8-S16. [doi: 10.18001/trs.3.2(suppl1).2] [Medline: 28317004]
16. Kong AY, Queen TL, Golden SD, Ribisl KM. Neighborhood Disparities in the Availability, Advertising, Promotion, and Youth Appeal of Little Cigars and Cigarillos, United States. Nicotine Tob Res. 2020 2020 Dec 12;22(12):2170-2177. [doi: 10.1093/ntr/ntaa005] [Medline: 31917833]
17. Smiley SL, Soto C, Cruz TB, Kintz N, Rodriguez YL, Sussman S, et al. Point-of-sale marketing of little cigars and cigarillos on and near California Tribal lands. Tob Control 2020 Jan;29(1):122-124 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054558] [Medline: 30385648]
18. Smiley SL, Kintz N, Rodriguez YL, Barahona R, Sussman S, Cruz TB, et al. Disparities in retail marketing for little cigars and cigarillos in Los Angeles, California. Addict Behav Rep 2019 Jun;9:100149 [FREE Full text] [doi: 10.1016/j.abrep.2018.100149] [Medline: 31193771]
19. Backwoods. Age verification. URL: https://backwoodscigars.com/age-verification/ [accessed 2020-06-19]
20. Smiley SL, Kim S, Mourali A, Allem JP, Unger JB, Boley Cruz T. Characterizing #Backwoods on Instagram: "The Number One Selling All Natural Cigar". Int J Environ Res Public Health 2020 Jun 25;17(12) [FREE Full text] [doi: 10.3390/ijerph17124584] [Medline: 32630567]
21. Allem JP, Escobedo P, Chu K, Boley Cruz T, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Jul 14;19(7):e255. [doi: 10.2196/jmir.7634] [Medline: 28710057]
22. Montgomery L, Yockey A. Rolling and scrolling: The portrayal of marijuana cigars (blunts) on YouTube. J Subst Use 2018;23(4):436-440 [FREE Full text] [doi: 10.1080/14659891.2018.1444108] [Medline: 29887756]
23. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-Related Posts on Instagram. Prev Sci 2016 Aug;17(6):710-720. [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
24. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
25. Imperial Tobacco Group Brands. ITG Brands Homepage. URL: https://www.itgbrands.com [accessed 2020-11-27]
26. The CP. Auxly Cannabis signs $123M deal with tobacco giant Imperial Brands. CBC. URL: https://www.cbc.ca/news/business/auxly-cannabis-deal-with-tobacco-giant-imperial-brands-1.5224955 [accessed 2020-06-17]
27. Cavale S, Machado S. Imperial's premium cigar era draws to $1.33 billion end after slow burn sale.: Reuters; 2020 Apr 27. URL: https://www.reuters.com/article/imperial-brands-divestiture-idINKCN229114?edition-redirect=in [accessed 2020-11-26]
28. U.S. Food and Drug Administration. Deeming Tobacco products to be subject to the Federal Food, Drug, and Cosmetic Act, as amended by the Family Smoking Prevention and Tobacco Control Act; restrictions on the sale and distribution of tobacco products and required warning statements for tobacco products. 2016. URL: tinyurl.com/1ipo6e32 [accessed 2020-12-30]

XSL•FO
RenderX

Edited by T Sanchez; submitted 10.08.20; peer-reviewed by L Montgomery, R Sun, N Nguyen, S Park; comments to author 24.09.20; revised version received 27.11.20; accepted 04.01.21; published 09.02.21.

Please cite as:
Kim S, Mourali A, Allem JP, Unger JB, Boley Cruz T, Smiley SL
Instagram Posts Related to Backwoods Cigarillo Blunts: Content Analysis
JMIR Public Health Surveill 2021;7(2):e22946
URL: http://publichealth.jmir.org/2021/2/e22946/
doi:10.2196/22946
PMID:33560242

©Stephanie Kim, Alia Mourali, Jon-Patrick Allem, Jennifer B Unger, Tess Boley Cruz, Sabrina L Smiley. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 09.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Influence of Media Coverage and Governmental Policies on Google Queries Related to COVID-19 Cutaneous Symptoms: Infodemiology Study

Solene Huynh Dagher[1], MSc, MD; Guillaume Lamé[2], PhD; Thomas Hubiche[3], MD; Khaled Ezzedine[1,4], MD, PhD, Prof Dr; Tu Anh Duong[1,5], MD, PhD

[1]Assistance Publique des Hôpitaux de Paris (AP-HP), Département de dermatologie, Hôpital Henri Mondor, Créteil, France

[2]Laboratoire Génie Industriel, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

[3]Département de dermatologie, Centre hospitalier universitaire de Nice, Nice, France

[4]EA 7379, EpidermE, Université Paris-Est Créteil, Créteil, France

[5]Chaire Avenir Santé numérique, Équipe 8 IMRB U955, INSERM, Université Paris-Est Créteil, Créteil, France

Corresponding Author:
Tu Anh Duong, MD, PhD
Assistance Publique des Hôpitaux de Paris (AP-HP)
Département de dermatologie
Hôpital Henri Mondor
51 avenue du Maréchal de Lattre de Tassigny
Créteil, 94000
France
Phone: 33 149812512
Email: tu-anh.duong@aphp.fr

Abstract

Background: During COVID-19, studies have reported the appearance of internet searches for disease symptoms before their validation by the World Health Organization. This suggested that monitoring of these searches with tools including Google Trends may help monitor the pandemic itself. In Europe and North America, dermatologists reported an unexpected outbreak of cutaneous acral lesions (eg, chilblain-like lesions) in April 2020. However, external factors such as public communications may also hinder the use of Google Trends as an infodemiology tool.

Objective: The study aimed to assess the impact of media announcements and lockdown enforcement on internet searches related to cutaneous acral lesions during the COVID-19 outbreak in 2020.

Methods: Two searches on Google Trends, including daily relative search volumes for (1) “toe” or “chilblains” and (2) “coronavirus,” were performed from January 1 to May 16, 2020, with the United States, the United Kingdom, France, Italy, Spain, and Germany as the countries of choice. The ratio of interest over time in “chilblains” and “coronavirus” was plotted. To assess the impact of lockdown enforcement and media coverage on these internet searches, we performed an interrupted time-series analysis for each country.

Results: The ratio of interest over time in “chilblains” to “coronavirus” showed a constant upward trend. In France, Italy, and the United Kingdom, lockdown enforcement was associated with a significant slope change for “chilblain” searches with a variation coefficient of 1.06 (SE 0.42) (P=0.01), 1.04 (SE 0.28) (P<.01), and 1.21 (SE 0.44) (P=0.01), respectively. After media announcements, these ratios significantly increased in France, Spain, Italy, and the United States with variation coefficients of 18.95 (SE 5.77) (P=.001), 31.31 (SE 6.31) (P<.001), 14.57 (SE 6.33) (P=.02), and 11.24 (SE 4.93) (P=.02), respectively, followed by a significant downward trend in France (–1.82 [SE 0.45]), Spain (–1.10 [SE 0.38]), and Italy (–0.93 [SE 0.33]) (P<.001, P=0.004, and P<.001, respectively). The adjusted R^2 values were 0.311, 0.351, 0.325, and 0.305 for France, Spain, Italy, and the United States, respectively, suggesting an average correlation between time and the search volume; however, this correlation was weak for Germany and the United Kingdom.

Conclusions: To date, the association between chilblain-like lesions and COVID-19 remains controversial; however, our results indicate that Google queries of “chilblain” were highly influenced by media coverage and government policies, indicating that caution should be exercised when using Google Trends as a monitoring tool for emerging diseases.

XSL•FO
RenderX

(JMIR Public Health Surveill 2021;7(2):e25651) doi:10.2196/25651

KEYWORDS

chilblains; COVID-19; dermatology; Google Trends; infodemiology; lesion; media; media coverage; online health information; skin lesions; trend

Introduction

The use of social media and search engines for information on diseases and their diagnosis increased considerably in the last decade [1-3]. Infodemiology can be defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [4-6]. Queries on search engines including Google and tracking of specific health research terms could constitute an early warning signal for the emergence of diseases, as already shown during the COVID-19 pandemic [7,8].

The first-described COVID-19 symptoms were fever, fatigue, dyspnea, dry cough, and anorexia [9]. Later, physicians described other symptoms associated with the disease, such as anosmia or acral cutaneous lesions [10]. A group of French dermatologists reported an outbreak of these lesions, describing them as erythemas with vesicles or pustules located on extremities (eg, chilblain-like lesions) in paucisymptomatic patients, in a WhatsApp group [11]; subsequently, several studies reported this uncommon cutaneous manifestation without confirming its association with COVID-19 [12,13]. Chilblain-like lesions were the most frequently reported cutaneous findings during COVID-19.

Studies revealed a surge in individual Google searches related to chilblains in early 2020 during COVID-19 [14-16]. Kluger et al [14] and Hughes et al [15] reported an increase in searches during early March in France, the United States, and worldwide, suggesting a potential link between searches on chilblain-like lesions and an increased frequency of such lesions in the context of COVID-19. However, these studies do not consider alternative explanations for surges in these Google searches, concurrent with the curiosity generated by public announcements on new symptoms. Previous studies have highlighted the impact of these events on Google searches during COVID-19 [17]. As opposed to anosmia and other symptoms [18], chilblain-like lesions appear late during the infection, implying that internet queries are more likely to be affected by early media announcements.

To assess the possibility of monitoring the dynamics of COVID-19 by analyzing internet searches, we first need to understand whether the increase in Google searches for chilblains reflect actual symptoms or mere curiosity after public announcements. In particular, in the context of an emerging infectious disease, it is important to investigate the impact of external factors, such as media coverage and strong government policies, on Google queries. Using Google Trends (GT), we sought to study the influence of public announcements related to chilblain-like lesions and government decisions (eg, lockdown enforcement) on searches related to chilblains during COVID-19.

Methods

Study Design

We used GT to measure internet searches on chilblain-like symptoms during COVID-19 and used these data to assess the impact of media announcements and government policy decisions on internet search behaviors in accordance with existing guidelines [19].

Search Strategy

To select keywords, we selected commonly used search terms related to chilblain-like lesions, such as "toe" or "frostbite," in order to retrieve queries made by the general population. We did not include "finger," since excessive use of sanitizers during the pandemic resulted in irritation and contact dermatitis. Furthermore, we measured searches of medical terms such as "chilblains" using both the singular and plural forms of the search term. We did not take misspellings into account. Since "toe" can be used in different contexts, we restricted our searches to the "health" category. We used quotation marks for queries including strings of >2 words (eg, "dedo del pie" in Spanish). Search strings included a combination of these terms with the "+" operator denoting the "OR" logical function (Table 1). "Coronavirus" and "COVID-19" were searched in a separate query, also restricted to the "health" category.

XSL•FO
RenderX

Table 1. Search inputs, date of lockdown enforcement, and date of the initial press release on acral lesions in France, Spain, Italy, Germany, the United Kingdom, and the United States.

Data points	France	Spain	Italy	Germany	United Kingdom	United States
Search input	Orteil + orteils + engelure + engelures	sabañones + "dedo del pie" + "dedos de les pies" + erupciones + pies + moretones	gelone + piedi + "dita dei piedi" + geloni	zehen + Hautausschlag + flecken + frostbeulen + erfrierungen	toe + toes + frost-bite + chilblain + chilblains + "chill burn"	toe + toes + frost-bite + chilblain + chilblains + "chill burn"
Lockdown enforce-ment date	March 17	March 13	March 9	March 22	March 23	March 21[a]
Press release date	April 6	April 9	April 9	April 20	April 14	April 14

[a]In the United States, we chose March 21 as the date of lockdown enforcement, since this date differed among states: California, March 19; Illinois, New Jersey, and New York, March 22; Ohio, March 23.

We included high-income countries whose populations are familiar with search engines and where Google is the leading search engine. Furthermore, we included countries with different government responses to the pandemic and different announcements related to chilblain-like lesions, including the United States, the United Kingdom, Spain, Italy, France, and Germany. We translated the search terms in the five languages (English, French, Spanish, Italian, and German) using both internet-based translators and support from native speakers.

Identification of External Factors

First, we hypothesized a change in people's internet behavior immediately after the lockdown enforcement, since the measure was a major disruption to their daily routine. Second, we hypothesized that the first announcement of a potential association between chilblain-like lesions and COVID-19 could have led to curiosity-driven internet searches. Therefore, GT data on acral lesions were assessed and compared through 3 periods: (1) from early 2020 until the lockdown was enforced, (2) from lockdown enforcement to the date when the association between acral lesions and COVID-19 was first reported, and (3) from the date of the first report of the aforementioned association until May 16, 2020.

For each country, we documented the dates of lockdown enforcement and the dates when an association between the acral lesions and COVID-19 was first reported (Table 1). The press releases or scientific communications on chilblain-like lesions are indicated in Table 2.

Table 2. First nationwide media release and scientific communication related to chilblain-like lesions in France, Spain, Italy, Germany, the United Kingdom, and the United States.

Country	First chilblain-like lesion–related communication	Date
France	Scientific communication: [20]; press release: [21]	April 6, 2020
Spain	Scientific communication: [22]	April 9, 2020
Italy	Press release: [23,24]	April 9, 2020
United Kingdom	Press release: [25-27]	April 14, 2020
United States	Press release: [28]	April 14, 2020
Germany	Press release: [29]	April 20, 2020

Generation of the Data Set

GT data from January 1 to May 16, 2020, were extracted for all six countries. We selected January 1 as the starting date to assess baseline interest before the pandemic in Europe, and May 16 because it allowed for data collection before, during, and after the lockdown (varying depending on the government measures in each country).

Day-to-day relative search volume (RSV) data were retrieved from GT on May 19, 2020. RSVs were calculated by GT as follows: search results are proportionate to the time and location of a query. Each data point is divided by the total searches from that geographical region and the time range it represents to compare relative popularity. The resulting numbers of a topic are then scaled 0-100 on the basis of its proportion to all searches on all topics.

Time-Series Analysis

For each country, the RSVs for acral lesions, COVID-19, and the ratio of the RSVs for acral lesions to those of COVID-19 were plotted from March 1 to May 16, 2020.

To assess the impact of external factors including lockdown enforcement and press releases, we used a segmented linear regression model with the approach described by Lagarde [30]. The dependent variable was RSVs, and the independent variable was time. The dates of lockdown enforcement and press release in each country were included in the model as interventions to evaluate their impact on internet search behavior. Table 3 describes the variables in this model.

The results are expressed as estimated variation coefficient (SE) values. Autocorrelation was expected owing to the nature of the data and confirmed through a Durbin-Watson test. Thus,

XSL•FO
RenderX

regression analysis was performed using the Prais-Winsten method, a generalized least squares estimator. Statistical analysis was performed using R software (version 4.0.0, The R Foundation). All tests were two-tailed, and *P*.05 indicated significance.

Media contamination among countries might have occurred herein. People in one country might be aware of chilblain-like COVID-19 symptoms through foreign media before their own national media reported such information. We conducted a sensitivity analysis to assess this phenomenon, using the date of the first public communication on chilblain-like symptoms (April 6, 2020) as the date of the press release (instead of the actual date of the first press release in each country).

Table 3. Variables in the segmented linear regression model.

Variables	Description	Details
Dependent variable: relative search volume	Relative search volumes	Extracted from Google Trends
Independent variables		
Time	Secular trend	Number of days since January 1, 2020
Media announcement	Change in level due to media announcement	Dichotomous variable indicating whether the first media announcement in each country had already occurred
Time since media announcement	Change in trend due to media announcement	Number of days since the first media announcement in the country (0 before this date)
Lockdown	Change in level due to the lockdown	Dichotomous variable indicating whether the lockdown was already enforced in each country
Time since lockdown enforcement	Change in trend due to the lockdown	Number of days since nationwide lockdown enforcement (0 before this date)

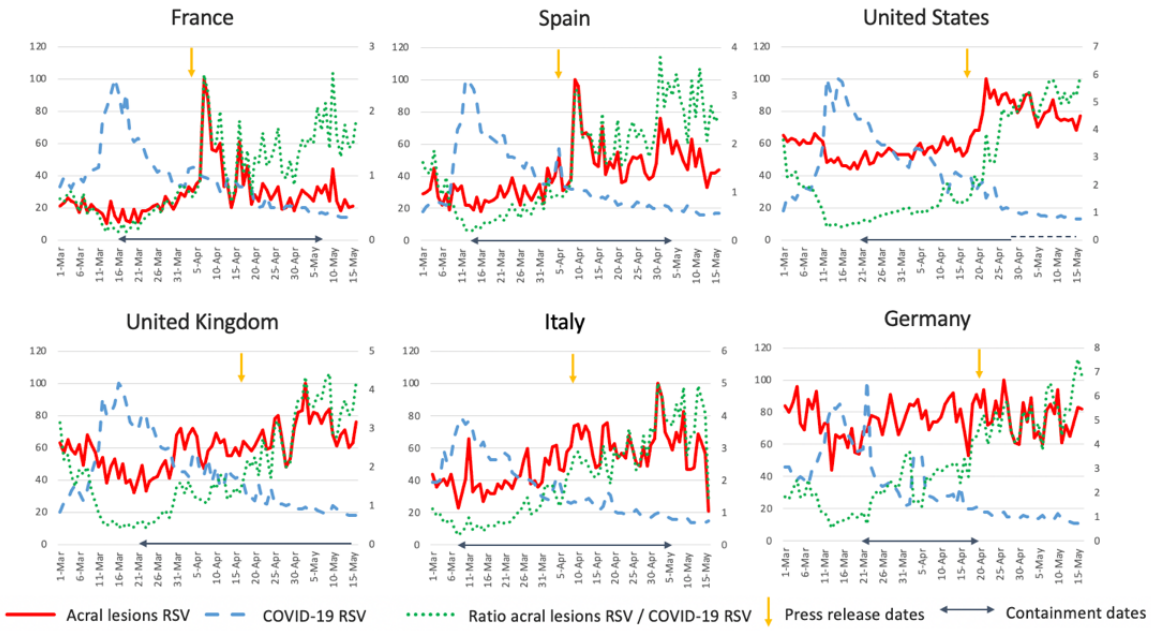
Results

RSV for COVID-19 or Acral Lesions

The frequency of queries for acral lesion–related terms did not vary from baseline values between January 1 and March 1, 2020. COVID-19–related queries increased from March 1 (Figure 1), peaking in all six countries between March 10 and 20, 2020, corresponding to the pandemic outbreak and lockdown enforcement policies in Europe. In three countries (France, Spain, and the United States), acral lesion RSVs peaked immediately after press releases. The ratios between acral lesion RSVs and COVID-19 RSVs show a constant upward trend, suggesting a relative increase in acral lesion–related searches rather than for COVID-19–related searches (Figure 1).

XSL•FO
RenderX

Figure 1. Graphs representing chilblain-like lesion relative search volumes (RSV), COVID-19 RSV, and the ratio of chilblain-like lesion RSV to COVID-19 RSV for France, Italy, Spain, Germany, the United States, and the United Kingdom between March 1 and May 16, 2020. Dates of press release are indicated with yellow arrows.

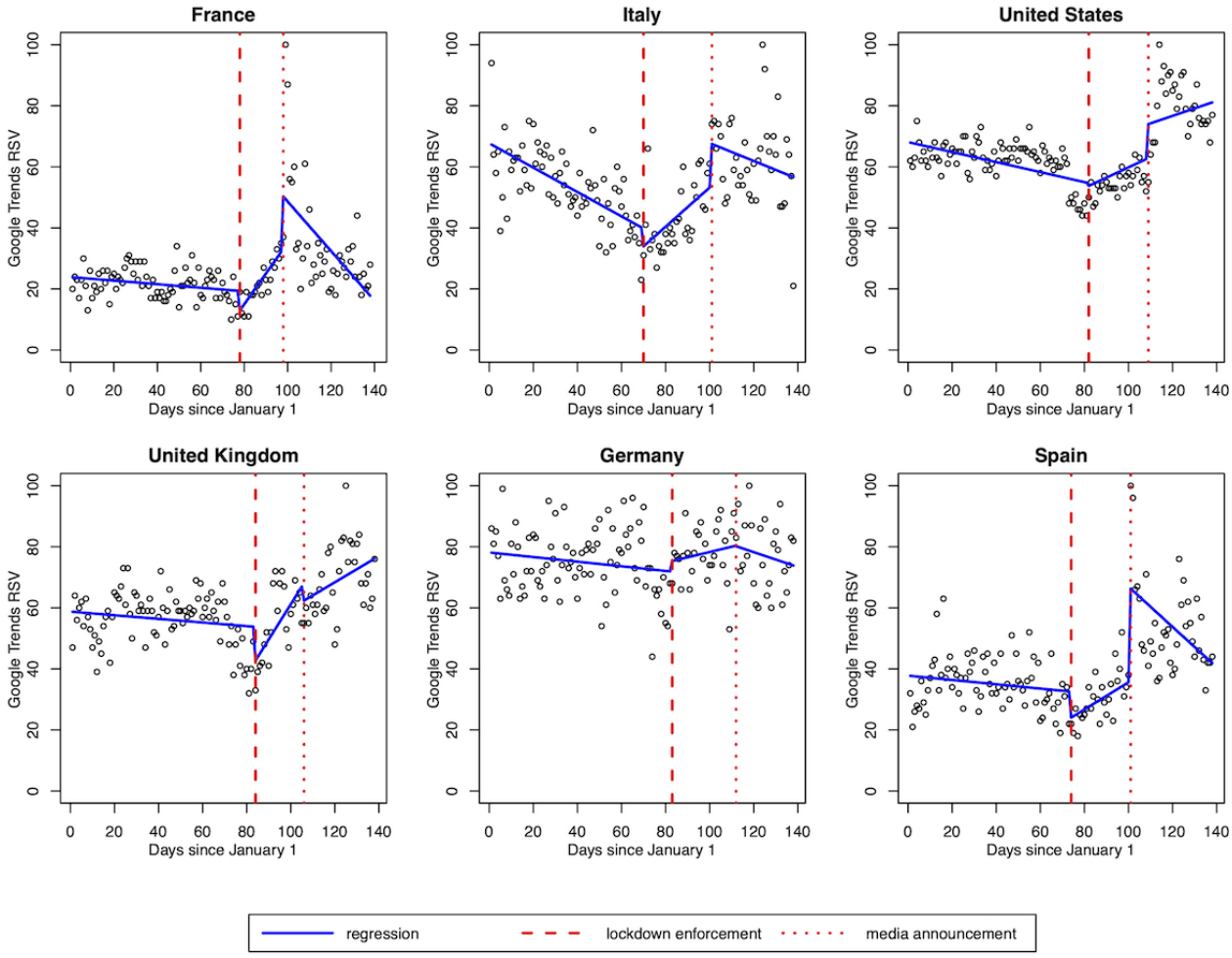


Interrupted Time-Series Analysis

On segmented linear regression analysis, lockdown enforcement was associated with a significant slope change in acral lesion–related searches in France, Italy, and the United Kingdom, with variation coefficients of 1.06 (SE 0.42), 1.04 (SE 0.28), and 1.21 (SE 0.44), respectively. A nonsignificant upward trend was observed in Germany, the United States, and Spain (Figure 2).

XSL•FO
RenderX

Figure 2. Segmented linear regression model for acral lesion–related relative search volumes. The model integrates the dates of lockdown enforcement and acral lesion–related press releases for France, Italy, Spain, Germany, the United States, and the United Kingdom. RSV: relative search volume.



RSVs significantly increased before press releases in France, Spain, Italy, and the United States, with variation coefficients of 18.95 (SE 5.77), 31.31 (SE 6.31), 14.57 (SE 6.33), and 11.24 (SE 4.93), respectively, with no significant changes in Germany and the United Kingdom (0.23 [SE 6.48]; *P*=.97 and –4.99 [SE 6.68]; *P*=.46, respectively). RSVs displayed a significant downward trend after press releases in France, Spain, and Italy (variation coefficients –1.82 [SE 0.45], –1.10 [SE 0.38], and –0.93 [SE 0.33], respectively). In the United States and the United Kingdom, RSVs displayed an upward trend after press releases, albeit in a nonsignificant manner (variation coefficients –0.09 [SE 0.42]; *P*=.83 and –0.72 [SE 0.53]; *P*=.18, respectively).

Adjusted R^2 values were 0.311, 0.351, 0.325, and 0.305 for France, Spain, Italy, and the United States, respectively, suggesting an average correlation; those for Germany and the United Kingdom were 0.017 and 0.147, respectively, suggesting a weak correlation (Tables 4 and 5).

XSL•FO
RenderX

Table 4. Coefficient values and adjusted R^2 values of the linear segmented regression model integrating dates of nationwide lockdown enforcement and of acral lesion–related press releases in France, Spain, and the United States.

Trends	France		Spain		United States	
	Variation coefficient (SE)	*P* value	Variation coefficient (SE)	*P* value	Variation coefficient (SE)	*P* value
Constant	23.89[a] (2.65)	<.001	37.77[a] (3.21)	<.001	68.08[a] (3.20)	<.001
Secular trend	–0.05 (0.05)	.32	–0.07 (0.07)	.35	–0.16[b] (0.06)	.01
Change in level after lockdown	–7.45 (5.61)	.19	–9.01 (5.87)	.13	–1.38 (4.63)	.77
Change in trend after lockdown	1.06[b] (0.42)	.01	0.51 (0.33)	.12	0.49 (0.29)	.09
Change in level after press releases	18.95[b] (5.77)	.001	31.31[a] (6.31)	<.001	11.24[b] (4.93)	.02
Change in trends after press releases	–1.82[a] (0.45)	<.001	–1.10[b] (0.38)	.004	–0.09 (0.42)	.83
Adjusted R^2	0.311	__[c]	0.351	—	0.305	—

[a]*P*<.001

[b]*P*<.05.

[c]Not applicable.

Table 5. Coefficient values and adjusted R^2 values of the linear segmented regression model integrating dates of nationwide lockdown enforcement and of acral lesion–related press releases in the United Kingdom, Italy, and Germany.

Trends	United Kingdom[a]		Italy		Germany	
	Variation coefficient (SE)	*P* value	Variation coefficient (SE)	*P* value	Variation coefficient (SE)	*P* value
Constant	58.76[b] (3.35)	<.001	67.7[b] (3.32)	<.001	78.14[b] (2.77)	<.001
Secular trend	–0.06 (0.07)	.39	–0.39[b] (0.08)	<.001	–0.07 (0.05)	.20
Change in level after lockdown	–12.24 (6.27)	.05	–6.85 (5.81)	.24	3.21 (5.37)	.55
Change in trend after lockdown	1.21[c] (0.44)	.01	1.04[b] (0.28)	<.001	0.24 (0.27)	.37
Change in level after press releases	–4.99 (6.68)	.46	14.57[c] (6.33)	.02	0.23 (6.48)	.97
Change in trends after press releases	–0.72 (0.53)	.18	–0.93[b] (0.33)	<.001	–0.41 (0.41)	.31
Adjusted R^2	0.147	__[d]	0.325	—	0.017	—

[a]A sensitivity analysis was performed considering April 6, 2020, as the date of press release. This modification did not improve the fit for 4 of 6 countries (France, Italy, the United States, and Spain). For Germany, the adjusted R^2 value increased from 0.017 to 0.018; United Kingdom, from 0.147 to 0.209. We obtained the following estimates for model variables: constant 59.12 (SE 3.05; *P*<.001), secular trend –0.07 (SE 0.063; *P*=.26), change in level after lockdown –20.5 (SE 6.78; *P*=.003), change in trend after lockdown 2.90 (SE 0.75; *P*<.001), change in level after press releases –15.4 (SE 6.63; *P*=.02), and change in trend after press releases –2.34 (SE 0.787; *P*=.003).

[b]*P*<.001

[c]*P*<.05.

[d]Not applicable.

Discussion

Principal Findings

Our study shows that GT searches for chilblain-like lesions are influenced by lockdown enforcement and media coverage. Chilblain-like lesion–related searches were the highest in France, Spain, Italy, and the United States, seemingly affected by press releases about an outbreak of acral lesions.

While a downward trend was observed in European countries after the press release, regression analysis revealed an upward trend in the United States. However, visual and regression analyses suggest that the model does not fit the US data well, probably owing to the time parameters selected for the model because the dates of lockdown enforcement, press releases, and disease progression differed among states in the United States.

XSL•FO
RenderX

This could also reflect the variations in COVID-19 dynamics nationwide.

In France, the United Kingdom, and Italy, the lockdown policies significantly impacted GT dynamics. Lockdown enforcement did not have a significant impact in other countries. In Germany, neither the lockdown nor press releases influenced individual searches related to acral lesions. In all other countries, individual RSV searches increased before the press release. Simultaneously, the search ratio for acral lesions vs COVID-19 remained high, indicating a high interest in acral lesions rather than in COVID-19.

Studies reporting an increase in public interest in COVID-19 before the outbreak in several countries supported the potential role of GT as a surveillance tool [31]. However, for novel diseases such as COVID-19, media coverage or significant announcements by the World Health Organization may strongly influence generic searches. Studies have reported a correlation between new COVID-19 cases or deaths and queries related to COVID-19 symptoms (specific: fever, cough, or pneumonia; unspecific: anosmia or ageusia). However, media releases regarding new clinical signs also strongly influenced internet searches [32-34]. This calls into question the potential to successfully predict the pandemic by analyzing GT data. Regarding anosmia and ageusia, an increase in RSV searches related to acral lesions preceded public scientific communications [18,35,36]. Our results confirm the generation of public interest before press releases, with a peak corresponding to media coverage in several European countries; however, caution must be exercised when using GT as an epidemiological monitoring tool for new or unknown diseases.

Beyond epidemic monitoring, our results illustrate the use of GT in monitoring the reactions of the general population to health-related communications during a period with a high incidence of anxiety and disrupted health care provision. Spikes in interest generated by public announcements on potential new manifestations of the disease should be matched with appropriate health-related information [37]. Some studies have revealed a change in search patterns during COVID-19 for a broad range of health-related topics, including otorhinolaryngologic, lung, or dermatological diseases [33,38,39]. These queries should be matched with appropriate health-related communication to ensure that people can easily obtain the right information and take appropriate action. The increase in searches for a broad range of health-related topics might also suggest that curiosity and inquisitiveness played an essential role in these internet searches, which would undermine attempts to use GT to identify changes in disease dynamics. Moreover, physicians might have also carried out internet searches for chilblains.

We did not perform correlation analyses with the number of COVID-19–related deaths. Indeed, chilblain-like lesions are associated with mild or asymptomatic SARS-CoV-2 infection [40,41], thus undermining correlation analyses with the number of COVID-19–related deaths, and correlations with the number of detected COVID-19 cases are unreliable owing to the limited number of tests performed at the beginning of the pandemic in many countries. Therefore, caution must be exercised when performing correlation analyses between GT data on chilblain-like symptoms and COVID-19 progression.

Limitations

This study has some limitations of note. First, the results depend on the quality of the initial GT searches, especially keywords. Query RSVs vary greatly in accordance with the search terms. We attempted to pragmatically use search terms that could be used by medical practitioners (such as “chilblain”) and nonmedical individuals (such as “toes”). It was difficult to identify dermatological keywords for the study. We inquired about the potential terms the general population might have used to search for chilblain-like lesions (showing images of chilblains). Most of our participants with a nonmedical background would have searched “eczema” or “urticaria” on Google. In dermatology, it would be ideal to conduct the same study with individuals clicking on pictures resembling their cutaneous lesions [10].

We used direct language translations and did not include misspellings, accents, and special character variations. Thus, we may have missed certain nuances and additional queries. However, we believe that using only the correct spelling might underestimate the RSV, making our results even more robust. To evaluate the potential limitation due to keywords, we assessed the results of “related queries” (top and rising) for the same dates and regions and compared them with GT search terms. It appears that “coronavirus” or “covid” is often associated with “toe” or “frostbite.” However, we intentionally excluded COVID-19–related keywords in order to capture searches performed by individuals who did not consider a potential association between acral lesions and COVID-19.

Furthermore, our data set is limited to the pandemic's first wave between January and May 2020. Studies using these initial data should be repeated when more data are available to determine whether the results remain applicable for subsequent waves of the pandemic.

In addition, our nationwide analyses did not consider the heterogeneity of the regional effects of the pandemic in these countries. This holds particularly true for Italy and the United States, which have been severely affected by the pandemic (eg, Lombardy and New York state, respectively). Thus, state-level analysis of pandemic hotspots might have yielded more reliable results. In some countries including Germany, the incidence of COVID-19 was relatively low. Another study limitation is our use of only one search engine; nonetheless, Google is the most frequently used internet search engine in these six countries [42].

Moreover, internet users are not necessarily representative of the general population, especially in high-income countries with an increasingly elderly population. However, chilblain-like lesions are observed in young patients, who are more prone to searching information on the internet.

Finally, the pathophysiological link between chilblain-like lesions and COVID-19 remains unclear [16,40,43] and is beyond the scope of this study.

XSL•FO
RenderX

Conclusions

This study suggests that during the first wave of COVID-19, public announcements and government policies served as driving factors for Google queries related to acral lesions. For emerging diseases, the use of GT to detect new symptoms and monitor disease progression is limited by high media coverage levels and government policies.

Acknowledgments

We are grateful to Dr Jason Shourick for his advice and Dr Charbel Skayem for English editing.

Authors' Contributions

SHD, GL, TH, KE, and TAD reviewed all the study data and ensured data integrity and accuracy of the data analysis. SHD and TAD conceived and designed the study. SHD acquired the data. All authors analyzed and interpreted the data. SHD, KE, and TAD drafted the manuscript. All authors critically revised the manuscript. SHD and GL performed the statistical analysis. TAD provided administrative, technical, and material support and supervised the study. All the authors reviewed and approved the final version submitted for publication.

Conflicts of Interest

None declared.

References

1. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The Use of Google Trends in Health Care Research: A Systematic Review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
2. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
3. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping Review on Search Queries and Social Media for Disease Surveillance: A Chronology of Innovation. J Med Internet Res 2013 Jul;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
4. Mavragani A. Infodemiology and Infoveillance: Scoping Review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
5. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
6. Eysenbach G. Infodemiology and Infoveillance Tracking Online Health Information and Cyberbehavior for Public Health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
7. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
8. Cousins HC, Cousins CC, Harris A, Pasquale LR. Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns. J Med Internet Res 2020 Jul 30;22(7):e19483 [FREE Full text] [doi: 10.2196/19483] [Medline: 32692691]
9. Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 Novel Coronavirus (SARS-CoV-2) Based on Current Evidence. Int J Antimicrob Agents 2020 Jun;55(6):105948 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105948] [Medline: 32201353]
10. Darlenski R, Tsankov N. COVID-19 Pandemic and the Skin: What Should Dermatologists Know? Clin Dermatol 2020;38(6):785-787 [FREE Full text] [doi: 10.1016/j.clindermatol.2020.03.012] [Medline: 33341217]
11. Duong TA, Velter C, Rybojad M, Comte C, Bagot M, Sulimovic L, et al. Did Whatsapp Reveal a New Cutaneous COVID-19 Manifestation? J Eur Acad Dermatol Venereol 2020 Aug;34(8):e348-e350 [FREE Full text] [doi: 10.1111/jdv.16534] [Medline: 32330322]
12. Bouaziz J, Duong T, Jachiet M, Velter C, Lestang P, Cassius C, et al. Vascular Skin Symptoms in COVID-19: a French Observational Study. J Eur Acad Dermatol Venereol 2020 Sep;34(9):e451-e452 [FREE Full text] [doi: 10.1111/jdv.16544] [Medline: 32339344]
13. Marzano AV, Genovese G. Response to: "Reply to 'Varicella-like exanthem as a specific COVID-19-associated skin manifestation: multicenter case series of 22 patients': To consider varicella-like exanthem associated with COVID-19, virus varicella zoster and virus herpes simplex must be ruled out". J Am Acad Dermatol 2020 Sep;83(3):e255-e256 [FREE Full text] [doi: 10.1016/j.jaad.2020.05.072] [Medline: 32442697]
14. Kluger N, Scrivener JN. The Use of Google Trends for Acral Symptoms During COVID-19 Outbreak in France. J Eur Acad Dermatol Venereol 2020 Aug;34(8):e358-e360. [doi: 10.1111/jdv.16572] [Medline: 32358813]

XSL•FO
RenderX

15. Hughes M, Rogers S, Lepri G, Bruni C, Matucci-Cerinic M. Further Evidence that Chilblains are a Cutaneous Manifestation of COVID-19 Infection. Br J Dermatol 2020 Sep;183(3):596-598 [FREE Full text] [doi: 10.1111/bjd.19243] [Medline: 32460349]
16. Kluger N. Why are Chilblains Underreported in Nordic Countries During the COVID-19 Pandemic? An Analysis of Google Trends. J Eur Acad Dermatol Venereol 2021 Feb;35(2):e100-e101 [FREE Full text] [doi: 10.1111/jdv.16974] [Medline: 33010073]
17. Sousa-Pinto B, Anto A, Czarlewski W, Anto J, Fonseca A, Bousquet J. Assessment of the Impact of Media Coverage on COVID-19-Related Google Trends Data: Infodemiology Study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
18. Walker A, Hopkins C, Surda P. Use of Google Trends to Investigate Loss-of-Smell-Related Searches During the COVID-19 Outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
19. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
20. Le Syndicat National des Dermatologues-Vénéréologues (SNDV) alerte sur les manifestations cutanées du COVID 19. Syndicat National des Dermatologues - Vénéréologues. 2020 Apr 06. URL: https://www.syndicatdermatos.org/wp-content/uploads/2020/04/cp-covid-peau-6-avril.pdf [accessed 2021-02-09]
21. Pecnard J, Frouville M. Les dermatologues alertent sur de nouveaux symptômes cutanés du COVID-19. BFM TV. 2020 Apr 06. URL: https://www.bfmtv.com/sante/les-dermatologues-alertent-sur-de-nouveaux-symptomes-cutanes-du-covid-19_AN-202004060271.html [accessed 2021-02-09]
22. Registro De Casos Compatibles COVID-19. Consejo General de Colegios Oficiales de Podólogos. 2020 Apr 09. URL: https://cgcop.es/registro-de-casos-compatibles-covid-19/ [accessed 2021-02-09]
23. No está demostrado que haya niños con sintomatología de Coronavirus en manos y pies. EITB.EUS. 2020 Apr 09. URL: https://www.eitb.eus/es/noticias/sociedad/detalle/7156303/no-esta-demostrado-haya-ninos-sintomas-coronavirus-manos-pies/ [accessed 2021-02-09]
24. Detectadas lesiones dermatológicas en los pies de niños y adolescentes con coronavirus. 20Minutos. 2020 Apr 09. URL: https://www.20minutos.es/noticia/4222618/0/detectadas-lesiones-dermatologicas-pies-ninos-adolescentes-coronavirus/ [accessed 2021-02-09]
25. Mullin G. NEW SIGN Bizarre marks on your feet could be ‘early coronavirus symptom’, experts discover. The Sun. 2020 Apr 14. URL: https://www.thesun.co.uk/news/11392789/marks-on-feet-early-coronavirus-symptom/ [accessed 2021-02-09]
26. Chalmers V. Could bruises on your FEET be a sign of coronavirus? Doctors suspect a link after patients with chickenpox-like lesions on their toes test positive for the killer infection - as another study warns headaches and feeling dizzy may also be signs of COVID-19. Daily Mail. 2020 Apr 14. URL: https://www.dailymail.co.uk/news/article-8217901/Could-bruises-FEET-sign-coronavirus-Doctors-suspect-link.html [accessed 2021-02-09]
27. Pochin C. Possible sign of coronavirus to look out for on feet - even if you don't have symptoms. Mirror. 2020 Apr 14. URL: https://www.mirror.co.uk/lifestyle/health/possible-sign-coronavirus-look-out-21862683 [accessed 2021-02-09]
28. Salo J. Foot sores could be an early sign of coronavirus, experts say. New York Post. 2020 Apr 14. URL: https://nypost.com/2020/04/14/foot-sores-could-be-an-early-sign-of-coronavirus-experts-say/ [accessed 2021-02-09]
29. Blaue Flecken an Zehen könnten auf Corona hindeuten. Bild. 2020 Apr 20. URL: https://www.bild.de/ratgeber/gesundheit/gesundheit/coronavirus-blaue-flecken-an-zehen-koennen-auf-covid-19-hindeuten-70151296.bild.html [accessed 2021-02-09]
30. Lagarde M. How to do (or not to do) ... Assessing the Impact of a Policy Change with Routine Longitudinal Data. Health Policy Plan 2012 Jan;27(1):76-83. [doi: 10.1093/heapol/czr004] [Medline: 21278077]
31. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 Pandemic With Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
32. Szmuda T, Ali S, Hetzger TV, Rosvall P, Słoniewski P. Are Online Searches for the Novel Coronavirus (COVID-19) Related to Media or Epidemiology? A Cross-sectional Study. Int J Infect Dis 2020 Aug;97:386-390 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.028] [Medline: 32535297]
33. Walker MD, Sulyok M. Online Behavioural Patterns for Coronavirus Disease 2019 (COVID-19) in the United Kingdom. Epidemiol Infect 2020 Jun 05;148:e110 [FREE Full text] [doi: 10.1017/S0950268820001193] [Medline: 32498731]
34. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence From Internet Search Data Shows Information-Seeking Responses to News of Local COVID-19 Cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
35. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. Otolaryngol Head Neck Surg 2020 Sep;163(3):491-497 [FREE Full text] [doi: 10.1177/0194599820932128] [Medline: 32484425]
36. Sousa-Pinto B, Anto A, Czarlewski W, Anto J, Fonseca J, Bousquet J. Assessment of the Impact of Media Coverage on COVID-19-Related Google Trends Data: Infodemiology Study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]

XSL•FO
RenderX

37. Lieberman-Cribbin W, Alpert N, Gonzalez A, Schwartz R, Taioli E. Three Months of Informational Trends in COVID-19 Across New York City. J Public Health (Oxf) 2020 Aug 18;42(3):448-450 [FREE Full text] [doi: 10.1093/pubmed/fdaa082] [Medline: 32542402]
38. Barbosa M, Morais-Almeida M, Sousa C, Bousquet J. The "Big Five" Lung Diseases in CoViD-19 Pandemic - a Google Trends analysis. Pulmonology 2021;27(1):71-72 [FREE Full text] [doi: 10.1016/j.pulmoe.2020.06.008] [Medline: 32636165]
39. Searle T, Al-Niaimi F, Ali FR. Google Searches in Patient Self-Care Increase During the COVID-19 Lockdown. J Am Acad Dermatol 2020 Nov;83(5):e349 [FREE Full text] [doi: 10.1016/j.jaad.2020.06.1028] [Medline: 32707255]
40. Hubiche T, Le Duff F, Chiaverini C, Giordanengo V, Passeron T. Negative SARS-CoV-2 PCR in Patients With Chilblain-like Lesions. Lancet Infect Dis 2020 Jun 18 [FREE Full text] [doi: 10.1016/S1473-3099(20)30518-1] [Medline: 32563281]
41. Hubiche T, Cardot-Leccia N, Le Duff F, Seitz-Polski B, Giordana P, Chiaverini C, et al. Clinical, Laboratory, and Interferon-Alpha Response Characteristics of Patients With Chilblain-like Lesions During the COVID-19 Pandemic. JAMA Dermatol 2020 Nov 25. [doi: 10.1001/jamadermatol.2020.4324] [Medline: 33237291]
42. The Top 500 Sites on the Web. Alexa. URL: https://www.alexa.com/topsites/countries [accessed 2021-02-09]
43. Herman A, Peeters C, Verroken A, Tromme I, Tennstedt D, Marot L, et al. Evaluation of Chilblains as a Manifestation of the COVID-19 Pandemic. JAMA Dermatol 2020 Sep 01;156(9):998-1003. [doi: 10.1001/jamadermatol.2020.2368] [Medline: 32584377]

Abbreviations

GT: Google Trends
RSV: relative search volume

Edited by G Eysenbach; submitted 10.11.20; peer-reviewed by B Sousa-Pinto, A Mavragani; comments to author 25.11.20; revised version received 15.12.20; accepted 22.01.21; published 25.02.21.

Please cite as:
Huynh Dagher S, Lamé G, Hubiche T, Ezzedine K, Duong TA
The Influence of Media Coverage and Governmental Policies on Google Queries Related to COVID-19 Cutaneous Symptoms: Infodemiology Study
JMIR Public Health Surveill 2021;7(2):e25651
URL: https://publichealth.jmir.org/2021/2/e25651
doi:10.2196/25651
PMID:33513563

©Solene Huynh Dagher, Guillaume Lamé, Thomas Hubiche, Khaled Ezzedine, Tu Anh Duong. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 25.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis

Katja Reuter[1,2*], PhD; Melissa L Wilson[3*], PhD; Meghan Moran[4], PhD; NamQuyen Le[5], MPH; Praveen Angyan[5], MS; Anuja Majmundar[6], MA, MBA, PhD; Elsi M Kaiser[7*], MS, PhD; Jennifer B Unger[2*], PhD

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, United States

[2]Institute for Health Promotion and Disease Prevention Research, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[3]Division of Disease Prevention, Policy and Global Health, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[4]Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[6]Economic and Health Policy Research, American Cancer Society, Washington, DC, United States

[7]Department of Linguistics, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Katja Reuter, PhD
Department of Public Health & Preventive Medicine
SUNY Upstate Medical University
766 Irving Avenue
Syracuse, NY, 13210
United States
Phone: 1 315 464 1520
Email: reuterk@upstate.edu

Related Article:

This is a corrected version. See correction statement: https://publichealth.jmir.org/2021/2/e28131/

Abstract

Background: Public health organizations have begun to use social media to increase awareness of health harm and positively improve health behavior. Little is known about effective strategies to disseminate health education messages digitally and ultimately achieve optimal audience engagement.

Objective: This study aims to assess the difference in audience engagement with identical antismoking health messages on three social media sites (Twitter, Facebook, and Instagram) and with a referring link to a tobacco prevention website cited in these messages. We hypothesized that health messages might not receive the same user engagement on these media, although these messages were identical and distributed at the same time.

Methods: We measured the effect of health promotion messages on the risk of smoking among users of three social media sites (Twitter, Facebook, and Instagram) and disseminated 1275 health messages between April 19 and July 12, 2017 (85 days). The identical messages were distributed at the same time and as organic (unpaid) and advertised (paid) messages, each including a link to an educational website with more information about the topic. Outcome measures included message engagement (ie, the click-through rate [CTR] of the social media messages) and educational website engagement (ie, the CTR on the educational website [wCTR]). To analyze the data and model relationships, we used mixed effects negative binomial regression, z-statistic, and the Hosmer-Lemeshow goodness-of-fit test.

XSL•FO
RenderX

Results: Comparisons between social media sites showed that CTRs for identical antitobacco health messages differed significantly across social media (*P*<.001 for all). Instagram showed the statistically significant highest overall mean message engagement (CTR=0.0037; 95% CI 0.0032-0.0042), followed by Facebook (CTR=0.0026; 95% CI 0.0022-0.0030) and Twitter (CTR=0.0015; 95% CI 0.0013-0.0017). Facebook showed the highest as well as the lowest CTR for any individual message. However, the message CTR is not indicative of user engagement with the educational website content. Pairwise comparisons of the social media sites differed with respect to the wCTR (*P*<.001 for all). Messages on Twitter showed the lowest CTR, but they resulted in the highest level of website engagement (wCTR=0.6308; 95% CI 0.5640-0.6975), followed by Facebook (wCTR=0.2213; 95% CI 0.1932-0.2495) and Instagram (wCTR=0.0334; 95% CI 0.0230-0.0438). We found a statistically significant higher CTR for organic (unpaid) messages (CTR=0.0074; 95% CI 0.0047-0.0100) compared with paid advertisements (CTR=0.0022; 95% CI 0.0017-0.0027; *P*<.001 and *P*<.001, respectively).

Conclusions: Our study provides evidence-based insights to guide the design of health promotion efforts on social media. Future studies should examine the platform-specific impact of psycholinguistic message variations on user engagement, include newer sites such as Snapchat and TikTok, and study the correlation between web-based behavior and real-world health behavior change. The need is urgent in light of increased health-related marketing and misinformation on social media.

(JMIR Public Health Surveill 2021;7(2):e24429) doi:10.2196/24429

KEYWORDS

affordance; digital; dissemination of science; Facebook; health communication; health promotion; Instagram; online; smoking; social media; tobacco; Twitter; user engagement

Introduction

With the emergence of social media, public health organizations face new opportunities and challenges. Social media include widely accessible web-based and mobile information tools that allow users to view, create, and share messages with others on the web [1]. Overall, 72% of American adults and 97% of teens aged 13-17 years teens say they use at least one social media site, many of them daily [2-4]. Public health groups can use social media to instantly reach more people than ever [5-8]. On the other hand, social media users are increasingly exposed to health-related misinformation, polarization, and targeted commercial marketing of potentially health-harming products and practices, and previous work suggests that the public is exposed to widespread antivaccination messages [9,10], e-cigarette endorsements [11-13], and medical misinformation about the COVID-19 pandemic [14]. Messages that promote tobacco, for example, outnumber antitobacco posts on social media, raising concerns about their effects on the users of these sites, especially members of vulnerable populations such as youth [12,15-17]. Public health groups will need to find innovative and cost-effective ways to increase their information output as one way to counterbalance the overabundance of marketing and misinformation.

Public health groups have started to use social media for health promotion to increase awareness of health harms and positively change behavioral intent [18-20]. Evidence-based health promotion and interventions on social media are an area of growing interest among public health groups. A growing number of systematic reviews have examined social media–based interventions for a variety of health topics, diseases, and behavioral risks [21-24]. As suggested by previous research and outlined by social media measurement standards, audience engagement is an important component of social media–based interventions. Engagement goes beyond mere exposure to a social media post and involves the interactions between an audience and an organization and includes activities that indicate acceptance and involvement with a message, such as liking or sharing a post or clicking a link [25-27]. In general, audience engagement with social media messages indicates their interest and involvement and offers possibilities for widespread message dissemination to peers within their networks [28].

Differences in structural layers of communication across social media platforms can play a crucial role in determining the extent of audience engagement with health messages [29]. Different social media platforms have different features that may facilitate audience engagement and, in particular, may facilitate the ability for campaigns to drive audiences to websites or link them with health education programs. For instance, Instagram users may engage with ephemeral content (eg, *stories*) or static content (eg, *posts*). Users may have different privileges regarding their ability to directly link audiences to web content outside of the Instagram platform. Twitter users may subscribe to and engage with content associated with specific hashtags posted by nonnetwork peers, which could facilitate a campaign's ability to reach a wider audience and connect them to resources offered outside of the social media platform. Facebook users may engage with content posted on private or public community group *pages*, allowing an organization to connect with their audience. Thus, different social media platforms may have different capacity to engage audiences. However, to date, no prior work has systematically examined audience engagement with health promotion messages across social media platforms.

In a health promotion context, addressing the above-mentioned gap is crucial to inform future health promotion communication and intervention strategies [30,31]. Methodologically rigorous studies to investigate the effects of social media as part of health promotion and prevention campaigns are critically needed [24,32] to answer questions such as, Are some social media platforms more effective for public health campaigns than others in terms of getting users' engagement? Is there a relationship between the number of clicks on health messages and user engagement with the referred to educational website content? Should health groups limit health promotion campaigns to paid

XSL•FO
RenderX

advertisements or is it worthwhile investing time in organic (unpaid) social media efforts that rely more heavily on developing and engaging a community of followers? Past social media research has either focused predominantly on general advertising research or on the assessment of health campaigns within the context of a single social media platform [32]. Few studies compared the performance of social media messages across different platforms [33].

This study assesses the difference in user engagement with identical health messages on three social media sites, including Twitter, Facebook, and Instagram. We defined the message effectiveness as digital user engagement, which was assessed via two primary outcome variables: (1) health message engagement: the click-through rate (CTR) of the social media messages and (2) educational website engagement: the CTR of the educational website the messages linked to (wCTR). CTRs are important social media metrics because they indicate the extent to which the audience finds the message appealing or interesting [34] and because many campaigns make social media posts in hopes of driving the audience to educational websites. We hypothesized that the distributed health messages might not get the same engagement on the different social media sites, although these messages are identical and distributed at the same time. More specifically, we aimed to evaluate (1) the effect of the social media platform on user engagement and (2) the impact of the type of message (paid or organic) on user engagement. We chose these three platforms (ie, Twitter, Facebook, and Instagram) because they are among the most popular sites in the United States and are used daily by broad populations across the different age groups [2-4].

Our work contributes to developing a scientific approach for the selection of the appropriate social media platform for a health promotion or intervention and aligns with recent calls for more transparency of the processes and mechanisms that make digital health promotion feasible and effective [35]. If supported, this would provide important knowledge to improve the design of social media–based health education campaigns.

Methods

Study Overview

This study included the dissemination and analysis of a total of 1275 antismoking health messages posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017, as previously described [36]. The messages were focused on the risks of using combustible tobacco products. The study target population included English-speaking social media users on Facebook, Twitter, and Instagram.

For this experiment, we developed parameterized text message templates (n=102) and extracted images (n=315) from two government-sponsored health education campaigns [36]. Copy-protected images from these campaigns were replaced with similar images from a public photo repository, Stocksnap, and topic-related hashtags (n=4) from Twitter (eg, #cigs, *#cigarettes*, #smoking, *#tobaccofree)*. All messages were antismoking messages focused on the risk of using combustible tobacco products. However, the messages referred to three different themes: (1) health or appearance or addiction, (2) money, and (3) family, for example, health: *do not let #cigs cut your life short. Smoking #cigarettes can claim more than 10 years of your life;* money: *Smoking half a pack per day costs about $1000/year. Smoking can do serious damage to your wallet;* family: *about 50% of 3- to 11-year-olds are exposed to secondhand smoke. Look out for the lil ones by keeping it #tobaccofree*.

Theoretical Background

Our hypothesis draws on two theories: the Selective Exposure Self- and Affect-Management (SESAM) model, which posits that selective exposure to media content is driven by pre-existing self-concept, motivations, and affect [37], and the *affordance theory* (also called affordances framework) [38,39], which suggests that social media users' engagement with health messages depends not only on their needs but also on the characteristics of the social media site. The user interface and features of social media sites influence whether users may or may not perceive or attend to the affordances of a social media site [40-42]. Studies on Facebook, for example, showed that posts requiring a simple user response such as polls might elicit the highest engagement, whereas the most common form of engagement is the use of the *like* feature [43]. This study sheds light on whether there is a relationship between the social media platform and the user engagement with the public health messages distributed on the platform.

Procedures

Each message was posted at the most once each month for 85 days. To increase the number of message variations, we used related linguistic message variations, for example, using *we* versus *you* versus *they*. As a result, some of the messages appear similar. The messages were randomized and posted once a month. Multimedia Appendix 1 [36] provides the entire list of parameterized message templates used in the experiment. The details of the technology-enhanced implementation of the experiment were previously published, and examples of messages with images for each platform can be found in the technical paper mentioned previously [36].

We used a web-based tool (Trial Promoter) [36] to randomize the order of the messages and disseminated them at the same time in identical form as organic (unpaid) messages and paid advertisements on each social media site. Organic messages are not paid for; they are usually seen by followers and those who are interested in the same topic. In contrast, advertisements are paid messages that can be targeted to broad and hard-to-reach groups of the population based on proprietary information on user demographics and interests owned by the social media site. Paid and organic messages were posted to separate project accounts on each social media site. For the paid messages, we used a set of targeting criteria, such as gender, location, language, and age (as shown in Multimedia Appendix 2).

The daily message volume per social media platform was 6 on Facebook (advertisements and organic), 6 on Twitter (advertisements and organic), and 3 on Instagram (advertisements only because Instagram does not support referral

XSL•FO
RenderX

URLs in organic, unpaid messages). On the basis of market research showing that messages sent at these times receive the most user engagement [44,45], we posted messages on Facebook at 9 AM, 1 PM, and 3 PM PST; on Twitter at noon, 3 PM, and 5 PM PST; and on Instagram at 8 AM, 9 AM, and 5 PM PST. The length of the pilot project (85 days) was determined by the available pilot project budget for social media advertisements.

On seeing the message, users could engage with the post by commenting, sharing, liking, and clicking on the link in the message, which directed them to an educational website operational during the campaign period. The website provided more information about the risks of tobacco products, which was based on government-sponsored health education campaigns.

Data Collection and Confidentiality

Analytics were collected for each distributed message to determine the engagement among social media users with the message and on the referred educational webpage [36]. The information we analyzed for this study is aggregate and nonidentifiable, such as message clicks, message impressions, and website clicks. The data were stored in Trial Promoter [36], which was hosted by the cloud-based hosting provider, Heroku, a Salesforce application. Salesforce has passed security and privacy-related audits and certifications, including the EU-US Privacy Shield Framework and TRUSTe Certification [46]. Study approval was obtained from the institutional review board at the University of Southern California (protocol #HS-16-00660).

Calculation of Message and Website CTRs

Message effectiveness was defined as digital user engagement, which was assessed via two primary outcome variables: (1) health message engagement: the CTR of the social media messages and (2) educational website engagement: the CTR of the educational website the messages linked to (wCTR). The CTR was defined as the total number of clicks on the message link divided by the total number of impressions for a specific message. Impressions describe the number of times a message was served to potential viewers, as reported by the social media platform. The wCTR was defined as the proportion of those who, after clicking the message link, visited the educational website and scrolled to peruse the content on the landing page.

It was calculated as the number of scrolls on the website divided by the total number of sessions on the website. A session is a group of interactions that a user takes within 30 minutes on a website.

Data Analysis

Differences in Audience Engagement Across Social Media Platforms

We evaluated the message effectiveness across different social media platforms (Facebook, Twitter, and Instagram). The CTR and wCTR were calculated for each social media platform, as described. To evaluate whether the social media platform was associated with the CTR and wCTR, we used mixed effects negative binomial regression with the exposure being the social media type and specifying the message variant as a random effect. We also included the number of impressions or sessions in the model, depending on whether we were estimating the CTR or wCTR, respectively, but constrained the coefficient to 1.0. The models for the wCTR had an overdispersion parameter set to constant=1+delta. A negative binomial Poisson regression model was selected because the data consist of click counts and a large number of messages received 0 clicks, thereby creating an overdispersed distribution.

On the basis of each model, we obtained adjusted predictive margins for the CTR or wCTR by social media site and tested the null hypothesis that there were no differences in the adjusted predictive margins between groups using a z-statistic. No data on user characteristics were collected (ie, data were aggregated by each appearance of the message), and thus, there were no user-specific predictors or confounders available for inclusion in the model. Where data were sparse, we present the results descriptively.

For all models, fit was evaluated via the Hosmer-Lemeshow goodness-of-fit test, plots of residuals, and inspection of model outliers. The Hosmer-Lemeshow test evaluates how well the observed and predicted values from the model align. Both the deviance statistic and Pearson statistic are reported. When the model is inappropriate, the test is statistically significant. Despite multiple comparisons, an unadjusted P value of .05 was considered statistically significant, as this study aims to provide preliminary data on this topic. All analyses were conducted using Stata 15.0 (StataCorp).

Messages With the Highest Mean Audience Engagement Per Platform

Message ranking is based on mean CTR and wCTR, calculated by the message within each social media platform. The 5 highest ranking messages are reported descriptively for each platform because of the sparseness of the data.

Differences in Audience Engagement With Organic Messages and Paid Advertisements

We also examined the effect of organic, unpaid messages versus paid advertisements on the CTR, independent of the social media platform. The data were too sparse to examine the independent effects on wCTR, so we present unadjusted values. As described previously, we used multilevel mixed effects negative binomial regression to model the relationship between the CTR and wCTR and the type of message (organic message vs paid advertisement). In the CTR model only, we included the covariate social media type. Predictive margins were calculated, and differences were tested between groups as described earlier.

Statistical Power

Statistical power was calculated to evaluate the message effectiveness via evaluation of the CTR across the selected social media platforms (Twitter, Facebook, and Instagram). We calculated the sample size using a mixed effects negative binomial model. Data from a web-based smoking cessation study through Facebook (FB) suggested a CTR^{FB} of 0.18% [47]. A publication reporting data from an academic health care Twitter (T) account noted a CTR^{T} of 2.01% [48]. Owing to the wide range of reported CTRs and the paucity of the available

XSL•FO
RenderX

data, we considered it prudent to use a slightly more conservative estimate for CTR^{T} of 1.50%. There is currently no reported literature providing data from which to calculate the $CTR^{Instagram}$ for Instagram. On the basis of the estimates provided, as well as assuming 80% power, an α level of .05, an exposure time of 1.0, and a negative binomial dispersion of 1.0, we obtained an estimate of 1176 messages required. One of the goals of this study is to estimate the effect size for future, well-powered studies. Power was calculated using PASS software, version 14 (NCSS, LLC).

Exclusion of Data

Paid messages required approval on all three platforms. The number of days we ran each advertisement was limited to 3 days. However, some paid messages received zero impressions because of the delay in approval of up to 3 days. We excluded these missing values generated for the CTR (n=148) and wCTR (n=524) because of impressions or sessions (the denominator) equaling 0, making it impossible to calculate CTR or wCTR, respectively. To evaluate the influence of the message theme, we further excluded observations (CTR: n=29; wCTR: n=21), where the message theme (ie, health, appearance, money, or family) was unclear. Finally, we excluded observations for the CTR, where the number of impressions was completely missing (n=6). The remaining number of messages for analysis was 1062 for the CTR and 700 for the wCTR.

Results

CTR by Social Media Platform

This health communication experiment included a total of 1275 antismoking health messages that were distributed across three social media sites: Twitter, Facebook, and Instagram. All comparisons between the types of social media used in this experiment showed CTRs that differed significantly from one social media platform to the other (*P*<.001 for all; Table 1). More specifically, the CTR for Instagram was the highest, followed by Facebook and Twitter (Multimedia Appendix 3).

Table 1. Comparisons of CTRs for the analysis of 1275 antismoking health messages that were posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.

Social media type	Total clicks	Total impressions	CTR[a] (95% CI)	*P* values[b] for the comparisons of the CTRs among the three social media types		
				Facebook	Instagram	Twitter
Facebook	510	504	0.0026 (0.0022-0.0030)	N/A[c]	<.001	<.001
Instagram	255	251	0.0037 (0.0032-0.0042)	N/A	N/A	<.001
Twitter	510	484	0.0015 (0.0013-0.0017)	N/A	N/A	N/A

[a]CTR: click-through rate.

[b]*P* values were obtained using multilevel mixed effects negative binomial regression.

[c]N/A: not applicable.

Website CTR by Social Media Type

Pairwise comparisons of the social media types differed with respect to the wCTR (*P*<.001 for all; Table 2). The wCTR for Twitter was the highest, followed by Facebook and Instagram (Multimedia Appendix 3).

Table 2. Comparison of website CTR by social media type for analysis of 1275 antismoking health messages that were posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.

Social media type	CTR[a] (95% CI)	*P* values[b] for the comparisons of the CTRs among the three social media types		
		Facebook	Instagram	Twitter
Facebook	0.2213 (0.1932-0.2495)	N/A[c]	<.001	<.001
Instagram	0.0334 (0.0230-0.0438)	N/A	N/A	<.001
Twitter	0.6308 (0.5640-0.6975)	N/A	N/A	N/A

[a]CTR: click-through rate.

[b]*P* values were obtained using multilevel mixed effects negative binomial regression.

[c]N/A: not applicable.

Independent Effects of Paid Advertisements Versus Organic Messages

After adjusting for social media type, we found that the type of post, organic (unpaid) or paid, was statistically significantly associated with the CTR (Table 3). Specifically, organic messages had a higher CTR than paid advertisements. In addition, the adjusted CTRs for social media types were also significantly different from one another (*P*<.001 for all), with Instagram having the highest CTR, followed by Facebook and Twitter.

We also found that the type of post, organic (unpaid), or paid, was statistically significantly associated with wCTR (Multimedia Appendix 4). However, in contrast to the message

XSL•FO
RenderX

engagement (CTR), organic messages had a lower wCTR compared with paid advertisements (*P*<.001). We were not able to adjust the wCTRs for social media types because the data became sparse in some categories, and the statistical model would not converge.

Table 3. Effects of message type on click-through rate adjusted for social media type for the analysis of 1275 antismoking health messages that were posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.

Variable	n (%)	Click-through rate[a] (95% CI; n=1062)	Comparison	*P* value[b]
Advertisement type				
Organic	346 (27.1)	0.0074 (0.0047-0.0100)	Organic versus paid	<.001
Paid	716 (56.2)	0.0022 (0.0017-0.0027)	Organic versus paid	<.001
Social media type				
Facebook	346 (27.1)	0.0043 (0.0030-0.0057)	Instagram versus Facebook	<.001
Instagram	245 (19.2)	0.0064 (0.0045-0.0084)	Instagram versus Twitter	<.001
Twitter	471 (36.9)	0.0022 (0.0016-0.0028)	Twitter versus Facebook	<.001

[a]All estimates are mutually adjusted.

[b]*P* values were obtained using multilevel mixed effects negative binomial regression, followed by the calculation of marginal means.

Effects of Message and Image Themes on CTR and wCTR

All messages were antismoking messages. However, they referred to three different themes: (1) health or health and community or health and family, (2) money, and (3) addiction. We did not find an effect of the message theme (love of family [LOF] vs no LOF) on the CTR or wCTR (CTRLOF=0.0024, 95% CI 0.0019-0.0029 versus CTRnoLOF=0.0027, 95% CI 0.0023-0.0030; *P*=.33) or wCTR (wCTRLOF=0.1451, 95% CI 0.1039-0.1863 vs wCTRnoLOF=0.1746, 95% CI 0.1549-0.1942; *P*=.20). We further investigated whether an interaction existed between social media type and image theme for both CTR and wCTR but found no statistically significant interaction with either CTR (*P*=.48) or wCTR (*P*=.21).

Messages With Highest CTR

Of the 1275 distributed messages, the message with the highest CTR was “Smoking can destroy the tiny hairs that help keep the lungs clear, giving a person a smoker’s cough” on Facebook. However, Facebook also had some of the lowest CTRs, resulting in a lower mean overall CTR (Table 4). Only the health message “Polonium-210 is a chemical in #cigarette smoke. It’s also found in nuclear reactors” was found to have a high mean CTR on more than one platform.

XSL•FO
RenderX

Table 4. Top 5 messages by social media platform that showed the highest click-through rates for analysis of 1275 antismoking health messages that were posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.

Platform and health message	Theme	Mean click-through rate	Number of clicks	Number of impressions[a]
Facebook				
Smoking can destroy the tiny hairs that help keep the lungs clear, giving a person a smoker's cough.	Health	0.201	2	1919
How does #smoking take a decade of life away? Smokers die about 12% earlier than nonsmokers.	Health	0.007	10	1595
Nicotine can change the way a person's brain works, causing them to crave more and more nicotine.	Addiction	0.006	8	1599
Smoking half a pack per day costs about $1000/year. Smoking can do serious damage to your wallet.	Money	0.006	12	1942
Tobacco use causes 1300 US deaths daily-more than AIDS, alcohol, car accidents, homicides & illegal drugs combined.	Health	0.006	5	934
Instagram				
Don't let #cigs cut your life short. Smoking #cigarettes can claim more than 10 years of your life.	Health	0.023	15	622
About 20% of all US deaths are caused by a #smoking-related disease. Forget death, chase life.	Health	0.020	14	707
Smoking can cause cancer almost anywhere in the body. 160,000+ US cancer deaths every year are linked to #smoking.	Health	0.014	19	3573
Smoking #cigarettes can claim more than 10 years of your life. Don't let #cigs cut your life short.	Health	0.014	19	2869
Polonium-210 is a chemical in #cigarette smoke. It's also found in nuclear reactors.	Health	0.013	12	1292
Twitter				
480,000 US deaths are caused by a #smoking-related disease every year. Forget death, chase life.	Health	0.035	10	5121
Polonium-210 is a chemical in #cigarette smoke. It's also found in nuclear reactors.	Health	0.032	9	851
About 40% of nonsmokers in this country are exposed to toxic secondhand smoke.	Health	0.014	12	7434
Over 100 million nonsmokers in this country are exposed to toxic secondhand smoke.	Health	0.014	6	4658
Over 160,000 cancer deaths in the US every year are linked to #smoking.	Health	0.014	9	10000

[a]Impressions: number of times a post or advertisement is displayed, whether or not the post is clicked.

Overall, Twitter had the highest wCTRs, followed by Facebook (Table 5). A variety of themes were represented in the messages that received website clicks and were thus examined by the wCTR, although the predominant themes remained to be health-related, including appearance and addiction (12/15, 80%) as opposed to money or family. One health message had a high mean wCTR on both Twitter and Facebook, "Smokers die about 10 years younger than nonsmokers. When someone dies from #tobacco use, we lose them too soon," suggesting potential resonance with a cross-platform population.

XSL•FO
RenderX

Table 5. Top 5 messages by social media type that showed the highest website click-through rate for the analysis of 1275 antismoking health messages that were posted across three social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.

Platform and health message	Theme	Mean website click-through rate	Number of clicks	Number of sessions[a]
Facebook				
Smoking can shorten your life by 10 years. If you smoke, you may be cutting your time with the fam short	Health and family	0.750	4	7
Smokers die about 10 years younger than nonsmokers. When someone dies from #tobacco use, we lose them too soon	Health and community	0.578	8	10
On average, every cig reduces your life by 11 minutes. Even occasional #smoking can hurt you	Health	0.515	15	17
#Smoking can damage our wallets. Smoking half a pack per day costs $1000 per year on average	Money	0.512	10	15
Polonium-210 is a chemical in nuclear reactors. It's also found in #cigarette smoke	Health	0.500	2	4
Instagram				
How does #smoking take a decade of life away? Smokers die about 12% earlier than nonsmokers	Health	0.500	1	6
In the US, 480,000 deaths are caused by a #smoking-related disease every year	Health	0.383	9	13
Nicotine can change the way your brain works, causing you to crave more and more nicotine	Addiction	0.357	13	14
Over 100 million nonsmokers in this country are exposed to toxic secondhand smoke	Health	0.333	14	14
#Smoking can weaken the immune system, leaving a person more vulnerable to bronchitis & pneumonia	Health	0.200	8	17
Twitter				
There is no safe level of exposure to secondhand smoke. Even a short time can harm people's health	Health	1.0	4	2
There is no safe level of exposure to secondhand smoke. Even a short time can harm our health	Health	1.0	7	2
Teens underestimate how addictive #cigarettes are. 3 out of 4 teen smokers become adult smokers	Addiction	1.0	7	2
Smoking can cause cancer almost anywhere in the body. 160,000+ US cancer deaths every year are linked to #smoking	Health	1.0	6	2
Smokers die about 10 years younger than nonsmokers. When someone dies from #tobacco use, we lose them too soon	Health and community	1.0	4	2

[a]Sessions: a session is defined as a group of interactions that a user takes within a time frame of 30 minutes on a website.

Discussion

Principal Findings

This study tested 1275 antitobacco public health messages, targeting English-speaking users in the United States on three popular social media platforms (Twitter, Facebook, and Instagram). The results demonstrate that the same public health message received different levels of engagement from social media users, depending on the platform. Instagram, a platform that focuses on helping users to share photos and video content, showed the statistically significant highest overall mean CTR compared with Facebook and Twitter. However, for any individual message, Facebook resulted in the highest and lowest CTRs, indicating that Facebook might generate the highest level of user engagement for an individual message while also posing the highest risk of underperforming messages.

We also assessed whether users visited the educational website and engaged with its content (ie, scrolling down to peruse the content on the landing page). The findings show that engagement with a health message on social media (ie, measured as CTR in this study) does not indicate user engagement on the website. Health messages on Twitter showed low CTRs, but they resulted in the highest level of website engagement (wCTR), followed by Facebook and Instagram. Therefore, it is recommended that both metrics (CTR and wCTR) should be taken into account when designing health promotion strategies.

The difference in user engagement, which we found for paid, advertised versus organic messages, was surprising. Paid advertisements on social media allow the targeting of special

XSL•FO
RenderX

group characteristics, such as age, gender, language, interests, and location. At the same time, organic messages rely more heavily on developing and engaging a community of followers. Our data suggest that unpaid, organic messages deserve equal engagement when designing health interventions. In this study, organic messages showed significantly higher CTRs than paid advertisements, suggesting the importance of combining organic and advertising messages in health promotion campaigns. Users might distrust advertisements as messages that are designed to manipulate them into buying something. However, we believe this is less likely, as the opposite seems to be the trend. Marketers and publishers are increasingly using forms of *native advertising*, that is, content that bears a similarity to the news, feature articles, product reviews, entertainment, and other materials that surround it on the web. In 2015, before this study was conducted, the Federal Trade Commission Act prohibited deceptive or unfair practices. It issued an Enforcement Policy Statement on Deceptively Formatted Advertisements that explains how the agency applies established truth-in-advertising standards in this context [49]. At the same time, the increased emphasis on advertising transparency on social media platforms was initiated after this study was conducted. Twitter, for example, updated its Political Content Policy on November 22, 2019 [50]. Future research needs to examine the extent to which this affects the advertising of public health messages.

Application of Theory

We kept the message content, including the image and distribution modus (ie, date of a message, time of day distributed) consistent across the three social media sites. Hence, we discuss the results under the assumption that users' probability of being exposed to the health messages did not differ across social media sites. This allows us to discuss our findings in light of potential platform-specific factors that may have contributed to the difference in user engagement with the messages: first, user factors that may affect their selected exposure to content, and second, technical site features.

The SESAM model posits that user motivations for selectively exposing themselves to media content vary situationally [37]. This is supported by research showing that users' motivation to use a specific social media site may vary, which may, in turn, be associated with different levels of engagement with health promotion messages. The Pew Research Center, for example, reported that Twitter is one of the social media sites with the most news-focused users [51]. According to the Pew Research Centre, "around seven-in-ten American adult Twitter users (71%) get news on the site" [51]. Thus, Twitter users may be more focused on their use of the platform and click more exclusively on content that serves this particular motivation. This could explain why, despite a lower CTR, the Twitter messages in this study resulted in the highest website engagement level. However, Facebook is a site where more than half of the users are exposed to news-related content (67%) [52]. These findings indicate that news-like posts may perform better on these platforms. Future research could test the hypothesis of whether health messages that apply characteristics of news are more effective in generating message engagement on these platforms.

On the other hand, Instagram has less of a news focus among its user base [52]. Users' primary motives for using the site have been described, for example, surveillance and knowledge about others, documentation, coolness, creativity, and escapism [53,54]. Thus, users may have clicked on the content we posted for these motivations (ie, curiosity or to find out what the content was about), but not been motivated to click further through to the website because the initial motivation (surveillance) was satisfied by the initial click. The different kinds of user intent could explain the higher CTR but lower engagement with informative content. Our data suggest that the nature of intent may play a role in engaging with subsequent educational content promoted. Health promotion messages on Instagram could be less focused on referring users to a website. Instead, they deliver instant access to the information on Instagram itself. If future research demonstrates the effectiveness of this approach, organizations could consider using Instagram posts that immediately convey the relevant information to users, rather than using posts that require users to connect through to a website to obtain the relevant information. In addition, organizations could consider posting Instagram content that is visually compelling. All three social media sites allow users to include visual media (ie, images, videos) in their posts, but Instagram stands out as a platform that emphasizes visual content over text [55]. The emphasis on visual content on Instagram might have affected the users' propensity to engage with the health messages. Research has shown that digital forms of media monopolize individuals' engagement and attention spans, using visual strategies that demand our interactions [56].

Second, specific platform features might have contributed to the difference in user message engagement and behavior. Bucher and Helmond [39] introduced the *concept of affordance* of social media platforms to understand and analyze social media interfaces and the relations between technology and its users. They suggested that social media users may or may not *perceive or attend to the affordance* of a social media site according to their needs but also to the affordances (ie, technical features) of the social media site—in this case, engage with the message and click on its link. However, the technical features of the three social media sites used in this study are similar, allowing users to interact with content through *likes*, shares (retweets), and comments (replies) to a message. The success of applications such as Facebook relies on the simplicity and immersive design of their interface. Internet platforms are designed to capture viewers' engagement [56]. We argue that this applies to the three platforms used here. Nonetheless, there may still be nuances in site design that facilitate clicks, for example, Instagram's focus on visual imagery [57] may contribute to a higher CTR.

Study Limitations

This pilot study has multiple limitations. Owing to the limited advertisement budget, we had to limit the study to 85 days, which limited the amount of data we could gather and analyze. The analysis is further limited to the digital data that we could access, not including potential behavior change and other real-world implications. The data itself are in aggregate, which do not allow the analysis of demographic or behavioral factors that might have influenced the outcomes assessed here. To

XSL•FO
RenderX

increase the number of message variations, we used linguistic variations, for example, using *we* versus *you* versus *they*. However, for some variables, data were too sparse to adjust for all covariates of interest. The study was also underpowered to detect significant interactions of interest.

Furthermore, the generalizability of this study is limited. The messages we tested were antitobacco messages with nuanced messaging focused on either the health effect of smoking, the cost of tobacco products, or the negative impact on family members such as siblings. Health education campaigns targeting other health behaviors might show different results.

Finally, in this study, it is difficult to discern the effect of the message text from the message image's impact on the CTR and wCTR. Although our data do not suggest that the image theme influenced the CTR or wCTR, other research has demonstrated imagery as a strong predictor of user engagement [58]. Rus and Cameron [58] showed that messages with images had higher rates of liking and sharing relative to messages without images on Facebook. The formats and demographics of social media sites are constantly changing and have evolved since the study was conducted.

Conclusions

Our study provides evidence-based insights to guide the design of health promotion efforts on social media. Using the full potential of social media for health promotion efforts will require a deeper understanding of the factors that drive user message engagement across different social media to ultimately support informed health decisions and positive behavior change.

Future studies should examine the platform-specific impact of psycholinguistic message variations on user engagement and include newer sites such as Snapchat and TikTok. We suggest a focus on more rigorous studies and a move toward *evidence-based science communication* [59] and study the correlation between web-based behavior and real-world health behavior change. The need is urgent in light of increased health-related marketing and misinformation on social media, as evidenced most recently through the COVID-19 infodemic [60], a global epidemic of misinformation spreading rapidly through social media and other outlets that poses a serious problem for public health.

Acknowledgments

The authors would like to acknowledge and thank Wendy Mack, PhD, for her helpful advice regarding data analysis and modeling. This work was supported by the National Institutes of Health (NIH) National Cancer Institute (NCI), the Food and Drug Administration (FDA) Center for Tobacco Products (Tobacco Center of Regulatory Science grant #5P50CA180905), and the Southern California Clinical and Translational Science Institute (grant #UL1TR000130) from the National Center for Advancing Translational Sciences (NCATS) of the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCI, FDA, and NIH.

Conflicts of Interest

MM serves as a paid expert witness in litigation sponsored by the Public Health Advocacy Institute against RJ Reynolds. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies.

Multimedia Appendix 1
List of parameterized message templates used in the experiment. Examples of messages with images for each platform can be found in the technical implementation paper of the experiment mentioned above.
[DOCX File , 22 KB - publichealth_v7i2e24429_app1.docx]

Multimedia Appendix 2
Targeting criteria for paid messages (advertisements) on Twitter, Facebook, and Instagram.
[DOCX File , 10 KB - publichealth_v7i2e24429_app2.docx]

Multimedia Appendix 3
Message impressions, message clicks, website sessions, and website clicks per message.
[XLSX File (Microsoft Excel File), 132 KB - publichealth_v7i2e24429_app3.xlsx]

Multimedia Appendix 4
Effects of message type (paid advertisements vs organic messages) on website click-through rate for the analysis of 1275 antismoking health messages that were posted across 3 social media platforms (Twitter, Facebook, and Instagram) between April 19 and July 12, 2017.
[PDF File (Adobe PDF File), 13 KB - publichealth_v7i2e24429_app4.pdf]

References

XSL•FO
RenderX

1. Obar JA, Wildman S. Social media definition and the governance challenge: an introduction to the special issue. Telecommunications Policy 2015 Oct;39(9):745-750. [doi: 10.1016/j.telpol.2015.07.014]
2. Demographics of social media users and adoption in the United States internet. 2019. URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-06-28]
3. Teens, Social Media & Technology 2018 Internet. 2018. URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2019-10-18]
4. Teens? Social Media Habits and Experiences Internet. 2018. URL: https://www.pewinternet.org/2018/11/28/teens-social-media-habits-and-experiences/ [accessed 2019-10-20]
5. Gabarron E, Årsand E, Wynn R. Social media use in interventions for diabetes: rapid evidence-based review. J Med Internet Res 2018 Aug 10;20(8):10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
6. Strømme H, Mosdøl A, Nordheim LW, Vist GE. Effects of using social media in health awareness campaigns to promote healthy lifestyle habits. In: Knowledge Centre for the Health Services. Oslo, Norway: The Norwegian Institute of Public Health (NIPH); 2018:9788281219151.
7. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
8. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
9. Kearney MD, Selvan P, Hauer MK, Leader AE, Massey PM. Characterizing HPV vaccine sentiments and content on instagram. Health Educ Behav 2019 Dec;46(2_suppl):37-48. [doi: 10.1177/1090198119859412] [Medline: 31742459]
10. Jenkins MC, Moreno MA. Vaccination discussion among parents on social media: a content analysis of comments on parenting blogs. J Health Commun 2020 Mar 03;25(3):232-242. [doi: 10.1080/10810730.2020.1737761] [Medline: 32154770]
11. 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]
12. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One 2016;11(7):0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]
13. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep;26(5):575-578 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
14. 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):7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
15. Allem J, Ramanujam J, Lerman K, Chu K, Boley Cruz T, Unger JB. Identifying sentiment of Hookah-related posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
16. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:26-30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
17. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. JMIR Public Health Surveill 2016 Dec 12;2(2):171 [FREE Full text] [doi: 10.2196/publichealth.6551] [Medline: 27956376]
18. Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising awareness about cervical cancer using twitter: content analysis of the 2015 #smearforsmear campaign. J Med Internet Res 2017 Oct 16;19(10):344 [FREE Full text] [doi: 10.2196/jmir.8421] [Medline: 29038096]
19. Klippert H, Schaper A. Using Facebook to communicate mammography messages to rural audiences. Public Health Nurs 2019 Mar;36(2):164-171. [doi: 10.1111/phn.12556] [Medline: 30370687]
20. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014 Oct 16;16(10):238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
21. 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):9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]
22. Elaheebocus SMRA, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018 Feb 22;20(2):20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
23. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh 2013 Sep;45(3):256-264. [doi: 10.1111/jnu.12032] [Medline: 23676115]

XSL•FO
RenderX

24. Carson KV, Ameer F, Sayehmiri K, Hnin K, van Agteren JE, Sayehmiri F, et al. Mass media interventions for preventing smoking in young people. Cochrane Database Syst Rev 2017 Jun 02;6:CD001006 [FREE Full text] [doi: 10.1002/14651858.CD001006.pub3] [Medline: 28574573]
25. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar 27;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
26. Strekalova YA, Damiani RE. Message design and audience engagement with tobacco prevention posts on social media. J Cancer Educ 2018 Jun 10;33(3):668-672. [doi: 10.1007/s13187-016-1135-x] [Medline: 27832508]
27. Paine Publishing LLC. The conclave on social media measurement standards. Engagement & Conversation. 2021. URL: https://smmstandards.wixsite.com/smmstandards/engagement-and-conversation [accessed 2020-12-22]
28. Carah N, Angus D. Algorithmic brand culture: participatory labour, machine learning and branding on social media. In: Media, Culture & Society. Thousand Oaks, California, United States: SAGE Publications Ltd; Jan 22, 2018:178-194.
29. Bruns A, Mahrt M, Weller K, Burgess J, Puschmann C. Twitter and Society. In: Series: Digital Formations. Bern, Switzerland: Peter Lang Publishing; 2014:15.
30. Gold RS, Auld ME, Abroms LC, Smyser J, Yom-Tov E, Allegrante JP. Digital Health Communication Common Agenda 2.0: An Updated Consensus for the Public and Private Sectors to Advance Public Health. Health Educ Behav 2019 Dec;46(2_suppl):124-128. [doi: 10.1177/1090198119874086] [Medline: 31742457]
31. Gatewood J, Monks S, Singletary C, Vidrascu E, Moore J. Social media in public health: strategies to distill, package, and disseminate public health research. J Public Health Manag Pract Internet 2019 Oct 22. [doi: 10.1097/phh.0000000000001096]
32. Haardörfer R. Taking quantitative data analysis out of the positivist era: calling for theory-driven data-informed analysis. Health Educ Behav 2019 Aug;46(4):537-540. [doi: 10.1177/1090198119853536] [Medline: 31216879]
33. Knoll J. Advertising in social media: a review of empirical evidence. In: Int J of Advert. Oxfordshire, United Kingdom: Taylor & Francis; Mar 24, 2015:266-300.
34. Google.com. Clickthrough rate (CTR): Definition. Google.com 2021 [FREE Full text]
35. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
36. Reuter K, MacLennan A, Le N, Unger JB, Kaiser EM, Angyan P. A software tool aimed at automating the generation, distribution, and assessment of social media messages for health promotion and education research. JMIR Public Health Surveill 2019 May 07;5(2):11263 [FREE Full text] [doi: 10.2196/11263] [Medline: 31066708]
37. Knobloch-Westerwick S. The Selective Exposure Self- and Affect-Management (SESAM) Model. In: Communication Research. California, United States: SAGE Publications Inc; Jun 18, 2014:959-985.
38. Affordance theory in social media research: systematic review and synthesis of the literature internet. 25th European Conference on Information Systems (ECIS 2017). 2017. URL: https://papers.ssrn.com/abstract=2988394 [accessed 2021-01-22]
39. Moreno MA, D'Angelo J. Social media intervention design: applying an affordances framework. J Med Internet Res 2019 Mar 26;21(3):11014 [FREE Full text] [doi: 10.2196/11014] [Medline: 30912754]
40. The affordances of social media platforms. BucherHelmond_SocialMediaAffordances-preprint. URL: https://www.annehelmond.nl/wordpress/wp-content/uploads//2016/07/BucherHelmond_SocialMediaAffordances-preprint.pdf [accessed 2021-01-22]
41. Voorveld H, van Noort G, Muntinga D, Bronner F. Engagement with social media and social media advertising: the differentiating role of platform type. J of Advert 2018 Feb 13;47(1):38-54. [doi: 10.1080/00913367.2017.1405754]
42. Zhao Y, Liu J, Tang J, Zhu Q. Conceptualizing perceived affordances in social media interaction design. AP 2013 Mar;65(3):289-303. [doi: 10.1108/00012531311330656]
43. Edney S, Looyestyn J, Ryan J, Kernot J, Maher C. Posts, pics, or polls? Which post type generates the greatest engagement in a Facebook physical activity intervention? Transl Behav Med 2018 Nov 21;8(6):953-957. [doi: 10.1093/tbm/iby006] [Medline: 29660093]
44. The best times to post on social media in 2018 based on research internet. CoSchedule Blog. URL: https://coschedule.com/blog/best-times-to-post-on-social-media/ [accessed 2019-04-30]
45. Spark S. How often to post to Twitter, Facebook, Instagram, and Pinterest | Adobe Spark Internet. Adobe Spark Blog. URL: https://blog.adobespark.com/2016/05/06/how-often-to-post-to-twitter-facebook-instagram-and-pinterest/ [accessed 2019-04-30]
46. Abokhodair N, Yoo D, McDonald D. Dissecting a social botnet: growth, content and influence in Twitter internet. arXiv csCY. URL: http://arxiv.org/abs/1604.03627 [accessed 2021-01-22]
47. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
48. Widmer RJ, Engler NB, Geske JB, Klarich KW, Timimi FK. An academic healthcare Twitter account: The Mayo Clinic experience. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):360-366. [doi: 10.1089/cyber.2015.0607] [Medline: 27327062]

XSL•FO
RenderX

49. Native Advertising: A Guide for Businesses Internet. 2015. URL: https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses [accessed 2021-01-22]
50. Ads transparency center. Internet. 2020 Sep 9. URL: https://ads.twitter.com/transparency [accessed 2020-01-22]
51. 10 facts about Americans and Twitter internet. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/ [accessed 2021-01-22]
52. News use across social media platforms 2018 internet. Pew Research Center's Journalism Project. 2018. URL: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [accessed 2021-01-22]
53. Sheldon P, Bryant K. Instagram: motives for its use and relationship to narcissism and contextual age. Computers in Human Behavior 2016 May;58:89-97. [doi: 10.1016/j.chb.2015.12.059]
54. Lee E, Lee J, Moon JH, Sung Y. Pictures speak louder than words: motivations for using Instagram. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):552-556. [doi: 10.1089/cyber.2015.0157] [Medline: 26348817]
55. Lee AS, Hart JL, Sears CG, Walker KL, Siu A, Smith C. A picture is worth a thousand words: Electronic cigarette content on Instagram and Pinterest. Tob Prev Cessat 2017 Jul;3 [FREE Full text] [doi: 10.18332/tpc/74709] [Medline: 28815224]
56. Liu K. The attention crisis of digital interfaces and how to consume media more mindfully internet. Claremont Colleges. 2019. URL: https://scholarship.claremont.edu/scripps_theses/1251 [accessed 2019-10-01]
57. Shane-Simpson C, Manago A, Gaggi N, Gillespie-Lynch K. Why do college students prefer Facebook, Twitter, or Instagram? Site affordances, tensions between privacy and self-expression, and implications for social capital. Computers in Human Behavior 2018 Sep;86:276-288. [doi: 10.1016/j.chb.2018.04.041]
58. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related facebook pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
59. Jensen EA, Gerber A. Evidence-based science communication. Front Commun 2020 Jan 23;4. [doi: 10.3389/fcomm.2019.00078]
60. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]

Abbreviations

CTR: click-through rate
FB: Facebook
FDA: Food and Drug Administration
IG: Instagram
NIH: National Institutes of Health
SESAM: Selective Exposure Self- and Affect-Management
T: Twitter
wCTR: website click-through rate

Edited by T Sanchez; submitted 19.09.20; peer-reviewed by CY Lin, X Xia, A Johnson; comments to author 03.12.20; revised version received 29.12.20; accepted 06.01.21; published 19.02.21.

Please cite as:
Reuter K, Wilson ML, Moran M, Le N, Angyan P, Majmundar A, Kaiser EM, Unger JB
General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis
JMIR Public Health Surveill 2021;7(2):e24429
URL: http://publichealth.jmir.org/2021/2/e24429/
doi:10.2196/24429
PMID:33605890

©Katja Reuter, Melissa L Wilson, Meghan Moran, NamQuyen Le, Praveen Angyan, Anuja Majmundar, Elsi M Kaiser, Jennifer B Unger. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 19.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Key Population Hotspots in Nigeria for Targeted HIV Program Planning: Mapping, Validation, and Reconciliation

Julia Lo[1,2], MSPH; Samuel U Nwafor[3], MPH; Amee M Schwitters[4], MPH, PhD; Andrew Mitchell[1], BS; Victor Sebastian[4], MBBS, MPH, FMCPH; Kristen A Stafford[1,2,5], MPH, PhD; Idoteyin Ezirim[6], MPH; Man Charurat[1,2], PhD; Anne F McIntyre[7], MPH, PhD

[1]Institute of Human Virology, University of Maryland School of Medicine, Baltimore, MD, United States

[2]Center for International Health, Education, and Biosecurity, University of Maryland School of Medicine, Baltimore, MD, United States

[3]Center for International Health, Education, and Biosecurity, University of Maryland School of Medicine, Abuja, Nigeria

[4]Division of Global HIV & TB, Center for Global Health, Centers for Disease Control and Prevention, Abuja, Nigeria

[5]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, United States

[6]National Agency for the Control of AIDS, Abuja, Nigeria

[7]Division of Global HIV & TB, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA, United States

Corresponding Author:
Anne F McIntyre, MPH, PhD
Division of Global HIV & TB
Center for Global Health
Centers for Disease Control and Prevention
1600 Clifton Road NE MS E-30
Atlanta, GA, 30333
United States
Phone: 1 404 433 3601
Email: zat4@cdc.gov

Abstract

Background: With the fourth highest HIV burden globally, Nigeria is characterized as having a mixed HIV epidemic with high HIV prevalence among key populations, including female sex workers, men who have sex with men, and people who inject drugs. Reliable and accurate mapping of key population hotspots is necessary for strategic placement of services and allocation of limited resources for targeted interventions.

Objective: We aimed to map and develop a profile for the hotspots of female sex workers, men who have sex with men, and people who inject drugs in 7 states of Nigeria to inform HIV prevention and service programs and in preparation for a multiple-source capture-recapture population size estimation effort.

Methods: In August 2018, 261 trained data collectors from 36 key population–led community-based organizations mapped, validated, and profiled hotspots identified during the formative assessment in 7 priority states in Nigeria designated by the United States President's Emergency Plan for AIDS Relief. Hotspots were defined as physical venues wherein key population members frequent to socialize, seek clients, or engage in key population–defining behaviors. Hotspots were visited by data collectors, and each hotspot's name, local government area, address, type, geographic coordinates, peak times of activity, and estimated number of key population members was recorded. The number of key population hotspots per local government area was tabulated from the final list of hotspots.

Results: A total of 13,899 key population hotspots were identified and mapped in the 7 states, that is, 1297 in Akwa Ibom, 1714 in Benue, 2666 in Cross River, 2974 in Lagos, 1550 in Nasarawa, 2494 in Rivers, and 1204 in Federal Capital Territory. The most common hotspots were those frequented by female sex workers (9593/13,899, 69.0%), followed by people who inject drugs (2729/13,899, 19.6%) and men who have sex with men (1577/13,899, 11.3%). Although hotspots were identified in all local government areas visited, more hotspots were found in metropolitan local government areas and state capitals.

Conclusions: The number of key population hotspots identified in this study is more than that previously reported in similar studies in Nigeria. Close collaboration with key population–led community-based organizations facilitated identification of many new and previously undocumented key population hotspots in the 7 states. The smaller number of hotspots of men who have sex with men than that of female sex workers and that of people who inject drugs may reflect the social pressure and stigma faced

XSL•FO
RenderX

by this population since the enforcement of the 2014 Same Sex Marriage (Prohibition) Act, which prohibits engaging in intimate same-sex relationships, organizing meetings of gays, or patronizing gay businesses.

(JMIR Public Health Surveill 2021;7(2):e25623) doi:10.2196/25623

KEYWORDS

key population; female sex workers; men who have sex with men; people who inject drugs; hotspot mapping; HIV; Nigeria

Introduction

Key populations include female sex workers, men who have sex with men, and people who inject drugs, and they are particularly vulnerable to HIV. The high incidence and prevalence of HIV in key populations are well documented in the literature [1-7]. Together, key population members and their sexual partners account for 54% of the new HIV infections worldwide, with the risk of HIV acquisition being up to 22-fold higher for female sex workers, men who have sex with men, and people who inject drugs than the risk of HIV acquisition by the general population [8]. Stigma and discrimination, fear of legal prosecution, misinformation, travel time, and transportation costs are some of the barriers to HIV testing and treatment among the general population and key population members alike [9-14].

Nigeria is characterized as having a mixed HIV epidemic with high HIV prevalence among key population members and low prevalence in the general population [15,16]. Recently published results from the Nigeria HIV/AIDS Indicator and Impact Survey revealed an HIV prevalence of 1.4% among men and women in the age range of 15-49 years, which is considerably lower than previous estimates [16,17]. In contrast, the 2014 data from the most recent biobehavioral survey among key populations demonstrated HIV prevalence of 14.4% among female sex workers, 22.9% among men who have sex with men, and 3.4% among people who inject drugs [15]. Key populations and their sexual partners are estimated to account for 32% of the new HIV infections in Nigeria [18]. Thus, targeted interventions designed to serve and reach these individuals become a necessary strategy to control the HIV epidemic. In the National HIV and AIDS Strategic Framework 2017-2021, the Government of the Federal Republic of Nigeria outlined the need for interventions to increase testing and treatment for key populations in order to fast-track the national response toward ending AIDS in Nigeria by 2030 [18,19]. Reliable and accurate information on where key population members socialize is needed for the strategic placement of services and allocation of limited resources for targeted interventions.

It is recommended that key population mapping and population size estimation be conducted every 2-3 years to produce reliable and up-to-date data for HIV program planning [20]. There have been several efforts to map key population hotspots in Nigeria in 2009, 2013, and 2015 [21-23]. Several of these studies had shortcomings, including failure to report the number of hotspots of men who have sex with men identified and limitations in scope, reporting only on male sex workers and only in major cities. The objective of this study was to map the key population hotspots in 7 United States President's Emergency Plan for AIDS Relief (PEPFAR) priority states in Nigeria. These states were chosen in consultation with the Government of the Federal Republic of Nigeria and others due to the evidence of high HIV burden and unmet needs for HIV/AIDS treatment services at the time this study was conducted [24]. This study is the first of a two-part key population mapping and size estimation effort. The scope of this paper is to report on the mapping exercise only. The multiple-source capture-recapture (MS-CRC) population size estimation exercise is a separate report. The final list of the key population hotspots generated from this exercise served as a sampling frame for MS-CRC.

Methods

Study Sites

This study was conducted in 7 PEPFAR priority states of Akwa Ibom, Benue, Cross River, Lagos, Nasarawa, Rivers, and Federal Capital Territory and as part of a larger key population hotspot mapping and size estimation exercise that included a formative assessment and MS-CRC for population size estimation. This exercise was conducted in all 134 local government areas in the 7 states.

Study Procedures

For planning and logistical purposes, each state was divided into 3 smaller geographical areas, referred to as zones. Prior to the start of the field activities, a list of known hotspots of female sex workers, men who have sex with men, and people who inject drugs was compiled from stakeholders, including Nigeria National Agency for Control of AIDS (NACA) and implementing partners who provide services to key population members or were involved in previous key population hotspot mapping exercises (Society for Family Health, 2015) [23]. This list was stratified by study states and zones and reviewed during state-specific and zone-specific focus group discussions and key informant interviews as part of the formative assessment. During the formative assessment, additional key population hotspots were added to the list and those reported to be inactive were marked but not deleted.

In August 2018, 261 trained data collectors in the 7 states mapped, validated, and profiled the hotspots collated at the end of the formative assessment. During mapping, data collectors used tablets to electronically record the hotspot name, local government area, address, type, and geographic coordinates. In addition, data collectors identified 1 key informant in the area to solicit peak times of key population activity, presence of key population behaviors of interest, and an estimate of the minimum and maximum number of key population members found in the hotspot at the time of their visit. Key informants were bartenders, bouncers/security staff, madams/brothel owners, bunk owners/drug peddlers, or taxi drivers, and they displayed familiarity with key population activities in the hotspot. In

XSL•FO
RenderX

addition to validating known hotspots, data collectors identified new hotspots that had emerged since the previous mapping exercise in 2015.

For this study, hotspots were defined as physical locations in which key population members frequent to socialize, seek clients, or engage in key population–defining behaviors. The following definitions of hotspot types were used in this study. Those located in an outdoor area accessible to the public were categorized as street/public place. These included streets, under bridges, organized motor parks, unnamed drinking places, and bus stops. Hostel/campus was defined as an area near student living apartments/hostels of a secondary, polytechnic, or university level academic institution meant for student relaxation, academic meetings, or social gatherings. Hotspots not captured in the categories as described in the protocol were classified as "Other."

A desk review was conducted after data collection in the field to verify the information submitted and to remove duplicates. Hotspots were reviewed by data analysts, and potential duplicates were flagged based on similarity in name, address, geographic coordinates, and other relevant information. State supervisors then met with data collectors and representatives of the key population community to review the list of hotspots and to remove duplicate entries. We utilized this opportunity to further improve on information collected during the hotspot mapping exercise. Corrections on location information such as local government area, address, name of hotspot, and documented geocoordinates were made following suggestions from data collectors during MS-CRC. Information on additional duplicates was also solicited for documentation and flagged but not removed. No additional hotspots were added or removed from the initial list of hotspots at the start of MS-CRC.

Technical Team in This Study

A national technical team was formed to support and inform study design, implementation, monitoring, and dissemination. The national technical team, chaired by NACA, was composed of members from the Nigeria Federal Ministry of Health, the University of Maryland Baltimore, the US Centers for Disease Control and Prevention (CDC) in Nigeria, the National Key Affected Population Network, Centre for the Right to Health, Heartland Alliance International Nigeria, and Population Council. In addition to the national technical team, a state technical team was formed in each of the 7 study states to guide implementation, ensure security, provide close monitoring, and increase buy-in at the state level. Similar to the national technical team, the state technical teams were composed of representatives from the government, key population community, implementing partners who provide services to key population, and University of Maryland Baltimore, in addition to representatives from the Police Action Committee on HIV/AIDS or National Drug Law Enforcement Agency, to help ensure the safety of the data collectors.

Local Partnerships and Recruitment of Data Collectors

In each state, we engaged a minimum of 3 key population–affiliated community-based organizations, one for each key population group, to help implement and inform study planning and logistics. Most data collectors for the exercise were key population members and were recruited through local partner key population community-based organizations. The few data collectors who did not identify as key population members had experience working with key population members and were trusted within the local key population community. While recruiting data collectors for this exercise, the study team considered the data collector's level of influence within the key population community in facilitating hotspot access for the study team, familiarity with key population networks and hotspots, experience in previous surveys, service delivery, or research activities, and considered diversity in age, base of operation (eg, brothel-based or street-based female sex workers), geographical area of familiarity, and other relevant experience. Data collectors were essential in facilitating access to hotspots, and diversity proved to be especially valuable in facilitating access to hotspots frequented by minority key population groups, for example, older female sex workers. Following advice from local partners to prioritize safety and acceptance of the mappers, data collection teams were assigned to map hotspot types that aligned with the key population identity of the team members. For example, hotspots frequented by people who inject drugs were assigned to teams composed of data collectors collecting data on people who inject drugs. A total of 36 key population community-based organizations and 261 data collectors were formally engaged in the 7 states. Most data collectors were retained for the population size estimation component of the study following this hotspot mapping exercise.

Data Management and Quality Assurance

Hotspot information was collected using REDCap (Research Electronic Data Capture) data collection and management software [25,26]. Skip logic and data validation rules were built-in throughout the data collection tool to prevent entry of implausible data. All data collectors were trained in data entry using REDCap, study objectives, standard operating procedures, and the importance of discretion and confidentiality. Data monitoring officials and state supervisors observed the uploaded information to check for missing data fields and inconsistent or implausible values. In addition, members of the national technical team and state technical teams conducted field visits to ensure that data collectors were capturing project data according to study procedures.

Data Analysis and Map Development

The number of key population hotspots per local government area was tabulated from the final list of hotspots, and the number of hotspots per 100,000 population was calculated using 2016 population projections by local government area [27]. The latter was done to allow for comparison of the number of hotspots across local government areas independent of the population size of the local government area. The number of key population hotspots per 100,000 population by local government area was mapped using ArcMap 10.5.1 (Esri). For clarity without crowding the maps, quintiles were used to display differences in the number of hotspots per 100,000 population across the local government areas in the 7 states. State-level distributions of the hotspot types for each key population group were tabulated from the final list of hotspots. Data cleaning and basic

XSL•FO
RenderX

analyses as described above (tabulation, determining mean, median, and mode, calculation of rates, and cutoff points for maps) were performed using Stata 15 (StataCorp LLC).

Research Protection of Human Subjects

This study was approved by the National Health Research Ethics Committee, Nigeria and the Institutional Review Board of University of Maryland Baltimore. This study was reviewed in accordance with the US CDC human research protection procedures and determined to be research, but CDC investigators did not interact with the human subjects or have access to identifiable data for research purposes.

Results

A total of 13,899 key population hotspots were mapped and identified in the 7 states. States with the largest number of hotspots were Lagos, Cross River, and Rivers (Table 1). Although more hotspots were found in urban local government areas and state capitals, at least one was identified in each of the 134 local government areas visited (Figure 1). In Figure 1, the number of key population hotspots presented in the map represents findings at the end of hotspot mapping and validation, immediately before MS-CRC activities began. Hotspots were dynamic with frequent changes in activity status. Hotspot mapping, validation, and data reconciliation were performed between August 2018 and December 2018. The local government areas of key population hotspots were indicated by enumerators, with maps representing the number of key population hotspots per 100,000 population per local government area. Population estimates by local government area [27] were obtained on March 13, 2019. The shape files of Nigerian geographic boundaries [28] were obtained from the Office of the Surveyor General of the Federation, eHealth, United Nations Cartographic Section on February 23, 2017. Among the 134 local government areas, all had at least one female sex worker hotspot, 129 (96.3%) areas had at least one hotspot of people who inject drugs, and 118 (88.1%) areas had at least one hotspot of men who have sex with men. The number of hotspots found in each state and local government area varied by key population group and depended largely on the size of the general population. Compared to hotspots frequented by female sex workers and people who inject drugs, far fewer were found for men who have sex with men (Table 1). The number of hotspots mapped during our 2018 study is summarized with the results from the previous mapping exercises in 2013 and 2015 for comparison of the coverage by key population and state (Table 1).

Table 1. Number of hotspots identified in 7 US President's Emergency Plan for AIDS Relief priority states in Nigeria by key population in 2013, 2015, and 2018.

State	Female sex worker hotspots (n)			Men who have sex with men hotspots (n)			People who inject drugs hotspots (n)			Total hotspots (N)		
	2013[a]	2015[b]	2018	2013	2015	2018	2013	2015	2018	2013	2015	2018
Akwa Ibom	__[c]	150	708	—	—	276	—	89	313	—	239	1297
Benue	825	344	1098	57	—	265	32	117	351	914	461	1714
Cross River	692	497	1782	15	—	268	8	192	616	715	689	2666
Federal Capital Territory	1446	677	977	120	—	116	22	41	111	1588	718	1204
Lagos	4056	2534	2603	191	—	131	95	230	240	4342	2764	2974
Nasarawa	1409	575	990	19	—	246	12	375	314	1440	950	1550
Rivers	—	393	1435	—	—	275	—	141	784	—	534	2494
Total	8428	5170	9593	402	—	1577	169	1185	2729	8999	355	13,899

[a]Source for 2013 data [22].

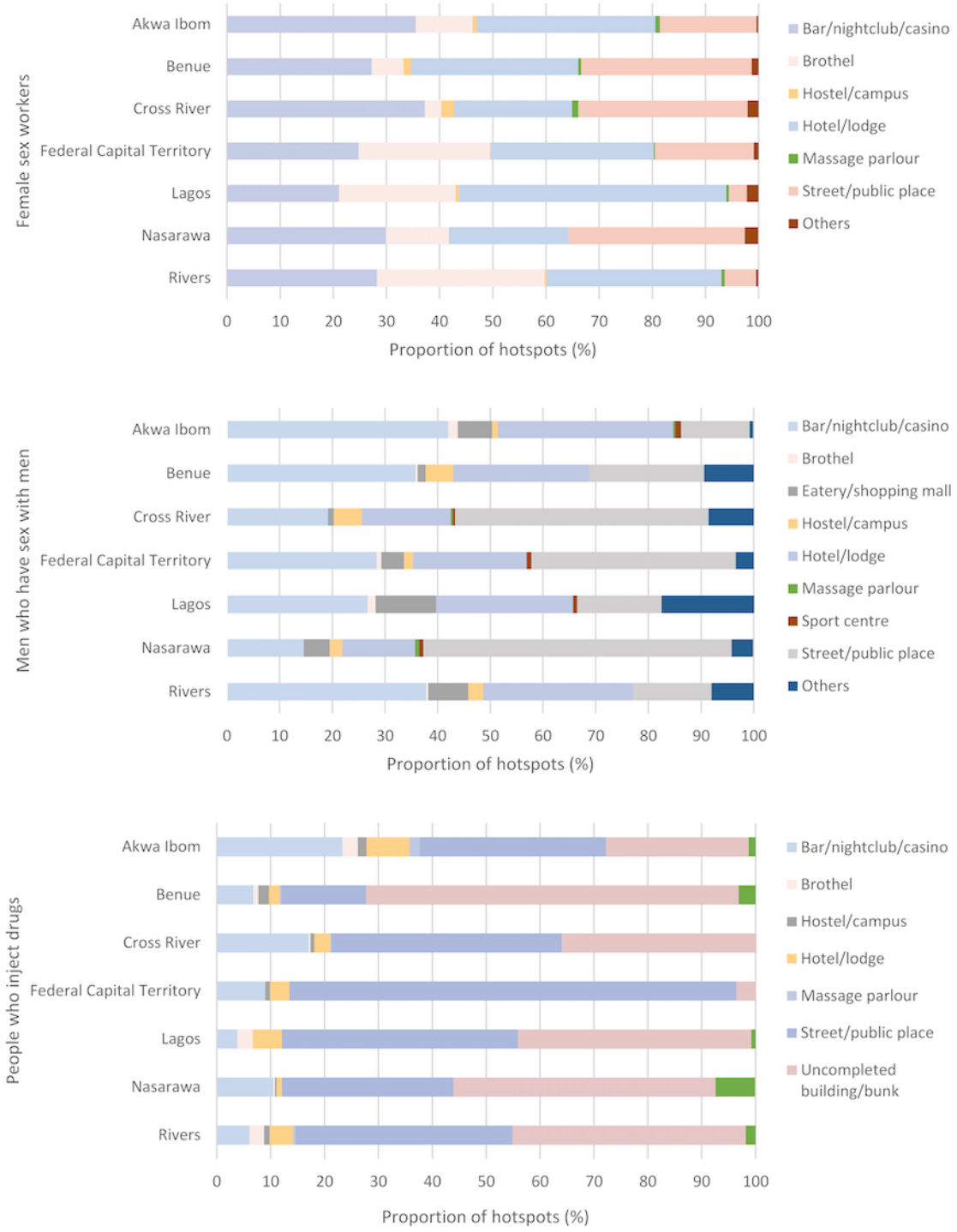
[b]Source for 2015 data [23].

[c]Not available.

Of all the female sex worker hotspots mapped in the 7 states, the majority (9253/9548, 96.9%) of the hotspot types fell into 1 of the following 4 categories: hotel (3266/9548, 34.2%), bar/nightclub/casino (2693/9548, 28.2%), street/public place (18.1%), and brothel (1569/9548, 16.4%). Popular categories of men who have sex with men hotspots in our study included street/public place (473/1575, 30.0%), bar/nightclub/casino (470/1575, 29.8%), and hotel/lodge (376/1575, 23.9%). Of the hotspots of people who inject drugs, 42.1% (1144/2718) mapped in our study were categorized as uncompleted buildings/bunks, 38.1% as street/public place, and 11.1% as bar/nightclub/casino. A small proportion of hotspots mapped in the 7 states were of unknown type and are not reflected in the proportions listed above. These included 0.5% (45/9953) of female sex worker hotspots, 0.1% (2/1577) of men who have sex with men hotspots, and 0.4% (11/2729) of people who inject drugs hotspots. The most popular type of hotspot varied by state and key population group (Figure 2).

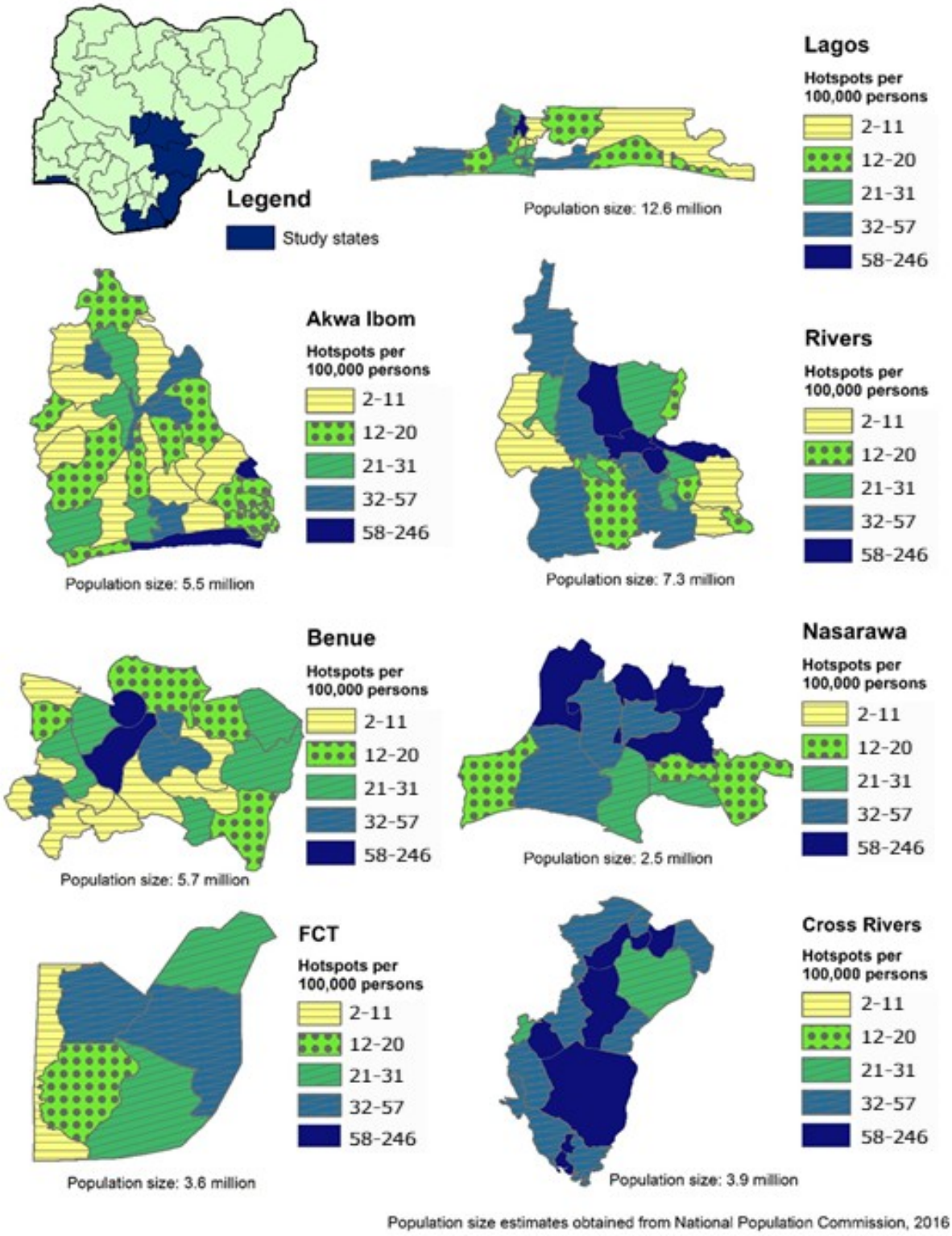
XSL•FO
RenderX

Figure 2. Type of key population hotspots by state.



XSL•FO
RenderX

Figure 1. Number of key population hotspots per 100,000 population by local government area in 7 US President's Emergency Plan for AIDS Relief Priority States in Nigeria.



Discussion

Overview of Our Findings

A total of 13,899 key population hotspots in 7 states were identified; these hotspots are more than the 6355 total hotspots previously reported from the last mapping exercise in the same 7 states of Nigeria during 2015 [22,23]. The sizable differences in the number of hotspots may be explained, in part, by divergent methods. Similar strategies were used to map hotspots of female sex workers and people who inject drugs in 2015 and 2018. We applied those strategies to map hotspots of men who have sex with men in 2018 as well. However, the 2015 efforts did not include hotspot mapping of men who have sex with men because investigators were concerned about the poor social visibility, and they focused efforts on capture-recapture to identify men who have sex with men for population size estimation. Although

XSL•FO
RenderX

there might be additional unknown factors that contributed to identifying more hotspots in 2018 than in 2015, community engagement played a major role in our efforts. Our strong collaboration with local key population–affiliated communitybased organizations facilitated entry into the key population community and enabled us to identify many new and previously undocumented hotspots. The diversity of the data collectors recruited enabled us to identify more hotspots frequented by minority key population groups, for example, older key population members or female people who inject drugs, in addition to many hotspots that were more commonly known among the public and the key population community.

We found fewer hotspots frequented by men who have sex with men compared to those frequented by female sex workers or people who inject drugs, and our data supported the notion that unlike female sex workers and people who inject drugs, public hotspots specific to men who have sex with men were less common, and based on information learned during the field activities from community members, this group tended to gather in private residences. This may be attributed to the widespread stigma against men who have sex with men in Nigeria that has made community members more reluctant to congregate in overt settings. We have previously described an increase in stigma following the enforcement of the 2014 Same Sex Marriage (Prohibition) Act, which made it "illegal not only to engage in an intimate relationship with a member of the same sex, but to attend or organize a meeting of gays, or patronize or operate any type of gay organization, including private clubs" [29-32]. Violence by the public and police against members of the lesbian, gay, bisexual, and transgender community was widely reported following the passing of the Same Sex Marriage (Prohibition) Act causing many lesbian, gay, bisexual, and transgender members, including men who have sex with men, to go into hiding [32,33].

Overall, hotspots of men who have sex with men and people who inject drugs were more difficult to locate compared to female sex worker hotspots and often required the knowledge of local key population informants. It is often advantageous for female sex worker hotspots to be visible and known to the public as a way for female sex workers to solicit their clients. The same concept does not apply for men who have sex with men and persons who inject drugs because of lower social acceptability and the potential for increased risk to their safety, although knowledge of male sex worker hotspots may be more common in the community of men who have sex with men. Although we report finding fewer hotspots of men who have sex with men and people who inject drugs than those of female sex workers, the absolute number of hotspots of men who have sex with men and people who inject drugs identified in this study is larger than that previously documented in Nigeria; 402 hotspots of men who have sex with men were reported by NACA in 2013 in the 5 states of Benue, Cross River, Lagos, Nasarawa, and Federal Capital Territory compared to 1026 hotspots of men who have sex with men that we report here (the Society for Family Health did not report the number of hotspots of men who have sex with men that they found in 2015) and 1184 hotspots of people who inject drugs were reported by the Society for Family Health in 2015 compared to 2729 hotspots of people who inject drugs that we report here [22,23].

We report more key population hotspots in state capitals and metropolitan city centers compared to those in less developed and metropolitan local government areas. This is true even after adjusting for the population size in the local government area. This finding is consistent with key population hotspot mapping and size estimation exercises conducted previously in Nigeria and in other countries [34]. This finding is also consistent with the expectations and experiences among stakeholders in the country and previous studies conducted in Nigeria [21-23]. Of the states in our study, Lagos was the most developed, commercial, and populous state with a projected population of 12.6 million. In comparison, our second, third, and fourth most populous states, that is, Rivers, Benue, and Akwa Ibom, had projected populations of 7.3 million, 5.7 million, and 5.5 million, respectively [27]. The projected population (2016) for Nigeria is 193 million [27]. Comparing the total number of hotspots by state with results published by the Society for Family Health in 2015, the largest difference in the number of hotspots of female sex workers and those of people who inject drugs were in Akwa Ibom, Benue, Cross River, and Rivers [23].

Our findings have important programming and policy implications. In addition to the substantially larger number of hotspots of female sex workers and people who inject drugs identified during 2018 compared to that identified in 2015, we also mapped the hotspots of men who have sex with men, filling a large data gap in Nigeria. Local government area health authorities are best positioned to identify other possible factors such as structural and behavioral changes that might have contributed to the growth of key population hotspots since 2015. Data may be used as is or linked by geocodes with a variety of other data sources (eg, health care facilities, testing sites, pharmacies, census projections, economic indicators) to inform outreach and prevention efforts, subnational HIV program target setting, or evaluation of service coverage and utilization. All results should be examined in the context of current HIV prevention programs and service delivery to ensure that response efforts are aligned and scaled appropriately to maximize impact and cost-effectiveness. Resources might need to be reallocated to broaden outreach and service coverage to include the newly identified hotspots. Hotspot profiles, including typology and peak activity, will assist in customizing efficient service delivery on the highest impact days and times at hotspots to reach specific subgroups within each key population. Our findings also provide rationale for increased advocacy for key population policy changes to reduce stigma and discrimination that might be hindering HIV response efforts.

These findings can be used by HIV service providers to identify strategic areas for placement of services and allocation of limited resources for targeted interventions. In addition to stigma and discrimination by health workers, travel time and transportation costs to treatment centers have been cited as barriers to treatment access and retention [9-14]. Placing prevention and treatment services that are conveniently located to venues where key populations socialize, seek clients, or inject drugs could increase testing and retention rates among key population members. Venue-based HIV testing and prevention services were

XSL•FO
RenderX

successful in reaching female sex workers and men who have sex with men who were unaware of their status in Malawi and Angola; 71% were not previously aware of their HIV status and received a diagnosis through the study [35]. Other studies report similar findings, supporting the success of venue-based testing and prevention services in reaching high-risk groups [36-39].

We engaged the key population community during formative assessment and continued partnership with key population community-based organizations throughout our activities. This effort proved to be enormously helpful in facilitating entrance to the key population community and activity buy-in. Close collaboration with local stakeholders, including government entities, law enforcement, and key population community-based organizations, minimized adverse outcomes from several security challenges during field activities. Our strong partnership with the local key population community-based organizations enabled us to identify more hotspots and strengthened partnerships for future collaboration with HIV programs.

This study was conducted as part of a larger key population hotspot mapping and size estimation activity that included an MS-CRC for population size estimation. The final list of hotspots produced from this exercise was used as the sampling frame for MS-CRC. MS-CRC also provided an opportunity to further correct and improve on hotspot location information such as local government area, address, and name of hotspot. Accurate documentation of addresses was essential as the list was used by different groups of enumerators to locate the same hotspot during MS-CRC.

Challenges and Limitations of This Study

The lack of formal addresses in our setting and multiple (slang) names for hotspots made de-duplication and locating of hotspots difficult. Although our partnership with the local stakeholders minimized security incidents and other challenges, there were still several security issues (eg, hotspots in or near areas where kidnapping, militancy, or other unrest were threats or travel to very remote hotspots requiring long drives into the night or an hour of water transportation via sea) and inclement weather conditions (ie, seasonal rains) during the data collection period that rendered several hotspots inaccessible. After fieldwork, the study team conducted additional reviews to verify and de-duplicate the list of key population hotspots. Our close partnerships with local key population community-based organizations were critical in this process. Although extensive time and resources were dedicated to de-duplicate, locate, and validate all hotspots, duplicates and inaccuracies may still be present.

As key population–defining behaviors are illegal in Nigeria and key population members often must evade law enforcement, the closing and relocation of hotspots were common. The numbers presented here reflect the situation at the time of our fieldwork. The states where our study was conducted were chosen as they were PEPFAR priority states at the time this study was conducted with evidence of high HIV burden and unmet needs for HIV/AIDS treatment services. These results should not be generalized to other states in Nigeria.

Conclusions

We identified many new and previously undocumented key population hotspots in the 7 priority states in Nigeria designated by the United States President's Emergency Plan for AIDS Relief. The implications of our findings have broad impact. Not only do they highlight the need for these states to strategically position and scale-up outreach and service programs, but also emphasize the need to effectively combat stigma and discrimination if Nigeria aims to have a fully successful HIV response. Engaging local key population community-based organizations throughout this activity allowed identification of a larger number of hotspots and strengthened partnerships for future collaboration with HIV programs. The small number of hotspots of men who have sex with men compared to that of female sex workers and people who inject drugs might be attributed to the 2014 Same Sex Marriage (Prohibition) Act or the practice of retreating into private residences, reflecting the low social visibility among these groups. Future studies may wish to consider broadening their definition of hotspots of men who have sex with men to include events in private residences or even web-based social platforms, which were increasing in popularity among the key populations in our setting. The information obtained from this exercise is expected to be used by the Government of the Federal Republic of Nigeria, donors, and implementing partners to design more strategically located and appropriately scaled key population–specific interventions. In addition, the final list of hotspots produced from this exercise was used as a sampling frame for MS-CRC for population size estimation, which was performed between October 2018 and December 2018 and summarized in a separate report.

Acknowledgments

The authors gratefully acknowledge Ochanya Ogbeh who led the informatics and management of REDCap data collection software for the study; Joyce Johnson who assisted in overall study management, budget, and finance support; University of Maryland Baltimore state supervisors Benjamin Obasa, Femi Samuel, Irene Iloegbunam, Margaret Gberikon, Obiageli Ezeabasili, Vera Nwandu, and Yemisi Kayode; Nigeria Ministry of Health, the National AIDS & STI Control Program; National Agency for the Control of AIDS; and State Agency for the Control of AIDS in Akwa Ibom, Benue, Cross River, Lagos, Nasarawa, Rivers, and Federal Capital Territory. This work was supported by the US President's Emergency Plan for AIDS Relief through the US CDC under the terms of cooperative agreement 5 U2G GH001976-02. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agencies.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Baral S, Beyrer C, Muessig K, Poteat T, Wirtz AL, Decker MR, et al. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. The Lancet Infectious Diseases 2012 Jul;12(7):538-549. [doi: 10.1016/s1473-3099(12)70066-x]
2. Sanders EJ, Okuku HS, Smith AD, Mwangome M, Wahome E, Fegan G, et al. High HIV-1 incidence, correlates of HIV-1 acquisition, and high viral loads following seroconversion among MSM. AIDS 2013;27(3):437-446. [doi: 10.1097/qad.0b013e32835b0f81]
3. Eluwa GI, Adebajo SB, Eluwa T, Ogbanufe O, Ilesanmi O, Nzelu C. Rising HIV prevalence among men who have sex with men in Nigeria: a trend analysis. BMC Public Health 2019 Sep 02;19(1):1201 [FREE Full text] [doi: 10.1186/s12889-019-7540-4] [Medline: 31477073]
4. Zalla LC, Herce ME, Edwards JK, Michel J, Weir SS. The burden of HIV among female sex workers, men who have sex with men and transgender women in Haiti: results from the 2016 Priorities for Local AIDS Control Efforts (PLACE) study. J Int AIDS Soc 2019 Jul;22(7):e25281 [FREE Full text] [doi: 10.1002/jia2.25281] [Medline: 31287624]
5. Baral S, Sifakis F, Cleghorn F, Beyrer C. Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000-2006: a systematic review. PLoS Med 2007 Dec;4(12):e339 [FREE Full text] [doi: 10.1371/journal.pmed.0040339] [Medline: 18052602]
6. 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. The Lancet 2012 Jul;380(9839):367-377. [doi: 10.1016/s0140-6736(12)60821-6]
7. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, Wirtz AL, et al. The increase in global HIV epidemics in MSM. AIDS 2013 Nov 13;27(17):2665-2678. [doi: 10.1097/01.aids.0000432449.30239.fe] [Medline: 23842129]
8. Fact sheet World AIDS Day global HIV statistics. The Joint United Nations Programme on HIV/AIDS. 2020 Dec 01. URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf [accessed 2020-12-27]
9. Pilgrim N, Musheke M, Raymond HF, Keating R, Mwale J, Banda L, et al. Quality of care and HIV service utilization among key populations in Zambia: a qualitative comparative analysis among female sex workers, men who have sex with men and people who use drugs. AIDS Care 2019 Apr;31(4):460-464. [doi: 10.1080/09540121.2018.1524119] [Medline: 30257574]
10. Nakanwagi S, Matovu J, Kintu B, Kaharuza F, Wanyenze R. Facilitators and Barriers to Linkage to HIV Care among Female Sex Workers Receiving HIV Testing Services at a Community-Based Organization in Periurban Uganda: A Qualitative Study. J Sex Transm Dis 2016 Jul 14. [doi: 10.1155/2016/7673014]
11. Mtetwa S, Busza J, Chidiya S, Mungofa S, Cowan F. “You are wasting our drugs”: health service barriers to HIV treatment for sex workers in Zimbabwe. BMC Public Health 2013 Jul 31;13(1):698 [FREE Full text] [doi: 10.1186/1471-2458-13-698] [Medline: 23898942]
12. Bogart LM, Chetty S, Giddy J, Sypek A, Sticklor L, Walensky RP, et al. Barriers to care among people living with HIV in South Africa: contrasts between patient and healthcare provider perspectives. AIDS Care 2013;25(7):843-853 [FREE Full text] [doi: 10.1080/09540121.2012.729808] [Medline: 23061894]
13. Kagee A, Remien R, Berkman A, Hoffman S, Campos L, Swartz L. Structural barriers to ART adherence in Southern Africa: Challenges and potential ways forward. Glob Public Health 2011;6(1):83-97 [FREE Full text] [doi: 10.1080/17441691003796387] [Medline: 20509066]
14. Yehia BR, Stewart L, Momplaisir F, Mody A, Holtzman CW, Jacobs LM, et al. Barriers and facilitators to patient retention in HIV care. BMC Infect Dis 2015 Jun 28;15:246 [FREE Full text] [doi: 10.1186/s12879-015-0990-0] [Medline: 26123158]
15. Final Nigeria integrated biological and behavioural surveillance survey (IBBSS), 2014. National Agency for the Control of AIDS. URL: https://naca.gov.ng/final-nigeria-ibbss-2014-report/ [accessed 2020-12-18]
16. New survey results indicate that Nigeria has an HIV prevalence of 1.4%. The Joint United Nations Programme on HIV/AIDS (UNAIDS). URL: https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2019/march/20190314_nigeria [accessed 2019-09-18]
17. Prevalence of HIV, total (% of population ages 15-49). The World Bank. URL: https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS [accessed 2019-10-08]
18. National HIV/AIDS Strategic Framework 2017-2021. National Agency for the Control of AIDS (NACA). URL: https://www.childrenandaids.org/sites/default/files/2017-11/NATIONAL-HIV-AND-AIDS-STRATEGIC-FRAMEWORK.pdf [accessed 2019-10-08]
19. Revised National HIV/AIDS Strategic Framework 2019-2021. National Agency for the Control of AIDS (NACA). URL: https://naca.gov.ng/revised-national-hiv-and-aids-strategic-framework-2019-2021/ [accessed 2019-10-08]
20. Hakim AJ, MacDonald V, Hladik W, Zhao J, Burnett J, Sabin K, et al. Gaps and opportunities: measuring the key population cascade through surveys and services to guide the HIV response. J Int AIDS Soc 2018 Jul;21 Suppl 5:e25119 [FREE Full text] [doi: 10.1002/jia2.25119] [Medline: 30033654]

XSL•FO
RenderX

21. Adebajo S, Eluwa G, Tocco J, Ahonsi B, Abiodun L, Anene O, et al. Estimating the number of male sex workers with the capture-recapture technique in Nigeria. Afr J Reprod Health 2013 Dec;17(4 Spec No):83-89. [Medline: 24689319]
22. HIV epidemic appraisals in Nigeria: evidence for prevention programme planning and implementation: Data from the first eight states. National Agency for the Control of AIDS (NACA). URL: https://tinyurl.com/4dcyx346 [accessed 2019-10-08]
23. Mapping and characterisations of key populations 2015: evidence for prevention programme planning and implementation in Nigeria. Society for Family Health. URL: http://www.sfhnigeria.org/wp-content/uploads/2017/08/National-Mapping-Characterisation-Report.pdf [accessed 2019-10-08]
24. Nigeria Country Operational Plan (COP) 2017 strategic direction summary. The United States President's Emergency Plan for AIDS Relief (PEPFAR). URL: https://www.pepfar.gov/documents/organization/272254.pdf [accessed 2019-08-21]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Nigeria: Administrative division, states and local governmental areas. Thomas Brinkhoff: City Population. URL: https://www.citypopulation.de/php/nigeria-admin.php [accessed 2019-09-18]
28. Nigeria - subnational administrative boundaries. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). URL: https://data.humdata.org/dataset/nga-administrative-boundaries [accessed 2017-02-23]
29. Gladstone R. Nigerian President signs ban on same-sex relationships. The New York Times. 2014 Jan 14. URL: https://www.nytimes.com/2014/01/14/world/africa/nigerian-president-signs-ban-on-same-sex-relationships.html [accessed 2019-09-18]
30. Tell Me Where I Can Be Safe: The impact of Nigeria's Same Sex Marriage (Prohibition) Act. Human Rights Watch. 2016 Oct 20. URL: https://www.hrw.org/report/2016/10/20/tell-me-where-i-can-be-safe/impact-nigerias-same-sex-marriage-prohibition-act [accessed 2019-09-18]
31. Nigeria Islamic court tries gay suspects in Bauchi. BBC News. 2014 Jan 15. URL: https://www.bbc.com/news/world-africa-25749308 [accessed 2019-09-18]
32. Schwartz SR, Nowak RG, Orazulike I, Keshinro B, Ake J, Kennedy S, et al. The immediate effect of the Same-Sex Marriage Prohibition Act on stigma, discrimination, and engagement on HIV prevention and treatment services in men who have sex with men in Nigeria: analysis of prospective data from the TRUST cohort. The Lancet HIV 2015 Jul;2(7):e299-e306. [doi: 10.1016/s2352-3018(15)00078-8]
33. Associated Press . Dozens arrested in Nigeria after anti-gay law passes. New York Post. URL: https://nypost.com/2014/01/14/dozens-arrested-in-nigeria-after-anti-gay-law-passes/ [accessed 2019-09-18]
34. Billong SC, Nguefack-Tsague G, Fokam J, Emmanuel F, Isac S, Fodjo RAT, et al. Mapping and size estimates of female sex workers in Cameroon: Toward informed policy for design and implementation in the national HIV program. PLoS One 2019;14(2):e0212315 [FREE Full text] [doi: 10.1371/journal.pone.0212315] [Medline: 30807616]
35. Herce M, Miller W, Bula A, Edwards J, Sapalalo P, Lancaster K, et al. Achieving the first 90 for key populations in sub-Saharan Africa through venue-based outreach: challenges and opportunities for HIV prevention based on PLACE study findings from Malawi and Angola. J Int AIDS Soc 2018 Jul;21 Suppl 5:e25132 [FREE Full text] [doi: 10.1002/jia2.25132] [Medline: 30033589]
36. Allan-Blitz L, Herrera MC, Calvo GM, Vargas SK, Caceres CF, Klausner JD, et al. Venue-Based HIV-Testing: An Effective Screening Strategy for High-Risk Populations in Lima, Peru. AIDS Behav 2019 Apr;23(4):813-819 [FREE Full text] [doi: 10.1007/s10461-018-2342-8] [Medline: 30506350]
37. Singh K, Brodish P, Mbai F, Kingola N, Rinyuri A, Njeru C, et al. A venue-based approach to reaching MSM, IDUs and the general population with VCT: a three study site in Kenya. AIDS Behav 2012 May;16(4):818-828 [FREE Full text] [doi: 10.1007/s10461-011-0103-z] [Medline: 22198312]
38. Weir SS, Figueroa JP, Scott M, Byfield L, Jones Cooper C, Hobbs MC. Reaching key populations through key venues: Insights from the Jamaica HIV Prevention Program. PLoS One 2018;13(11):e0206962 [FREE Full text] [doi: 10.1371/journal.pone.0206962] [Medline: 30475802]
39. Strömdahl S, Hoijer J, Eriksen J. Uptake of peer-led venue-based HIV testing sites in Sweden aimed at men who have sex with men (MSM) and trans persons: a cross-sectional survey. Sex Transm Infect 2019 Dec;95(8):575-579 [FREE Full text] [doi: 10.1136/sextrans-2019-054007] [Medline: 31113905]

Abbreviations

CDC: Centers for Disease Control and Prevention
MS-CRC: multiple-source capture-recapture
NACA: National Agency for Control of AIDS
PEPFAR: President’s Emergency Plan for AIDS Relief

XSL•FO
RenderX

REDCap: Research Electronic Data Capture

Edited by Y Khader; submitted 10.11.20; peer-reviewed by K Sabin, L Marc; comments to author 17.12.20; revised version received 31.12.20; accepted 02.01.21; published 22.02.21.

Please cite as:
Lo J, Nwafor SU, Schwitters AM, Mitchell A, Sebastian V, Stafford KA, Ezirim I, Charurat M, McIntyre AF
Key Population Hotspots in Nigeria for Targeted HIV Program Planning: Mapping, Validation, and Reconciliation
JMIR Public Health Surveill 2021;7(2):e25623
URL: https://publichealth.jmir.org/2021/2/e25623
doi:10.2196/25623
PMID:33616537

©Julia Lo, Samuel U Nwafor, Amee M Schwitters, Andrew Mitchell, Victor Sebastian, Kristen A Stafford, Idoteyin Ezirim, Man Charurat, Anne F McIntyre. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 22.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Protocol

Influenza and Respiratory Virus Surveillance, Vaccine Uptake, and Effectiveness at a Time of Cocirculating COVID-19: Protocol for the English Primary Care Sentinel System for 2020-2021

Simon de Lusignan[1*], MD, FRCGP; Jamie Lopez Bernal[2*], PhD; Rachel Byford[1*], BA; Gayatri Amirthalingam[2*], FMPH; Filipa Ferreira[1*], PhD; Oluwafunmi Akinyemi[1], PhD; Nick Andrews[2], PhD; Helen Campbell[2], PhD; Gavin Dabrera[2], MD; Alexandra Deeks[1], BSc; Alex J Elliot[2], PhD; Else Krajenbrink[3], MPH; Harshana Liyanage[1], PhD; Dylan McGagh[1], BSc; Cecilia Okusi[1], MRes; Vaishnavi Parimalanathan[1], MPH; Mary Ramsay[2], PhD; Gillian Smith[2], PhD; Manasa Tripathy[1], BSc, MSc; John Williams[1], MSc; William Victor[3], BSc; Maria Zambon[2], PhD; Gary Howsam[3], FRCGP; Brian David Nicholson[1], DPhil; Victoria Tzortziou Brown[3], PhD; Christopher C Butler[1], FMedSci; Mark Joy[1], PhD; FD Richard Hobbs[1], FMedSci

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

[2]Public Health England, London, United Kingdom

[3]Royal College of General Practitioners, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Simon de Lusignan, MD, FRCGP
Nuffield Department of Primary Care Health Sciences
University of Oxford
Eagle House
7 Walton Well Road
Oxford, OX2 6ED
United Kingdom
Phone: 44 1865617975
Email: simon.delusignan@phc.ox.ac.uk

Abstract

Background: The Oxford–Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) and Public Health England (PHE) are commencing their 54th season of collaboration at a time when SARS-CoV-2 infections are likely to be cocirculating with the usual winter infections.

Objective: The aim of this study is to conduct surveillance of influenza and other monitored respiratory conditions and to report on vaccine uptake and effectiveness using nationally representative surveillance data extracted from primary care computerized medical records systems. We also aim to have general practices collect virology and serology specimens and to participate in trials and other interventional research.

Methods: The RCGP RSC network comprises over 1700 general practices in England and Wales. We will extract pseudonymized data twice weekly and are migrating to a system of daily extracts. First, we will collect pseudonymized, routine, coded clinical data for the surveillance of monitored and unexpected conditions; data on vaccine exposure and adverse events of interest; and data on approved research study outcomes. Second, we will provide dashboards to give general practices feedback about levels of care and data quality, as compared to other network practices. We will focus on collecting data on influenza-like illness, upper and lower respiratory tract infections, and suspected COVID-19. Third, approximately 300 practices will participate in the 2020-2021 virology and serology surveillance; this will include responsive surveillance and long-term follow-up of previous SARS-CoV-2 infections. Fourth, member practices will be able to recruit volunteer patients to trials, including early interventions to improve COVID-19 outcomes and point-of-care testing. Lastly, the legal basis for our surveillance with PHE is Regulation 3 of the Health Service (Control of Patient Information) Regulations 2002; other studies require appropriate ethical approval.

Results: The RCGP RSC network has tripled in size; there were previously 100 virology practices and 500 practices overall in the network and we now have 322 and 1724, respectively. The Oxford–RCGP Clinical Informatics Digital Hub (ORCHID) secure

XSL•FO
RenderX

networks enable the daily analysis of the extended network; currently, 1076 practices are uploaded. We are implementing a central swab distribution system for patients self-swabbing at home in addition to in-practice sampling. We have converted all our primary care coding to Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) coding. Throughout spring and summer 2020, the network has continued to collect specimens in preparation for the winter or for any second wave of COVID-19 cases. We have collected 5404 swabs and detected 623 cases of COVID-19 through extended virological sampling, and 19,341 samples have been collected for serology. This shows our preparedness for the winter season.

Conclusions: The COVID-19 pandemic has been associated with a groundswell of general practices joining our network. It has also created a permissive environment in which we have developed the capacity and capability of the national primary care surveillance systems and our unique public health institute, the RCGP and University of Oxford collaboration.

(JMIR Public Health Surveill 2021;7(2):e24341) doi:10.2196/24341

KEYWORDS

COVID-19; general practice; influenza; computerized medical record systems; sentinel surveillance; coronavirus infections; records as topic; serology; virology

Introduction

Background

The Oxford–Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is a network of general practices and is now in its 54th season of surveillance and vaccine effectiveness in its long-standing collaboration with Public Health England (PHE) and predecessor bodies [1]. This season is complicated by the likely cocirculation of SARS-CoV-2 and its associated disease burden [2].

The RCGP RSC has been recruited as a nationally representative population [3] that provides pseudonymized data for surveillance of infectious diseases. However, the recent influx of practices may mean we have to adjust for practice distribution and ensure we recruit virology sampling practices that are evenly distributed. The disease surveillance program is commissioned by PHE and monitors 38 infectious diseases, including influenza; at the start of 2020, monitoring extended to include COVID-19 [4].

The recently expanded RCGP RSC extracts pseudonymized data from over 1700 general practices each week covering a nationally representative [3] population of over 13 million. Data from these practices are reported online in the RCGP's Weekly Return [5], which includes monitoring weekly rates of influenza-like illness (ILI) and other communicable and respiratory diseases for England. We also produce an annual report [6]. The RCGP RSC data set includes all coded data and all prescribed items, including vaccine exposure [3]. To manage our extended network of practices, we have set up a new analytics hub: the Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) [7].

The RCGP RSC conducts virology surveillance each influenza season. Given the onset of COVID-19, virological sampling has expanded to approximately 300 general practices (see Figure 1). We have reported on which groups of patients were more likely to test positive for COVID-19 [8], on excess mortality [9], and the shift away from face-to-face consultations during the first wave of COVID-19 pandemic [10].

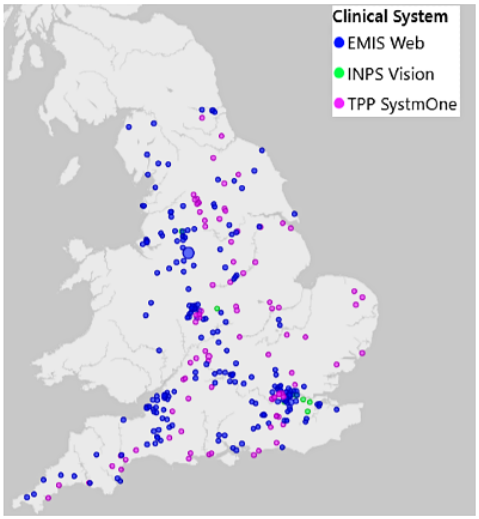
The RCGP RSC is also active in serosurveillance. During the 2018-2019 influenza season, we successfully conducted our first collection of serological samples from adults, linking them to their respective medical records [11,12]. Sentinel networks can provide a mechanism for systematic data collection and linkage to medical records and health outcomes [13], as well as sampling across the complete age range [14]. Current activity has included supporting PHE household studies and collecting convalescent sera from people previously infected with COVID-19.

Finally, the RCGP RSC network has supported trials and studies designated as high-priority public health research. These included recruiting to the PRINCIPLE (Platform Randomised trial of INterventions against COVID-19 In older peoPLE) trial [15] and supporting recruitment to RAPTOR-C19 (RAPid community Testing fOR COVID-19) [16].

This protocol sets out how the RCGP RSC will continue to provide enhanced passive surveillance, virological and serological surveillance from a subset of practices, and support high-priority public health research.

XSL•FO
RenderX

Figure 1. Royal College of General Practitioners Research and Surveillance Centre's virology sampling sites in England. Distribution by type of computerized medical record system is shown. EMIS: Egton Medical Information Systems; INPS: In Practice Systems; TPP: The Phoenix Partnership.



Aim

The aim of this protocol is to provide comprehensive surveillance of influenza, COVID-19, and other monitored conditions including health outcomes. The RCGP RSC network will also collect samples for virological and serological analysis, measure vaccine exposure and effectiveness, and support other high-priority public health research.

Objectives

Our six objectives are as follows:

1. To conduct surveillance of influenza and other monitored respiratory conditions, including COVID-19. This will be primarily from primary care data but will also include linkage to secondary care and mortality data sets to enable reporting of health outcomes. We will provide observatories, dashboards, and training to optimize data quality.
2. To collect virological and serological samples to report the presence and spread of monitored conditions from a nationally representative sample of practices.
3. To report vaccine uptake, estimate effectiveness, and investigate adverse events of interest (AEIs). We will focus on influenza vaccinations and possible new COVID-19 vaccines.
4. To support research to answer important public health questions, including trials of early interventions in the community, near-patient testing, and household spread.
5. Apply the Findable, Accessible, Interoperable, and Reusable (FAIR) principles so that our data are findable, accessible, interoperable, and reusable by other organizations.
6. To implement appropriate information and research governance.

Methods

Overview

The methods include six components: (1) surveillance of influenza, COVID-19, other monitored conditions, and associated health outcomes; (2) virological surveillance and serosurveillance; (3) monitoring vaccine uptake and effectiveness; (4) supporting priority public health research; (5) data curation; and (6) governance.

Surveillance of Influenza, COVID-19, Other Monitored Conditions, and Associated Health Outcomes

We will collect pseudonymized, routinely collected, coded clinical data, which has been coded into patient records, using UK Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) codes (SNOMED International), for enhanced passive surveillance. This will entail a wide range of clinical event data, including collection of monitored conditions, vaccine exposure, AEIs, hospital admissions, intensive care, and deaths, to monitor approved study outcomes and for any unexpected public health emergencies. Effective interoperability and communication will be facilitated by monthly pseudonymized linkage, facilitated by NHS (National Health Service) Digital and PHE [7], to secondary care and other data sources, including Secondary Uses Service, the continuously updated source of hospital data (see Textbox 1).

XSL•FO
RenderX

Textbox 1. Linkage to other data sets to support surveillance and research; individual data sets may have limitations on their use to specific projects and users.

NHS (National Health Service) Digital:

- Cancer Registration Data (reduced)
- Secondary Uses Service (SUS) Payment by Results Episodes
- SUS Payment by Results Outpatients
- SUS Payment by Results Accident and Emergency
- SUS Payment by Results Spells
- Mental Health Services Data Set
- Diagnostic Imaging Dataset
- Emergency Care Data Set (ECDS)
- COVID-19 Hospitalisation in England Surveillance System (CHESS) Dataset-CV19 NHS 111 Online
- Second Generation Surveillance System (SGSS): laboratory test reports
- Hospital Episode Statistics (HES; same granularity of supply as for SUS)
- Office of National Statistics (ONS) Mortality
- NHS 111 (free-to-call single nonemergency number medical helpline)

National Cancer Registration and Analysis Service (NCRAS):

- Cancer Registry (full)
- Chemotherapy data set
- Radiotherapy data set

We have converted all our primary care coding to SNOMED CT coding and are using this throughout our data processing [17] (see Multimedia Appendix 1). We have been at the forefront of ensuring that primary care systems code cases correctly and that we comprehensively capture historic data [18]. This has included the development of database-linked tools to programmatically search SNOMED CT for relevant tools. The hierarchical nature of SNOMED CT is that it requires machine searching rather than manual searching (see Multimedia Appendix 2). This will continue to deliver effective COVID-19 surveillance in parallel with the annual influenza surveillance this season.

We will enhance surveillance through the provision of dashboards and observatories to provide feedback about levels of care compared to other practices in our network and to improve data quality, clinical care, and patient safety. There will be a particular focus on data quality, including the recording of ILIs, the use of the RCGP RSC case definition (see Textbox 2 [19-21]), recording of upper and lower respiratory tract infections, and recording of suspected COVID-19 cases, with an emphasis on recording episode type (ie, whether this is a new case or a follow-up). We will programmatically adjust for when episode type is not recorded [18,22].

This year we will extend our virology sampling window to 10 days. We will also report on health outcomes, including mortality [9] (see Figure 2).

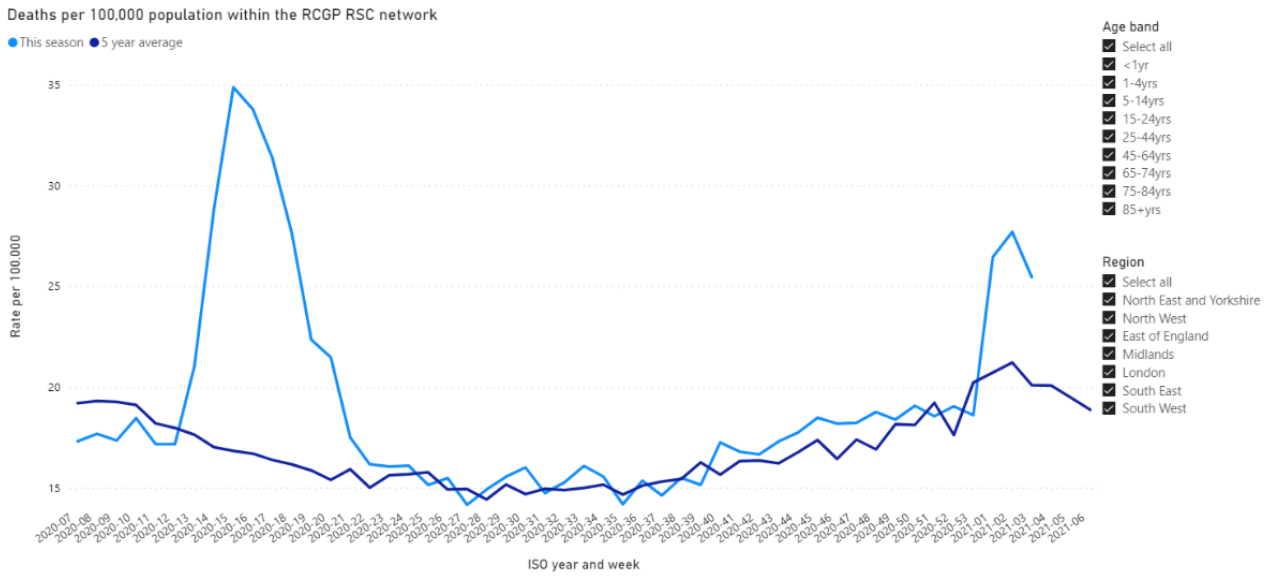
Textbox 2. Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) case definitions of influenza-like illness (ILI).

The RCGP RSC definition of ILI:

- An acute respiratory illness with a temperature measured as, reported as, or plausibly ≥38 °C and cough, with onset within the past 10 days. ILI cases should not have another more plausible diagnosis. ILI cases have a sudden onset, and there are often symptoms suggestive of systemic upset—myalgia, fatigue, malaise, headache, etc. RCGP RSC stresses the notion of symptoms within 10 days of onset to differentiate acute episodes, with swabs only wanted within 7 days of onset.
- This definition is compatible with the World Health Organization (WHO) and European Centre for Disease Prevention and Control (ECDC) definitions [19,20]. The WHO definition has the highest specificity (21.4%) and the ECDC definition has the highest sensitivity (96.1%) [21].

XSL•FO
RenderX

Figure 2. Royal College of General Practitioners Research and Surveillance Centre observatory showing mortality through the first wave of the COVID-19 pandemic in England. Deaths per 100,000 population are displayed over the International Organization for Standardization (ISO) weeks for 2019 and 2020.



Virological Surveillance and Serosurveillance

A volunteer subset of practices will conduct virological surveillance (see Multimedia Appendix 3). They will conduct or arrange virological sampling of patients presenting with ILI or lower respiratory tract infection, including acute bronchitis and bronchiolitis in children under the age of 5 years. We will include those suspected to have, or have been exposed to, COVID-19 who have persistent cough, loss of taste and/or smell, shortness of breath, history of fever, and/or the presence of a wheeze. We will look for practices to collect 20 samples per week; additional samples can be collected if the spread that is present is within a household or a communal establishment. Practices will receive feedback via dashboards [23] (see Textbox 3 [5,24-28]). Our target is to collect 600 samples per week across all age bands.

The anticipated changes in the coming year are as follows:

1. We will reduce the burden of information collected on the virology request form. Thus, a common but reduced set of data will be collected for all patients offered virology sampling. Vaccination history and a record of their comorbidities will have already been collected electronically as part of the standard data extract. However, we will stress the importance of recording the NHS number as a unique identifier, to help ensure efficient linkage of clinical data to test results.
2. People who volunteer to self-swab will be given a voucher code they can use online to arrange for a sample to be sent; those in the surgery will have a swab taken by their general practitioner or practice team, and this will be sent through the post to the PHE reference laboratory. Patients taking self-swabs should be given their NHS number to include in their swab request to improve data matching.
3. All results will be sent to practices online via the e-Lab system [29]; patients who self-swab and provide a mobile number will additionally receive their results by text message (see Figure 3).

XSL•FO
RenderX

Textbox 3. Dashboards, online support, and observatories—also called weekly returns—to monitor and support data recording and key public health projects.

Practice Dashboard

My Practice Dashboard [24]

All practices:

- Seasonal wellness: all practices can see vaccine uptake and incidence of key monitored conditions
- COVID-19: all practices will compare their incidence of COVID-19 with the rest of the network

Volunteer practices involved in specific surveillance or trials:

- Virology: for practices participating in specimen collection; in-practice and home tests
- Serology: for practices to observe their serology sampling; primarily by age band
- PRINCIPLE (Platform Randomised trial of INterventions against COVID-19 In older peoPLE) trial: to help practices know their level of recruitment
- Online support:
  - Facebook page [25]
  - Twitter account [26]
  - Podcast [27]

Observatories and Weekly Returns

Oxford–Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) weekly return:

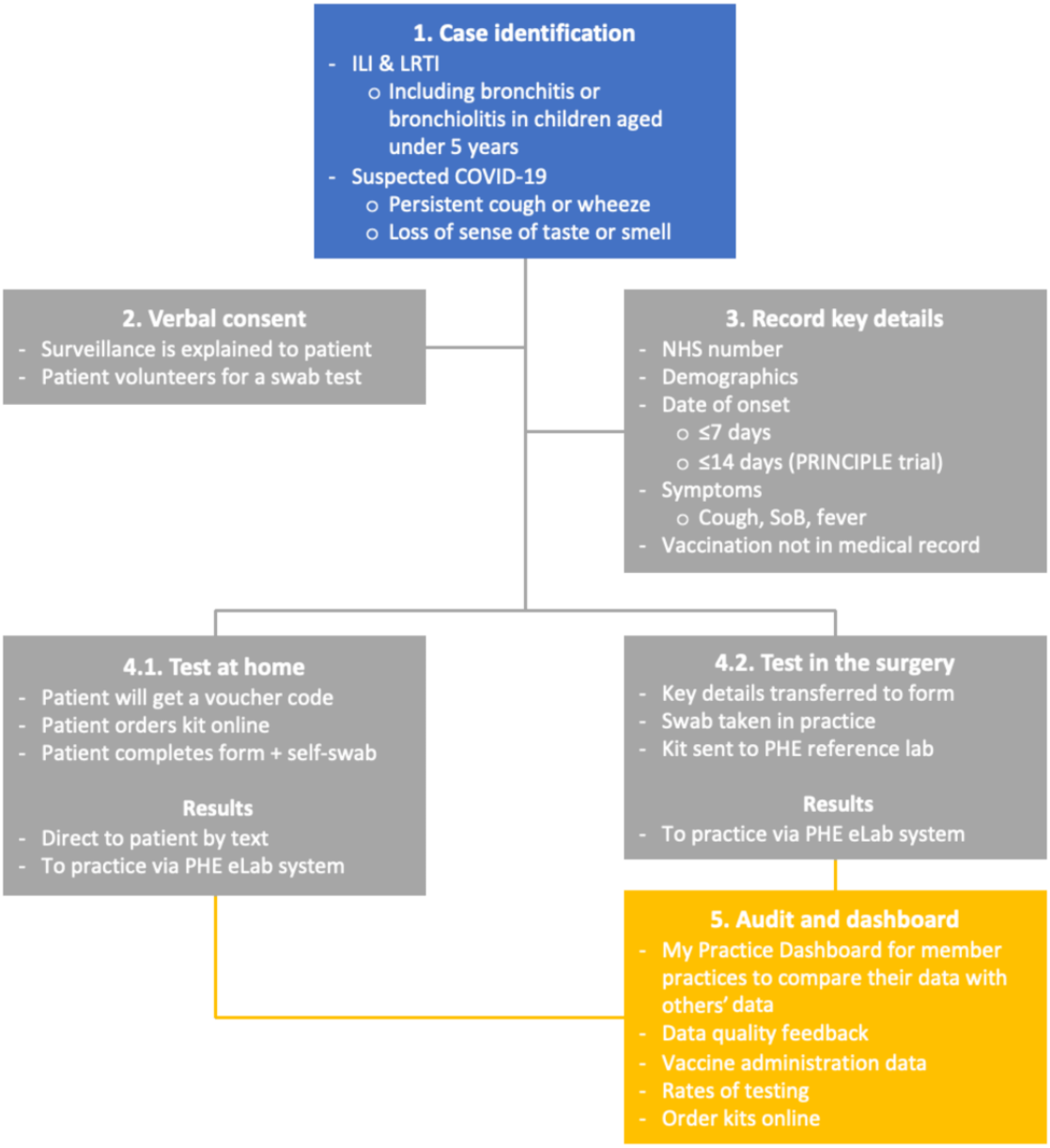
- Twice weekly report—during the pandemic—on influenza, COVID-19, and other monitored conditions; reports are published at the RCGP RSC website [5]—follow the "RSC Communicable and Respiratory Disease report" link on the "Public Health Data" page; this is the 54th year of this report

Observatories [28]. We produce weekly reports online for the observatories, which provide an interactive weekly return of national data; these are as follows:

- Seasonal wellness
- COVID-19
- Mortality
- Virology sampling
- Serology sampling
- PRINCIPLE trial eligible patients

XSL•FO
RenderX

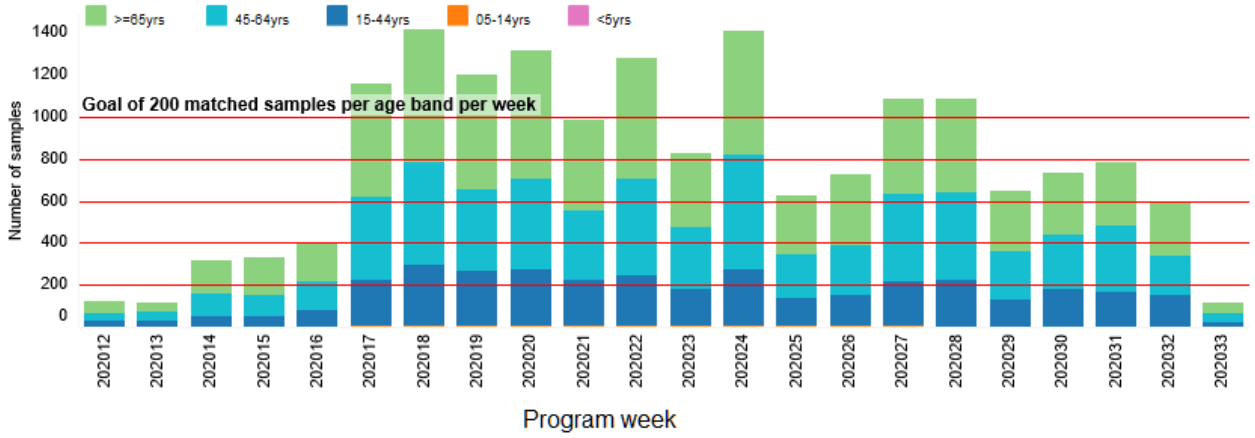
Figure 3. Flowchart for case identification and sampling for the 2020-2021 season surveillance. The National Health Service (NHS) number is a vital unique identifier for all forms across virology and serology, including for patients requesting a test kit at home. ILI: influenza-like illness; LRTI: lower respiratory tract infection; PHE: Public Health England; PRINCIPLE: Platform Randomised trial of INterventions against COVID-19 In older peoPLE; SoB: shortness of breath.



A volunteer group of practices will also participate in serological surveillance. The largest project will be the opportunistic collection of an extra blood sample from patients undergoing a routine blood test. Practices will be asked to collect across age bands (aged 10 years and above) (see Figure 4), with feedback provided to general practices via a dashboard and overall via an observatory [30]. We are now recruiting practices from areas with more ethnically mixed populations to provide greater insight into disease spread in ethnic minority populations. Other serology projects include extended serosurveillance in a subset of patients with positive virology tests (n=63) as well as recruiting and sampling patients included in the PHE-led household contacts of confirmed COVID-19 cases (HOCO) study [30]. We are piloting responsive surveillance in areas with higher disease incidence.

XSL•FO
RenderX

Figure 4. Number of serological surveillance samples collected per week. The program was initiated in week 11; we have been taking active steps to encourage similar numbers across age bands per week.



Monitoring Vaccine Uptake, Effectiveness, and Adverse Events of Interest

Capturing vaccine exposure is critical to conducting vaccine effectiveness studies. With an increasing range of vaccines available, capturing exposure data when not recorded within the practice, beyond the fact that a vaccine has been given, presents significant challenges [31]. Further issues include difficulties capturing vaccine batch numbers [32]. We will ask practices to preload the vaccine brands and batch numbers online.

We plan to continue to report AEIs following vaccination using routinely collected data, though we would have the capability to provide patient AEI reporting cards if commissioned for COVID-19 or other vaccines [33,34]. Practices will be provided with prompts to improve coding. Good data quality regarding AEI recording may identify any differences between influenza vaccine types, to elucidate whether egg- or cell-based vaccines with or without adjuvants are associated with different rates of AEIs. There may also be up to six types of COVID-19 vaccines available for which we may be able to provide uptake and AEI data.

Supporting Priority Public Health Research

Overview

The RCGP RSC will support a range of research (see Multimedia Appendix 4), with the PRINCIPLE [15] and RAPTOR-C19 [16] trials as the top priorities. The RCGP RSC is a member of a number of European studies, including a point-of-care testing study and data sharing as part of the DRIVE (Development of Robust and Innovative Vaccine Effectiveness) consortium [35,36] and a European study of respiratory syncytial virus epidemiology [37]. In partnership with PHE, we are also members of the I-MOVE (Influenza–Monitoring Vaccine Effectiveness in Europe) consortium, which has extended its work into COVID-19 [38].

We will support two ethically approved generic protocols for COVID-19 research: DECISION-COVID (DEfining the CharacterIStIcs Of Individuals with suspected Novel COronaVIrus Disease and risk factors for development of the disease) and MAINROUTE (Monitoring Attendance, INvestigation, Referral, and OUTcomEs in primary care: impact of and recovery from COVID-19 lockdown). DECISION-COVID provides a framework for looking at medications that might impact on COVID-19 outcomes; MAINROUTE assesses the impact of lockdown on the wider implications to care (see Multimedia Appendix 4). We have already reported on how lockdown has led to a reorganization of primary care service delivery [10]; in addition, we are part of the RECAP (Remote COVID-19 Assessment in Primary Care) early warning score validation study, which was set to start in September 2020 [39].

Statistical Methods

A range of statistical methods will be used that are appropriate to the area of surveillance or the research undertaken. We will use propensity matching and a wide range of other methods to work within the constraints of routine data [40].

Vaccine effectiveness will be assessed using test-negative design [41,42]. Patients meeting the predefined clinical criteria will undergo laboratory testing for specific circulating viruses. Patients who test positive will comprise the cases; patients who test negative to both influenza and COVID-19 will form the control group. In the clinical at-risk populations, as defined by NHS England, vaccine effectiveness will be calculated as follows: (1 − adjusted vaccination odds ratio) × 100% [42]. This will be estimated overall and stratified by influenza virus subtype for all ages. Cohort analysis will also be performed with vaccine effectiveness estimated as 1 – risk in vaccinated subjects and 1 – risk in unvaccinated subjects, with adjustment for confounding variables.

We will report AEIs using self-controlled case series design [43,44]. An exposure period will be identified shortly after vaccination, and a baseline period will form the remainder of the observation period. The frequency of AEIs will be compared to identify any invariant confounding [44,45]. We will use self-controlled case series to report the different rates of AEIs between vaccines; biases should be similar between vaccines. For influenza, we will continue to use the European Medicines Agency–specified AEIs; should COVID-19 vaccination commence this season, we will identify AEIs. Notwithstanding the enlarged size of the network, we are not large enough to detect rare events.

XSL•FO
RenderX

Data Curation

We strongly support the FAIR principles [46]. We are a founding member of the Health Data Research UK Data Alliance, and our data are included in their metadata, with our activities ranked as high priority [47]. We are creating a data set within ORCHID, an observational research platform that has themed databases that can be made readily available to researchers simultaneously; these will be used to support DECISION-COVID and other studies (see Multimedia Appendix 4) [48]. Additionally, we are linking over 50 years of virology and RCGP RSC data, as part of the Wellcome Trust–funded quinquagenarian database project, to make these data more accessible [49]. ORCHID [7] aims to provide a large, near-real-time–themed primary care health informatics hub for the use of data from consenting patients in clinical trials and to supplement existing disease surveillance using in situ network data without large-scale data extraction. We aim to make outputs available using common data models, which we will be providing to the European Health Data Evidence Network [50,51].

Governance

The surveillance for the coming year will be initiated by a joint letter to volunteer practices informing them of their role in the surveillance process. The legal basis for our surveillance with PHE is defined as Health Protection under Regulation 3 of the Health Service (Control of Patient Information) Regulations 2002 and is approved annually by the PHE Caldicott Guardian; other studies require appropriate ethical approval. The University of Oxford is compliant with the General Data Protection Regulation and the NHS Digital Data Security and Protection policy.

We want to minimize any risk associated with taking part in this surveillance. Infection prevention and control advice will follow extant national guidance. Our member practices already have most processes in place to meet Regulation 12 of the Health and Social Care Act 2008 (Regulated Activities) Regulations 2014, concerning safe care and treatment, and are periodically inspected by the Care Quality Commission [52]. Our training will include reminders about safe handling of specimens and will recommend revision of infection control measures, where needed.

Results

Surveillance Network Development

The COVID-19 pandemic has led to considerable numbers of practices joining the Oxford–RCGP RSC network. The collaboration with EMIS (Egton Medical Information Systems) Health has made it much easier for practices to contribute pseudonymized patient data at rapid speed and at a large scale. The network has tripled in size since January 2020, from 500 practices to 1724. Of these practices, 1076 are using the new ORCHID platform being developed in collaboration with EMIS for direct data extract. Pseudonymized data from over 13 million patients has enabled the growth of a rich database from which researchers are developing a deeper understanding of the COVID-19 pandemic and other diseases in general practice. We have opened a Facebook page and Twitter account to provide more online information and help (see Textbox 3).

Virological Surveillance and Serosurveillance

We formerly had 100 practices taking part in the national influenza surveillance scheme. This has expanded to 322 for COVID-19 surveillance. As of August 17, 2020, the surveillance system has collected 5357 swabs and detected 581 cases of COVID-19 through extensive virological sampling. A total of 273 practices have contributed toward serosurveillance, collecting a baseline of 2000 samples and an additional 1000 samples per week. A total of 19,761 samples have been collected for serology as of August 17, 2020. We are also piloting serosurveillance in two cities in the northwest of England with a high incidence of COVID-19.

Monitoring Vaccine Uptake and Effectiveness

We started collecting vaccine brands and batch numbers via our online portal on September 1, 2020, for the upcoming influenza season. We have made provisions in our dashboards and observatories for the changed vaccination policy for influenza and for virology sampling in the coming season. We have successfully piloted the secure transfer of data between ORCHID and PHE using the OxFile – Large File Exchange Service secure data transfer system [53].

Supporting Priority Public Health Research

We have recruited 517 practices into the PRINCIPLE trial. A total of 440 patients have had successful screening and have been randomized. As of August 13, 2020, 537 patients have been recruited via general practitioner practices set up as official recruiting sites. Oxford–RCGP RSC practices have recruited 81.9% of these patients.

Data Curation

Our key focus is the development of ORCHID and accessibility of our extended data set. Currently, 1076 practices use the ORCHID virtual server.

Governance

We will put in place processes to allow approved researchers access to ORCHID.

Discussion

Principal Findings

We have tripled the sign-up to the Oxford–RCGP RSC and are well on our way to operationalizing this new extended network ahead of the coming winter, in a season complicated by SARS-CoV-2 circulation and an extended flu vaccination program.

The Oxford–RCGP RSC now has in place the most comprehensive set of linked data in its history, supplemented by reference lab virology and serology. Over half of our network will have data available on a daily basis via unprecedented, timely, remote researcher access to our themed data sets.

We have also introduced self-swabbing to increase safety and convenience for patients and practices, to enable direct relay of COVID-19 results to patients, and to reduce the burden of form

XSL•FO
RenderX

filling. We have also extended the number and sophistication of our dashboards and observatories, adding interactivity and more extensive help.

Safety of practices remains our primary concern and we are not aware of any increased risk to practice staff or other patients from involvement in surveillance. Personal protective equipment was difficult to source at the start of the pandemic, as were SARS-CoV-2 tests. Both are now much more available to primary care.

Comparison With Previous Work

Pandemic preparedness is part of the role of the RCGP RSC. The RCGP RSC has operated for over 50 years and has been involved in collecting samples to monitor disease and vaccine effectiveness through the Hong Kong flu pandemic of 1968-1969, the Russian flu of 1977-1978, the 2009 swine flu pandemic [54,55], and the first wave of the COVID-19 pandemic earlier this year.

The introduction of self-swabbing kits this year will enable wider distribution, allowing for an increase in the number of people to be tested. People in self-isolation and quarantine can access testing kits without leaving the confines of their homes, negating the risk of transmission to others with whom they interact with in transit [6]. Self-swabbing appears to be reliable [56,57].

We have shown considerable adaptability so far this year, tackling a wide range of issues and producing publicly available dashboards as well as research outputs [6-8,10,22].

Strengths and Limitations

The strengths of our network lie in the willingness of practices to share data and the quality of those data. The Oxford–RCGP RSC team has a deep understanding of primary care data, the context of data recording, and the ability to process routine data. We have developed even closer relationships with computerized medical records system suppliers, especially EMIS. We have a long-term partnership with PHE and its predecessor bodies. There is a good understanding of the strengths and limitations of routine data and the capabilities of primary care.

The limitations of our system are those of all routine data, in that the data may not be complete or accurate. Ensuring we have good data quality is key for the reliability of our screening. There are few available comparators, but we know the network detects more new cases of some cardiovascular diseases than others [58]. Given our ambitious program of scaling, via a new pathway for virology surveillance based on patient-completed data, a further limitation is that these data are currently not linked back into their general practice record. We have also found that the success of open access testing means that individuals can be tested more quickly at a public testing station [59] than self-swabbing via a test through the surveillance system, thereby restricting the extent of possible collection.

While our network is large enough to report the relative incidence of common AEIs following influenza vaccination, comparing brands and vaccine types, it may be too small to comment on very rare events. Little is known about the anticipated AEIs with the new COVID-19 vaccine, so it is currently unclear what role we could play.

While we are part of many European and international networks, we did not have the opportunity to rapidly produce a shared international protocol. However, this document will include the learning from our decades of joint surveillance. It is being made available as a preprint to enable comments to be made and to allow learning to be shared.

Conclusions

The Oxford–RCGP RSC has had a groundswell of interest in membership during the first wave of the COVID-19 pandemic. We have grown in network size and functionality. Our challenges are to maintain data quality and to help support our practices in identifying sufficient volunteers to deliver the required number of samples as we take on more general practice members. Another challenge is the uncertainty of a winter season with cocirculating COVID-19. The pace of change in our understanding of COVID-19 requires us to move ahead rapidly, rather than in close coordination with others. This protocol sets out our approach to surveillance in the coming season and makes our approach widely available for comment or use by others involved in surveillance.

Acknowledgments

The authors would like to thank the participating practices and patients for providing the data for this cohort. We acknowledge collaboration with the general practitioner computer system suppliers—EMIS Health, The Phoenix Partnership, In Practice Systems, and Apollo Medical Systems—as well as with PHE and our other funders and collaborators. The RCGP RSC surveillance work is primarily funded by PHE.

Authors' Contributions

SdeL wrote the first and final draft of the manuscript with early feedback from NA. OA developed the next iteration. The rapid evolution of the surveillance system was led by SdeL, JLB, and GA, with substantial input from all authors in their areas of expertise. The SNOMED CT machine-readable tools were developed by RB and JW with input from HL, DMcG, and other team members.

Conflicts of Interest

SdeL has had unrelated projects in influenza and gastroenteritis funded by GSK, Takeda, and Seqirus and has been a member of Global Advisory Boards for Seqirus and Sanofi.

XSL•FO
RenderX

Multimedia Appendix 1
In-house developed Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) coding tools to permit development of variable lists that match coding in computerized medical record (CMR) systems.
[DOCX File , 571 KB - publichealth_v7i2e24341_app1.docx]

Multimedia Appendix 2
Newly released Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) concepts related to COVID-19 and SARS-CoV-2 grouped into relevant categories.
[DOCX File , 20 KB - publichealth_v7i2e24341_app2.docx]

Multimedia Appendix 3
Flowchart of the 2020-2021 influenza season surveillance.
[DOCX File , 128 KB - publichealth_v7i2e24341_app3.docx]

Multimedia Appendix 4
Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC)–Public Health England (PHE) Joint Virology Summary, including use of virology data, July 2020.
[DOCX File , 18 KB - publichealth_v7i2e24341_app4.docx]

References

1. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract 2017 Oct;67(663):440-441 [FREE Full text] [doi: 10.3399/bjgp17X692645] [Medline: 28963401]
2. Preparing for a Challenging Winter 2020-21. London, UK: The Academy of Medical Sciences; 2020 Jul 14. URL: https://acmedsci.ac.uk/file-download/51353957 [accessed 2020-08-17]
3. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: A cohort profile. BMJ Open 2016 Apr 20;6(4):e011092 [FREE Full text] [doi: 10.1136/bmjopen-2016-011092] [Medline: 27098827]
4. de Lusignan S, Lopez Bernal J, Zambon M, Akinyemi O, Amirthalingam G, Andrews N, et al. Emergence of a novel coronavirus (COVID-19): Protocol for extending surveillance used by the Royal College of General Practitioners Research and Surveillance Centre and Public Health England. JMIR Public Health Surveill 2020 Apr 02;6(2):e18606 [FREE Full text] [doi: 10.2196/18606] [Medline: 32240095]
5. RCGP Research and Surveillance Centre (RSC). Royal College of General Practitioners. URL: https://www.rcgp.org.uk/rsc [accessed 2020-03-07]
6. de Lusignan S, Correa A, Pathirannehelage S, Byford R, Yonova I, Elliot AJ, et al. RCGP Research and Surveillance Centre Annual Report 2014-2015: Disparities in presentations to primary care. Br J Gen Pract 2017 Jan;67(654):e29-e40 [FREE Full text] [doi: 10.3399/bjgp16X688573] [Medline: 27993900]
7. de Lusignan S, Jones N, Dorward J, Byford R, Liyanage H, Briggs J, et al. The Oxford Royal College of General Practitioners Clinical Informatics Digital Hub: Protocol to develop extended COVID-19 surveillance and trial platforms. JMIR Public Health Surveill 2020 Jul 02;6(3):e19773 [FREE Full text] [doi: 10.2196/19773] [Medline: 32484782]
8. de Lusignan S, Dorward J, Correa A, Jones N, Akinyemi O, Amirthalingam G, et al. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: A cross-sectional study. Lancet Infect Dis 2020 Sep;20(9):1034-1042 [FREE Full text] [doi: 10.1016/S1473-3099(20)30371-6] [Medline: 32422204]
9. Joy M, Hobbs FDR, McGagh D, Akinyemi O, de Lusignan S. Excess mortality from COVID-19 in an English sentinel network population. Lancet Infect Dis 2020 Aug 04:1 [FREE Full text] [doi: 10.1016/S1473-3099(20)30632-0] [Medline: 32763192]
10. Joy M, McGagh D, Jones N, Liyanage H, Sherlock J, Parimalanathan V, et al. Reorganisation of primary care for older adults during COVID-19: A cross-sectional database study in the UK. Br J Gen Pract 2020 Aug;70(697):e540-e547 [FREE Full text] [doi: 10.3399/bjgp20X710933] [Medline: 32661009]
11. de Lusignan S, Borrow R, Tripathy M, Linley E, Zambon M, Hoschler K, et al. Serological surveillance of influenza in an English sentinel network: Pilot study protocol. BMJ Open 2019 Mar 08;9(3):e024285 [FREE Full text] [doi: 10.1136/bmjopen-2018-024285] [Medline: 30852535]
12. Osborne K, Gay N, Hesketh L, Morgan-Capner P, Miller E. Ten years of serological surveillance in England and Wales: Methods, results, implications and action. Int J Epidemiol 2000 Apr;29(2):362-368. [doi: 10.1093/ije/29.2.362] [Medline: 10817137]

XSL•FO
RenderX

13. Metcalf CJE, Farrar J, Cutts FT, Basta NE, Graham AL, Lessler J, et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. Lancet 2016 Aug 13;388(10045):728-730 [FREE Full text] [doi: 10.1016/S0140-6736(16)30164-7] [Medline: 27059886]
14. de Lusignan S, Correa A. Opportunities and challenges of a World Serum Bank. Lancet 2017 Jan 21;389(10066):250-251. [doi: 10.1016/S0140-6736(17)30046-6] [Medline: 28118910]
15. PRINCIPLE: A trial evaluating treatments for suspected COVID-19 in people aged 50 years and above with pre-existing conditions and those aged 65 years and above. ISRCTN Registry. 2020. URL: http://www.isrctn.com/ISRCTN86534580 [accessed 2020-08-17]
16. Rapid community testing for COVID-19. ISRCTN Registry. 2020. URL: http://www.isrctn.com/ISRCTN14226970 [accessed 2020-08-17]
17. de Lusignan S. Codes, classifications, terminologies and nomenclatures: Definition, development and application in practice. Inform Prim Care 2005;13(1):65-70 [FREE Full text] [doi: 10.14236/jhi.v13i1.580] [Medline: 15949178]
18. de Lusignan S, Williams J. To monitor the COVID-19 pandemic we need better quality primary care data. BJGP Open 2020;4(2):1-3 [FREE Full text] [doi: 10.3399/bjgpopen20X101070] [Medline: 32295793]
19. Roderick P, Macfarlane A, Pollock AM. Getting back on track: Control of COVID-19 outbreaks in the community. BMJ 2020 Jun 25;369:m2484. [doi: 10.1136/bmj.m2484] [Medline: 32586845]
20. Fitzner J, Qasmieh S, Mounts AW, Alexander B, Besselaar T, Briand S, et al. Revision of clinical case definitions: Influenza-like illness and severe acute respiratory infection. Bull World Health Organ 2018 Feb 01;96(2):122-128 [FREE Full text] [doi: 10.2471/BLT.17.194514] [Medline: 29403115]
21. EU case definitions. European Centre for Disease Prevention and Control. 2018. URL: https://www.ecdc.europa.eu/en/surveillance-and-disease-data/eu-case-definitions [accessed 2020-08-17]
22. Smith N, Livina V, Byford R, Ferreira F, Yonova I, de Lusignan S. Automated differentiation of incident and prevalent cases in primary care computerised medical records (CMR). Stud Health Technol Inform 2018;247:151-155. [Medline: 29677941]
23. Liyanage H, Akinyemi O, Pathirannahelage S, Joy M, de Lusignan S. Near real time feedback of seasonal influenza vaccination and virological sampling: Dashboard utilisation in a primary care sentinel network. Stud Health Technol Inform 2020 Jun 16;270:1339-1340. [doi: 10.3233/SHTI200431] [Medline: 32570648]
24. RCGP RSC My Practice Dashboard. Tableau Public. 2021. URL: https://public.tableau.com/profile/orchid#!/vizhome/MyPracticeDashboard_ORCHID/Home [accessed 2021-01-28]
25. Oxford RCGP RSC. Facebook. URL: https://www.facebook.com/groups/320547995977306/ [accessed 2021-01-28]
26. ORCHID. Twitter. URL: https://twitter.com/ORCHID_Oxford [accessed 2021-01-28]
27. Podcasts. Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub. URL: https://orchid.phc.ox.ac.uk/index.php/podcasts/ [accessed 2021-01-28]
28. Clinical Informatics and Health Outcomes Research Group. Nuffield Department of Primary Care Health Sciences. URL: https://www.phc.ox.ac.uk/research/clinical-informatics-and-health-outcomes [accessed 2020-01-28]
29. e-Lab LabPortal. Public Health England. URL: https://phe-elab.phe.org.uk [accessed 2020-08-17]
30. Serology observatory. Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub. URL: https://orchid.phc.ox.ac.uk/index.php/serology/ [accessed 2021-01-26]
31. HOCO surveillance for RCGP RSC practices. Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub. URL: https://orchid.phc.ox.ac.uk/index.php/cov-19-hoco/ [accessed 2021-01-26]
32. de Lusignan S, Hoghton M, Rafi I. Flu vaccination by pharmacists leads to suboptimal medical records. BMJ 2017 Nov 09;359:j5084. [doi: 10.1136/bmj.j5084] [Medline: 29122753]
33. de Lusignan S, Ferreira F, Damaso S, Byford R, Pathirannehelage S, Yeakey A, et al. Enhanced passive surveillance of influenza vaccination in England, 2016-2017- An observational study using an adverse events reporting card. Hum Vaccin Immunother 2019;15(5):1048-1059 [FREE Full text] [doi: 10.1080/21645515.2019.1565258] [Medline: 30648923]
34. de Lusignan S, Damaso S, Ferreira F, Byford R, McGee C, Pathirannehelage S, et al. Brand-specific enhanced safety surveillance of GSK's Fluarix Tetra seasonal influenza vaccine in England: 2017/2018 season. Hum Vaccin Immunother 2020 Aug 02;16(8):1762-1771 [FREE Full text] [doi: 10.1080/21645515.2019.1705112] [Medline: 32118513]
35. de Lusignan S, Hoang U, Liyanage H, Yonova I, Ferreira F, Diez-Domingo J, et al. Feasibility of point-of-care testing for influenza within a national primary care sentinel surveillance network in England: Protocol for a mixed methods study. JMIR Res Protoc 2019 Nov 11;8(11):e14186 [FREE Full text] [doi: 10.2196/14186] [Medline: 31710303]
36. de Lusignan S, Hoang U, Liyanage H, Tripathy M, Yonova I, Byford R, et al. Integrating molecular point-of-care testing for influenza into primary care: A mixed-methods feasibility study. Br J Gen Pract 2020 Aug;70(697):e555-e562 [FREE Full text] [doi: 10.3399/bjgp20X710897] [Medline: 32661013]
37. RSV ComNet: Research on the burden of RSV in the community network. Nivel. 2019. URL: https://nivel.nl/en/RSVComNet [accessed 2020-08-17]
38. I-MOVE-COVID-19. I-MOVE. 2020. URL: https://www.imoveflu.org/i-move-covid-19/ [accessed 2020-08-17]

XSL•FO
RenderX

39. Greenhalgh T, Thompson P, Weiringa S, Neves AL, Husain L, Dunlop M, et al. What items should be included in an early warning score for remote assessment of suspected COVID-19? Qualitative and Delphi study. BMJ Open 2020 Nov 12;10(11):e042626 [FREE Full text] [doi: 10.1136/bmjopen-2020-042626] [Medline: 33184088]
40. de Lusignan S, Crawford L, Munro N. Creating and using real-world evidence to answer questions about clinical effectiveness. J Innov Health Inform 2015 Nov 04;22(3):368-373 [FREE Full text] [doi: 10.14236/jhi.v22i3.177] [Medline: 26577427]
41. Fielding JE. Influenza vaccine effectiveness using the test-negative design: Comparability and methodological challenges in meta-analyses. Lancet Respir Med 2017 Mar;5(3):161-162. [doi: 10.1016/S2213-2600(17)30044-9] [Medline: 28189521]
42. Shi M, An Q, Ainslie KEC, Haber M, Orenstein WA. A comparison of the test-negative and the traditional case-control study designs for estimation of influenza vaccine effectiveness under nonrandom vaccination. BMC Infect Dis 2017 Dec 08;17(1):757 [FREE Full text] [doi: 10.1186/s12879-017-2838-2] [Medline: 29216845]
43. Cross JW, Joy M, McGee C, Akinyemi O, Gatenby P, de Lusignan S. Adverse events of interest vary by influenza vaccine type and brand: Sentinel network study of eight seasons (2010-2018). Vaccine 2020 May 08;38(22):3869-3880 [FREE Full text] [doi: 10.1016/j.vaccine.2020.03.034] [Medline: 32299719]
44. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: An alternative to standard epidemiological study designs. BMJ 2016 Sep 12;354:i4515 [FREE Full text] [doi: 10.1136/bmj.i4515] [Medline: 27618829]
45. Foppa IM, Haber M, Ferdinands JM, Shay DK. The case test-negative design for studies of the effectiveness of influenza vaccine. Vaccine 2013 Jun 26;31(30):3104-3109. [doi: 10.1016/j.vaccine.2013.04.026] [Medline: 23624093]
46. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
47. Royal College of GPs Research and Surveillance Centre. UK Health Data Research Alliance. URL: https://ukhealthdata.org/members/rcgp/ [accessed 2020-08-17]
48. Using Oxford-RCGP RSC for observational studies. Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub. URL: https://orchid.phc.ox.ac.uk/index.php/orchid-data/ [accessed 2021-01-26]
49. de Lusignan S. Royal College of General Practitioners (RCGP), Research and Surveillance Centre (RSC) quinquagenarian (QQG) practice network. Creating a longitudinal linked sentinel database of 50 years clinical and virology data and prospective research platform. Europe PMC. URL: https://europepmc.org/grantfinder/grantdetails?query=pi%3A%22de%20Lusignan%20S%22%20gid%3A%22212763%22%20ga%3A%22Wellcome%20Trust%22 [accessed 2021-01-29]
50. Liyanage H, Liaw S, Jonnagaddala J, Hinton W, de Lusignan S. Common data models (CDMs) to enhance international big data analytics: A diabetes use case to compare three CDMs. Stud Health Technol Inform 2018;255:60-64. [Medline: 30306907]
51. EHDEN: European Health Data and Evidence Network. Innovative Medicines Initiative. URL: https://www.imi.europa.eu/projects-results/project-factsheets/ehden [accessed 2020-08-17]
52. Thompson RN. Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. J Clin Med 2020 Feb 11;9(2):498 [FREE Full text] [doi: 10.3390/jcm9020498] [Medline: 32054124]
53. OxFile - Large File Exchange Service. University of Oxford IT Services. URL: https://help.it.ox.ac.uk/services/oxfile/index [accessed 2020-08-17]
54. Altamirano J, Govindarajan P, Blomkalns AL, Kushner LE, Stevens BA, Pinsky BA, et al. Assessment of sensitivity and specificity of patient-collected lower nasal specimens for severe acute respiratory syndrome coronavirus 2 testing. JAMA Netw Open 2020 Jun 01;3(6):e2012005 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12005] [Medline: 32530469]
55. Fleming DM, Elliot AJ. Lessons from 40 years' surveillance of influenza in England and Wales. Epidemiol Infect 2008 Jul;136(7):866-875. [doi: 10.1017/S0950268807009910] [Medline: 18047750]
56. Hardelid P, Fleming DM, Andrews N, Barley M, Durnall H, Mangtani P, et al. Effectiveness of trivalent and pandemic influenza vaccines in England and Wales 2008-2010: Results from a cohort study in general practice. Vaccine 2012 Feb 08;30(7):1371-1378. [doi: 10.1016/j.vaccine.2011.12.038] [Medline: 22178524]
57. Elliot AJ, Bermingham A, Charlett A, Lackenby A, Ellis J, Sadler C, et al. Self-sampling for community respiratory illness: A new tool for national virological surveillance. Euro Surveill 2015 Mar 12;20(10):21058 [FREE Full text] [doi: 10.2807/1560-7917.es2015.20.10.21058] [Medline: 25788252]
58. Hinton W, McGovern A, Coyle R, Han TS, Sharma P, Correa A, et al. Incidence and prevalence of cardiovascular disease in English primary care: A cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC). BMJ Open 2018 Aug 20;8(8):e020282 [FREE Full text] [doi: 10.1136/bmjopen-2017-020282] [Medline: 30127048]
59. Casalegno J, Eibach D, Valette M, Enouf V, Daviaud I, Behillil S, et al. Performance of influenza case definitions for influenza community surveillance: Based on the French influenza surveillance network GROG, 2009-2014. Euro Surveill 2017 Apr 06;22(14):30504 [FREE Full text] [doi: 10.2807/1560-7917.ES.2017.22.14.30504] [Medline: 28422004]

XSL•FO
RenderX

Abbreviations

AEI: adverse event of interest
DECISION-COVID: DEfining the CharacterIStIcs Of Individuals with suspected Novel COronaVIrus Disease and risk factors for development of the disease
DRIVE: Development of Robust and Innovative Vaccine Effectiveness
EMIS: Egton Medical Information Systems
FAIR: Findable, Accessible, Interoperable, and Reusable
HOCO: household contacts of confirmed COVID-19 cases
ILI: influenza-like illness
I-MOVE: Influenza–Monitoring Vaccine Effectiveness in Europe
MAINROUTE: Monitoring Attendance, INvestigation, Referral, and OUTcomEs in primary care: impact of and recovery from COVID-19 lockdown
NHS: National Health Service
ORCHID: Oxford–Royal College of General Practitioners Clinical Informatics Digital Hub
PHE: Public Health England
PRINCIPLE: Platform Randomised trial of INterventions against COVID-19 In older peoPLE
RAPTOR-C19: RAPid community Testing fOR COVID-19
RECAP: Remote COVID-19 Assessment in Primary Care
RCGP: Royal College of General Practitioners
RSC: Research and Surveillance Centre
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms

Edited by G Eysenbach; submitted 15.09.20; peer-reviewed by B Bente; comments to author 10.11.20; revised version received 13.11.20; accepted 08.12.20; published 19.02.21.

Please cite as:
de Lusignan S, Lopez Bernal J, Byford R, Amirthalingam G, Ferreira F, Akinyemi O, Andrews N, Campbell H, Dabrera G, Deeks A, Elliot AJ, Krajenbrink E, Liyanage H, McGagh D, Okusi C, Parimalanathan V, Ramsay M, Smith G, Tripathy M, Williams J, Victor W, Zambon M, Howsam G, Nicholson BD, Tzortziou Brown V, Butler CC, Joy M, Hobbs FDR
Influenza and Respiratory Virus Surveillance, Vaccine Uptake, and Effectiveness at a Time of Cocirculating COVID-19: Protocol for the English Primary Care Sentinel System for 2020-2021
JMIR Public Health Surveill 2021;7(2):e24341
URL: http://publichealth.jmir.org/2021/2/e24341/
doi:10.2196/24341
PMID:33605892

©Simon de Lusignan, Jamie Lopez Bernal, Rachel Byford, Gayatri Amirthalingam, Filipa Ferreira, Oluwafunmi Akinyemi, Nick Andrews, Helen Campbell, Gavin Dabrera, Alexandra Deeks, Alex J Elliot, Else Krajenbrink, Harshana Liyanage, Dylan McGagh, Cecilia Okusi, Vaishnavi Parimalanathan, Mary Ramsay, Gillian Smith, Manasa Tripathy, John Williams, William Victor, Maria Zambon, Gary Howsam, Brian David Nicholson, Victoria Tzortziou Brown, Christopher C Butler, Mark Joy, FD Richard Hobbs. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 19.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Web-Based System for Exploring Cancer Risk With Long-term Use of Drugs: Logistic Regression Approach

Hsuan-Chia Yang[1,2,3], PhD; Md Mohaimenul Islam[1,2,3], PhD; Phung Anh Alex Nguyen[2], PhD; Ching-Huan Wang[1], MSc; Tahmina Nasrin Poly[1,2,3], MSc; Chih-Wei Huang[2], PhD; Yu-Chuan Jack Li[1,2,3,4,5], MD, PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

[3]Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[4]Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[5]TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Yu-Chuan Jack Li, MD, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
15 F, No. 172-1, Sec 2, Kellung Rd
Da'an District
Taipei, 110
Taiwan
Phone: 886 2 27361661 ext 7600
Email: jaak88@gmail.com

Abstract

Background: Existing epidemiological evidence regarding the association between the long-term use of drugs and cancer risk remains controversial.

Objective: We aimed to have a comprehensive view of the cancer risk of the long-term use of drugs.

Methods: A nationwide population-based, nested, case-control study was conducted within the National Health Insurance Research Database sample cohort of 1999 to 2013 in Taiwan. We identified cases in adults aged 20 years and older who were receiving treatment for at least two months before the index date. We randomly selected control patients from the patients without a cancer diagnosis during the 15 years (1999-2013) of the study period. Case and control patients were matched 1:4 based on age, sex, and visit date. Conditional logistic regression was used to estimate the association between drug exposure and cancer risk by adjusting potential confounders such as drugs and comorbidities.

Results: There were 79,245 cancer cases and 316,980 matched controls included in this study. Of the 45,368 associations, there were 2419, 1302, 662, and 366 associations found statistically significant at a level of $P<.05$, $P<.01$, $P<.001$, and $P<.0001$, respectively. Benzodiazepine derivatives were associated with an increased risk of brain cancer (adjusted odds ratio [AOR] 1.379, 95% CI 1.138-1.670; $P=.001$). Statins were associated with a reduced risk of liver cancer (AOR 0.470, 95% CI 0.426-0.517; $P<.0001$) and gastric cancer (AOR 0.781, 95% CI 0.678-0.900; $P<.001$). Our web-based system, which collected comprehensive data of associations, contained 2 domains: (1) the drug and cancer association page and (2) the overview page.

Conclusions: Our web-based system provides an overview of comprehensive quantified data of drug-cancer associations. With all the quantified data visualized, the system is expected to facilitate further research on cancer risk and prevention, potentially serving as a stepping-stone to consulting and exploring associations between the long-term use of drugs and cancer risk.

(JMIR Public Health Surveill 2021;7(2):e21401) doi:10.2196/21401

KEYWORDS

cancer; risk; prevention; chemoprevention; long-term–use drugs; drug; epidemiology; temporal model; modeling; web-based system

XSL•FO
RenderX

Introduction

In recent decades, the prevalence of chronic medical conditions such as arthritis, osteoporosis, diabetes, hypertension, and cardiovascular disease has increased [1]. Patients with multimorbidity are more likely to have multidrug treatments; it is, in fact, common in older populations. Patients also may need a longer duration of treatment to get rid of these conditions, which leaves these patients vulnerable to unwanted side effects [2-4]. Multiple studies have already reported that long-term use of drugs has been increasing tremendously [5,6]. Therefore, a growing concern regarding the safety issues associated with long-term prescriptions has recently attracted widespread media attention. Physicians are often asked about the appropriateness of long-term therapy for specific patients.

There has been a significant rise in the number of published studies in which commonly used medications were found to increase or decrease the risk of cancer [7]. Long-term use of insulin seems to be associated with an increased risk of pancreas, liver, kidney, and stomach cancers [8], whereas a protective association was observed between metformin use and colorectal cancer risk in patients with diabetes mellitus [9]. Additionally, antihypertensive drugs are associated with an increased risk of skin cancer [10], but they might have a possible beneficial effect on breast cancer risk [11]. However, there has been substantial controversy about these studies' validity, and investigations with varying study designs and populations have often arrived at different conclusions. A prudent decision is immediately needed in clinical practices because the use of commonly prescribed medications has been increasing, and the resulting burden of cancer can be substantial at the population level [12].

Big data approaches seem to offer an immense opportunity to generate strong evidence for taking insightful clinical and public health action [13]. Data from electronic medical records and other extended patient registries have been offering expanded research power, especially for analytic studies aiming at association, classification, and prediction [14]. Epidemiological studies have already established a temporal relationship between drugs and diseases and have evaluated a wide range of outcomes. Clinical knowledge is constantly developing as new drug-disease discoveries are made and practices are changed. The knowledge of these associations is valuable but often buried in texts within a range of published literature. Through a web-based approach, information associated with drugs and disease risk can be a great source to understanding the magnitude of the risk between them. The availability of a wide range of associations may be valuable for a variety of applications, including clinical decision support (eg, treatment recommendation), information retrieval, and data mining. We therefore developed a web-based system focusing on the long-term use of commonly prescribed medications, including antihypertensives, antihyperlipidemics, antidiabetics, antihyperuricemics, anxiolytics, hypnotics, sedatives, and nonsteroidal anti-inflammatory drugs (NSAIDs) and 18 different kinds of cancer risk. We developed a system sorted by age, gender, and duration of therapy for exploring cancer risk with commonly used medications.

Methods

Ethical Standard

This study is part of a larger project aimed at assessing the effect of most common medications on 20 cancer sites using a population-based nested case-control design. The National Health Insurance Research Database (NHIRD) safeguards the privacy and confidentiality of all beneficiaries and transfers health insurance data to health researchers after ethical approval has been obtained. In this analysis, access to the NHIRD was approved by the Taipei Medical University Joint Institutional Review Board.

Setting and Data Source

This case-control study was carried out using records from the Taiwan NHIRD, which was established in 1995 and has collects all claims of beneficiaries under the National Health Insurance (NHI) program. The program covers more than 99% of the total population (a total of 23,430,000) and has contracted with 97% of the hospitals and clinics in Taiwan. The NHIRD comprised claims data of 2,000,000 individuals randomly selected from all insured enrollees. This sample represents the original medical claims for all residents in Taiwan covered under the NHI program. The database included specific data on medications prescribed, laboratory and diagnostic test data, dates of visits, lengths of hospitalization, and diagnoses. Diagnoses were coded according to the International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM). Drugs were coded based on the World Health Organization Anatomical Therapeutic Chemical (ATC) classifications (WHO Collaborating Centre for Drug Statistics Methodology ATC/DDD Index). The database used in this study can be interlinked by the scrambled, unique, individual personal identification number.

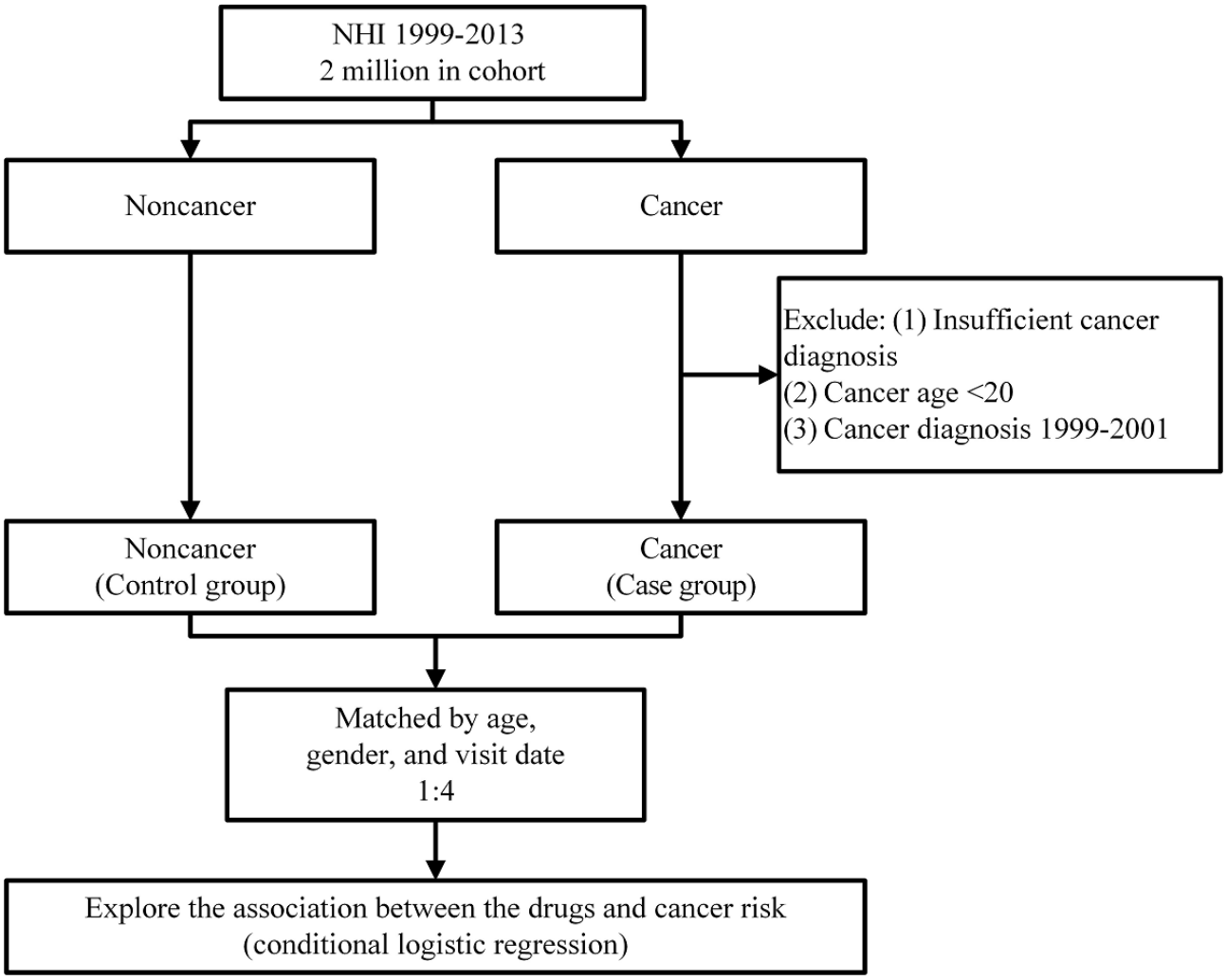
Case and Control Selection Criteria

We identified cases in adults who were aged 20 years or older and had received treatment at least two months before the index date. The index date was defined as the date of a cancer diagnosis. We used the ICD-9-CM to identify patients with cancer as cases. Among the NHIRD cases, eligibility criteria for case patients were (1) registration as patients with cancer in the catastrophic illness file, (2) diagnosis of primary cancer in inpatient admission, (3) treatment with any cancer drug from outpatient visits or inpatient admission, (4) a cancer-specific procedure from outpatient visits or inpatient admission, and (5) more than 4 cancer-specific examinations or more than 1 cancer-related procedure (radiotherapy, chemotherapy, or treatment tracking of cancer) from outpatient visits or inpatient admission.

We randomly selected control patients from patients without a cancer diagnosis during the 15 years (1999-2013) of the study period. Case and control patients were matched 1:4 based on age, sex, and visit date (Figure 1).

XSL•FO
RenderX

Figure 1. Workflow of the case-control study design. NHI: National Health Insurance.



Primary Outcomes

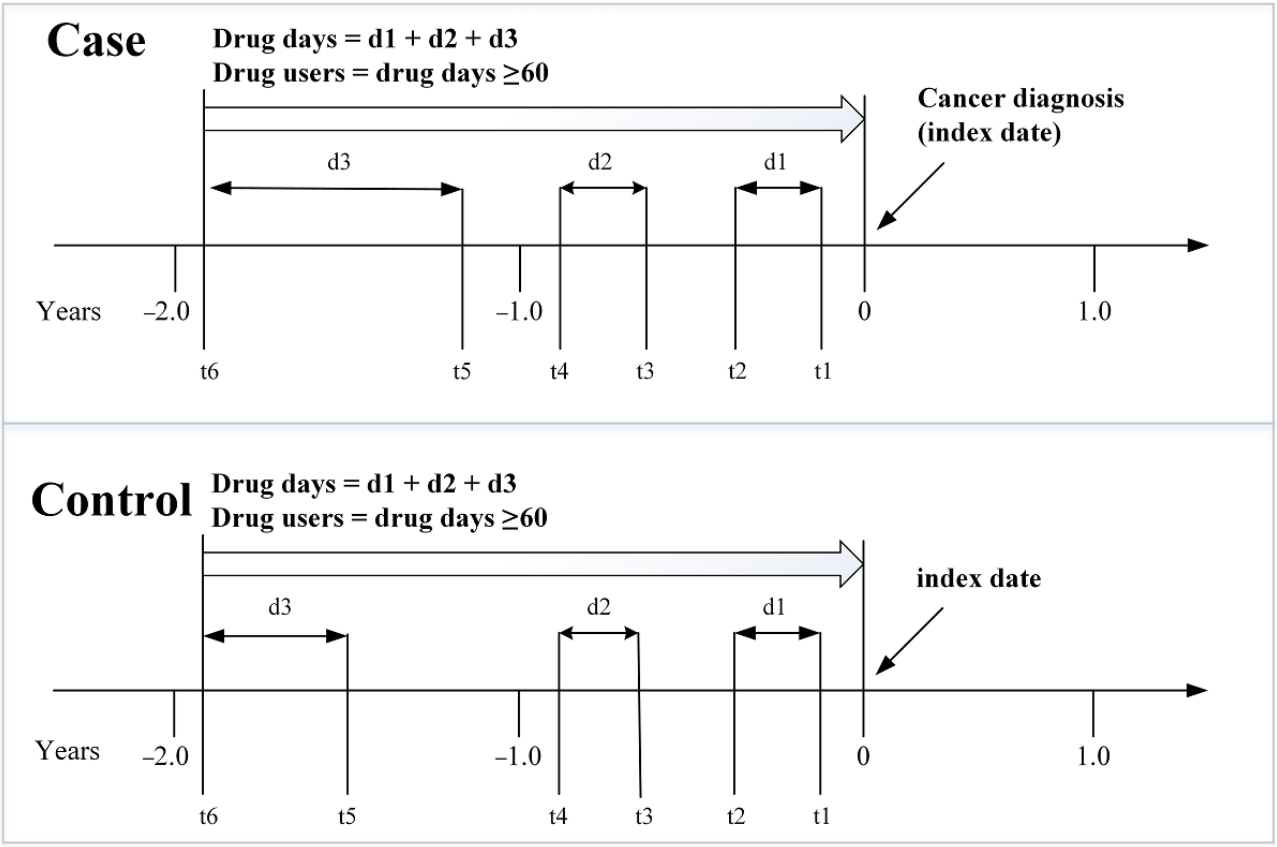
We quantified the risks of common cancers in Taiwan, comparing patients treated with (1) antihypertensives; (2) antihyperlipidemics; (3) antidiabetics; (4) antihyperuricemics; (5) anxiolytics, hypnotics, and sedatives; or (6) NSAIDs against those not prescribed any of these medications. We investigated the following cancers using their corresponding ICD-9-CM codes: oral cancer (140-149.xx, excluding 142.xx and 147.xx), esophageal cancer (150.xx), gastric cancer (151.xx), colorectal cancer (153.xx, rectum 154.xx), liver cancer (155.xx), pancreatic cancer (157.xx), lung cancer (162.xx), skin cancer (172-173.xx), female breast cancer (174.xx), cervical cancer (180.xx), endometrial cancer (182.xx), ovarian cancer (183.xx), prostate cancer (185.xx), bladder cancer (188.xx), kidney cancer (189.xx), brain cancer (191.xx), thyroid cancer (193.xx), non-Hodgkin disease (200.xx, 202.xx, 203.xx), leukemia (204-208.xx), and all cancers (140-208.xx).

Use of Drugs

We defined the index date as the date of a cancer diagnosis. The drug exposure was analyzed only before the index date, and we defined drug users as those who filled prescriptions of at least 60 days during admissions and outpatient visits within the 2 years before the index date (Figure 2). This definition was considered for 6 long-term drug groups, namely (1) antihypertensives; (2) antihyperlipidemics; (3) antidiabetics; (4) antihyperuricemics; (5) antianxiety agents, hypnotics, and sedatives; and (6) NSAIDs. Those having no exposure to or receiving these drugs for less than two months were classified as nondrug users.

XSL•FO
RenderX

Figure 2. Drugs exposure.



Potential Confounders

Comorbidities and medications identified as confounders were adjusted in this study. Comorbidities were defined using the Charlson Comorbidity Index [15], retrieved from outpatient visits before the index date. Any of the comorbidities as confounders were required to be diagnosed at least twice within the 2 years (720 days) before the index date, except for chronic pulmonary disease, which required a diagnosis made more than 4 times. In addition, aspirin (ATC: B01AC06), statins (ATC: C10AA), and metformin (ATC: A10BA2) were also confounders because they might potentially influence the risk of cancers. Exposure to a confounding drug was defined as treatment with the drug for at least 60 days during the 2 years before the index date.

Statistical Analysis

The McNamara test and paired *t* test were used to test the difference between the case and control groups [16]. Next, conditional logistic regression was conducted to estimate the association between drug exposure and cancer risk by adjusting potential confounders [17]. Table 1 shows our study variables, and conditional logistic regression (temporal model) was adopted to investigate the association between the long-term use of drugs and cancer risk. Age was divided into 4 categories: 20 to 39 years, 40 to 64 years, ≥65 years, and ≥20 years. Gender was classified as male, female, and both. The basic equation of the model was as below, and it may have been slightly modified in different study drug groups.

XSL•FO
RenderX

Table 1. Study variables.

Variables	Type	Descriptive	Statistical model	Definition
ID	Nominal	N/A[a]	No	N/A
Age	Ordinal	N/A	No	N/A
Sex	Binomial	1: Yes, 0: No	No	N/A
Study drug (exposure)	Binomial	1: Yes, 0: No	Yes	Independent
Cancer (outcome)	Binomial	1: Yes, 0: No	Yes	Dependent
Myocardial infarction	Binomial	1: Yes, 0: No	Yes	Confounding
Congestive heart failure	Binomial	1: Yes, 0: No	Yes	Confounding
Peripheral vascular disease	Binomial	1: Yes, 0: No	Yes	Confounding
Cerebrovascular disease	Binomial	1: Yes, 0: No	Yes	Confounding
Dementia	Binomial	1: Yes, 0: No	Yes	Confounding
Chronic pulmonary disease	Binomial	1: Yes, 0: No	Yes	Confounding
Rheumatic disease	Binomial	1: Yes, 0: No	Yes	Confounding
Peptic ulcer disease	Binomial	1: Yes, 0: No	Yes	Confounding
Liver disease (mild, moderate, and severe)	Binomial	1: Yes, 0: No	Yes	Confounding
Diabetes (with or without chronic complication)	Binomial	1: Yes, 0: No	Yes	Confounding
Hemiplegia or paraplegia	Binomial	1: Yes, 0: No	Yes	Confounding
Renal disease	Binomial	1: Yes, 0: No	Yes	Confounding
CCI[b] scores	Ordinal	N/A	Yes	Confounding
Metformin	Binomial	1: Yes, 0: No	Yes	Confounding
Aspirin	Binomial	1: Yes, 0: No	Yes	Confounding
Statin	Binomial	1: Yes, 0: No	Yes	Confounding
Matching number (case match control)	Nominal	N/A	Yes	Stratified

[a]N/A: not applicable.

[b]Charlson Comorbidity Index.

Data analysis and results were performed using SAS software (version 9.4; SAS Institute) [18]. The results were expressed in adjusted odds ratios (AORs), which is e^{β_1} with different confidence intervals, like 95%, 99%, and 99.9%. All statistical tests were 2-sided.

Web-Based System

After analyzing the associations between the long-term use of drugs and cancer risk, we built a web-based system to include all associations [19]. The back end of this web-based system includes a server, database, and application. We used Apache (Apache Software Foundation), MySQL (Oracle Corporation), and PHP (Hypertext Preprocessor) framework for developing the server. Apache is the most commonly used web server software and supports a variety of compiled modules. MySQL is a relational database management system, and PHP is a programming language designed primarily for web development. In this study, we inputted the associations into the database (MySQL) and used PHP to access and process the database. The web-based interface was designed accordingly with HTML, CSS, Bootstrap, JavaScript, and React.

Results

Baseline Characteristics

We identified 79,245 participants newly diagnosed with cancer between 2002 and 2013 from 2 million people (Table 2). After matching each case with 4 controls, we included 316,980 matched control patients in this study. The mean age was 59.2 years for both the case and control groups. A slight majority of the participants were male (201,295/396,225, 50.80%), and most participants (202,544/396,225, 51.12%) were aged 40 to 64 years. The prevalence of peptic ulcer disease (12,760/79,245, 16.10% vs 34,283/316,980, 10.82%) and liver disease (11,671/79,245, 14.73% vs 20,647/316,980, 6.51%) in the case group was higher than that in the control group. After associations between long-term use of drugs and cancer risk were comprehensively analyzed and stratified by age and sex, we obtained 45,368 associations in total, of which 2419, 1302, 662, and 366 associations were found statistically significant at a level of $P<.05$, $P<.01$, $P<.001$, and $P<.0001$, respectively (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Baseline characteristics of case patients and control patients.

Characteristics	Case patients (with cancer) (n=79,245)	Control patients (without cancer) (n=316,980)	*P* value
Age (years)			
Age (years), mean (SD)	59.20 (15.23)	59.21 (15.24)	N/A[a]
20-39, n (%)	8,292 (10.5)	33,168 (10.5)	N/A
40-64, n (%)	40,504 (51.1)	162,040 (51.1)	N/A
≥65, n (%)	30,449 (38.4)	121,772 (38.4)	N/A
Gender, n (%)			
Male	40,259 (50.8)	161,036 (50.8)	N/A
Female	38,986 (49.2)	155,944 (49.2)	N/A
Comorbid conditions, n (%)			
Myocardial infarction	372 (0.47)	1729 (0.55)	<.0001
Congestive heart failure	2557 (3.23)	9344 (2.95)	<.0001
Peripheral vascular disease	1143 (1.44)	4187 (1.32)	<.0001
Cerebrovascular disease	5306 (6.70)	22,609 (7.13)	<.0001
Dementia	1205 (1.52)	5522 (1.74)	<.0001
Chronic pulmonary disease	5951 (7.51)	20,423 (6.44)	<.0001
Rheumatic disease	1138 (1.44)	3901 (1.23)	<.0001
Peptic ulcer disease	12,760 (16.10)	34,283 (10.82)	<.0001
Liver disease	11,671 (14.73)	20,647 (6.51)	<.0001
Diabetes	9143 (11.54)	32,532 (10.26)	<.0001
Hemiplegia or paraplegia	490 (0.62)	2300 (0.73)	<.0001
Renal disease	2793 (3.52)	7453 (2.35)	<.0001
Other drugs, n (%)			
Metformin	8236 (10.39)	33,375 (10.53)	.07
Aspirin	9826 (12.40)	41,726 (13.16)	<.0001
Statin	8336 (10.52)	37,395 (11.80)	<.0001

[a]N/A: not applicable.

Web-Based System

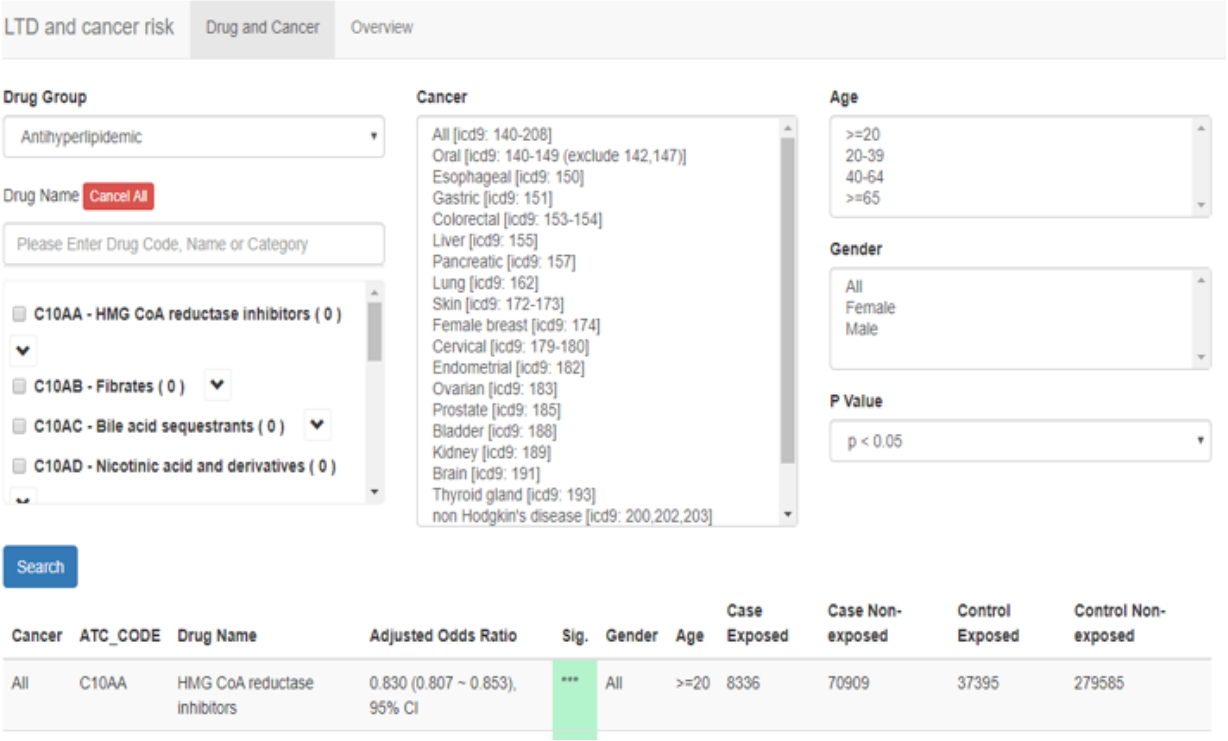
We successfully developed a web-based system [19], which contains 2 domains: (1) the drug and cancer page and (2) the overview page. The associations between long-term use of drugs and cancer diseases are shown in AORs with 95% CIs. The associations are visualized on the overview page, where researchers can compare and contrast the quantified personalized risk of multiple types of cancers for long-term users of the 6 groups of medications at the same time.

The Drug and Cancer Association Page

Drugs were categorized into 6 groups: (1) antihypertensives; (2) antihyperlipidemics; (3) antidiabetics; (4) antihyperuricemics; (5) NSAIDs; and (6) anxiolytics, hypnotics, and sedatives. As exemplified in Figure 3, the web-based system shows an AOR of 0.830 (95% CI 0.807-0.853) between statins (3-hydroxy-3-methyl-glutaryl coenzyme A [HMG-CoA] reductase inhibitors) and the overall cancer risk for those aged 20 years or older. Table 3 shows the associations between different drugs and cancers among different age groups gathered from the developed web-based system.

XSL•FO
RenderX

Figure 3. Display of drug and cancer risk.



XSL•FO
RenderX

Table 3. Associations between different drugs and cancers among different age groups.

Drug (ATC[a] code), cancer type, and age (years)	Adjusted odd ratio (95% CI)	Case patients, n		Control patients, n	
		Exposure	Nonexposure	Exposure	Nonexposure
Aspirin (B01AC06)					
All					
≥65	0.954** (0.923-0.985)	6964	23,485	28,665	93,131
40-64	0.871*** (0.831-0.913)	2825	37,679	12,925	149,091
20-39	1.000 (0.679-1.472)	37	8255	136	33,032
≥20	0.924*** (0.900-0.949)	9826	69,419	41,726	275,254
Metformin (A10BA02)					
Colorectal					
≥65	0.881* (0.794-0.979)	797	4532	3231	18,085
40-64	0.799*** (0.701-0.912)	456	4757	2076	18,776
20-39	0.448 (0.148-1.358)	4	776	28	3092
≥20	0.845*** (0.779-0.916)	1257	10,065	5335	39,953
Sitagliptin (A10BH01)					
Pancreatic					
≥65	1.901* (1.125-3.213)	28	632	50	2590
40-64	2.303* (1.109-4.781)	16	565	27	2297
20-39	N/A[b]	0	91	0	364
≥20	1.981** (1.298-3.024)	44	1288	77	5251
Benzodiazepine derivatives (N05)					
Brain					
≥65	1.090 (0.787-1.508)	78	201	256	860
40-64	1.456** (1.112-1.905)	111	379	287	1673
20-39	2.409** (1.364-4.257)	24	201	46	854
≥20	1.379** (1.138-1.670)	213	781	589	3387

[a]ATC: Anatomical Therapeutic Chemical.

[b]N/A: not applicable.

*P<.05.

**P<.01.

***P<.001.

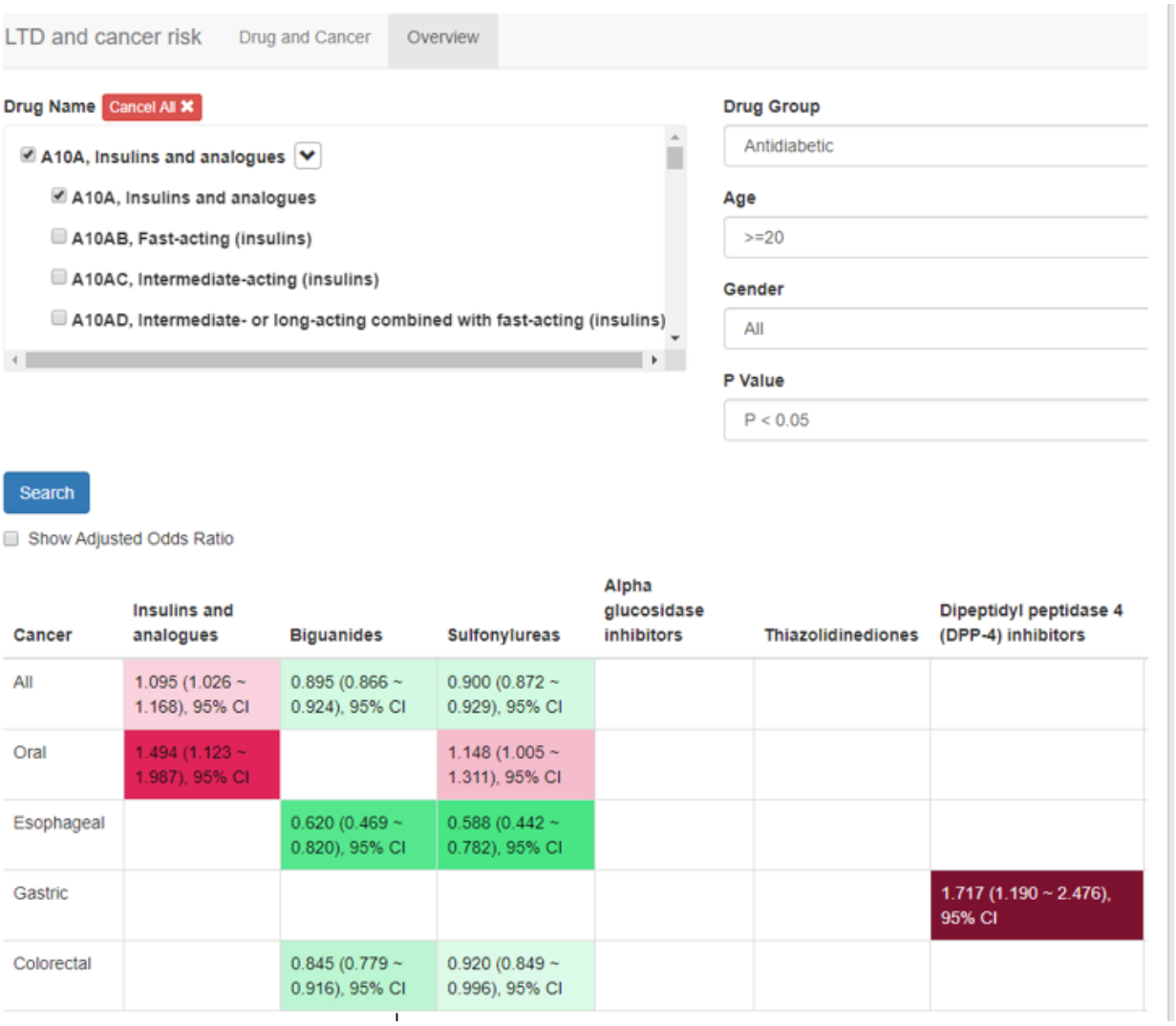
The Overview Page

The web-based system provided an overview of associations between cancers and medications sorted by age, gender, P value, and ATC class of medications (Figure 4). In the cells are AORs of each cancer for different medications, and a confidence interval of 95%, 99%, or 99.9% can be selected by users based on different P values (P<.05, P<.01, P<.001). The web-based system highlights the cells with different colors—from green to white to red—to demonstrate the direction and the extent of these drug-cancer associations. When a user clicks on the Show Adjusted OR button, the system is able to show all associations. However, if a drug-cancer association pair has a sample size less than 10 patients, the cell stays blank.

Green colors symbolize a significant association between a drug and a cancer with an AOR less than 1. The darker a green color is, the farther from 1 the AOR is. Red colors symbolize a significant association between a drug and a cancer with an AOR greater than 1. The darker a red color is, the farther from 1 the AOR is. White indicates no significant association between a cancer and a drug.

XSL•FO
RenderX

Figure 4. Display of overview.



Discussion

Main Outcomes

In this nationwide longitudinal retrospective study, we evaluated 79,245 patients with cancer and 316,980 control patients matched by variables including age, sex, and visit date at a 1:4 ratio from the NHIRD in Taiwan, including the follow-up data from 2001 to 2013 of 2 million individuals aged 20 years or older. This web-based system aimed to provide information of medication-cancer associations for users (researchers) to choose potentially clinically relevant ones for further studies (eg, a meta-analysis) and offered a filter by *P* value. We found aspirin and metformin were significantly associated with reduced cancer risk in those aged 40 to 64 years and 65 years or older, but no significant association was uncovered in those aged 20 to 39 years. A partial explanation for this may lie in the fact that the low prescribing rate or the low cancer incidence among those aged 20 to 39 years rendered it impossible for us to reject the null hypothesis that there were no associations between aspirin and all cancers or between metformin and colorectal cancer.

The long-term use of some drugs was associated with increased risk of certain cancers, such as sitagliptin with pancreatic cancer and benzodiazepines (BZDs) with brain cancer. For example, patients aged 40 to 64 years and 65 years or older treated with sitagliptin had a high risk for pancreatic cancer, but there was not sufficient information for us to estimate such risk among patients aged 20 to 39 years. On the contrary, those aged 20 to 39 years receiving BZDs had a higher risk of brain cancer (AOR 2.409, 95% CI 1.364-4.257; *P*=.003) compared with the overall population (AOR 1.379, 95% CI 1.138-1.670; *P*=.001), but there was no significant association between BZDs and brain cancer among those aged 65 years or older.

Biological Mechanisms

Despite mechanisms between the long-term use of drugs and cancer risk remaining not well understood, our findings were consistent with possible mechanisms proposed in previous studies. Aspirin, metformin, and statins are examples of this. According to previous studies, aspirin reduces prostaglandin generation, which is associated with decreased cellular proliferation, by inhibiting cyclooxygenase isozymes [20]. Metformin activates adenosine monophosphate–activated protein kinase (AMPK), which is a major sensor of whole-body energy metabolism, and activation of AMPK helps to reduce the proliferation of human colon cancer cells [21]. Statins can also decrease intracellular cholesterol production by inhibiting HMG-CoA and may also limit the cellular proliferation required

XSL•FO
RenderX

for cancer growth [22]. In this study, aspirin, metformin, and statins were found to have significant associations with overall cancer risk, with AORs of 0.924 (95% CI 0.900-0.949; *P*<.001), 0.845 (95% CI 0.779-0.916; *P*<.001), and 0.830 (95% CI 0.807-0.853; *P*<.001).

Additionally, sitagliptin has been suggested to have an association with elevated risk of pancreatitis and pancreatic cancer [23]. In our results, sitagliptin was also significantly associated with pancreatic cancer risk (AOR 1.981, 95% CI 1.298-3.024; *P*=.002). Another intriguing finding in our study was that cancer risk was significantly associated with angiotensin-converting enzyme inhibitors (ACEIs) (AOR 0.854, 95% CI 0.829-0.880; *P*<.001) but not with angiotensin II antagonists (ARBs), even though both ACEIs and ARBs pharmacologically share a similar pathway. A possible explanation is that the stronger inhibitory effect of ACEIs compared with ARBs on angiotensinogen may be associated with cancer risk. However, our system determined associations only, not causalities, between long-term use of medications and risk of cancers. For some of the significant associations, questions about their mechanisms are still left unanswered. Since the results of this study were associations that did not determine the causality, these associations will need further work to confirm mechanisms and causal relationship between long-term use of drugs and cancer risk.

Clinical Implication

Despite the immense investment in anticancer therapy, cancer remains the leading cause of death globally [24]. Development of an anticancer drug is resource intensive and takes an average of 13 years at a cost up to US $2.6 billion [25]. The rapidly growing cost and development time have made the pharmaceutical industry a less profitable choice for many investors [26]. Although little attention has been paid to identifying new chemoprevention drugs from existing available drugs, the strategy of using one drug to treat several indications has shown potential success and become an attractive proposition in many areas of medicine, especially in complex disorders [27].

Aspirin is widely used to treat fever and mild pain, but its long-term use may prevent development of squamous cell carcinoma [28], colorectal cancer [29], and hepatocellular carcinoma [30]. Statin use is associated with a reduced risk of pancreatic ductal adenocarcinoma [31] and hepatocellular carcinoma for patients with risk factors [22]. Moreover, metformin, an antidiabetic medication, has drawn attention, since it exhibits an effect on the prevention and treatment of cancers such as colorectal cancer as beneficial as an independent anticancer drug [32]. However, there has been substantial controversy about whether aspirin, statins, and metformin really have anticancer preventive or therapeutic effects on cancers, and often investigations with varying study designs and populations have reached different conclusions. A prudent decision is immediately needed in clinical practices because the use of commonly prescribed medications has been increasing and the resulting burden of cancer can be substantial at the population level [33]. Our study attempted to investigate the magnitude of cancer risk and the benefits of 6 groups of commonly prescribed medications using a large database and appropriate methodology. Our web-based system could potentially show hints of clinical interest for users such as researchers and health care professionals to propose new hypotheses and further undertake research to identify mechanisms or causalities of associations.

Strengths and Limitations

Strengths of this study include the retrospective study design, long-term follow-up, proper identification of case and control patients, and measurement of the magnitude of association between 6 commonly used groups of medications and cancer risks. Furthermore, confounding factors were appropriately adjusted to reduce the study bias.

We also acknowledge that our research has limitations that need to be addressed. First, drug adherence, self-payment, laboratory data, and lifestyles characteristics such as body mass index, smoking, and family history of cancer were unavailable in the NHIRD. Second, other risk factors for cancer, such as phenotype, genotype, and exposure type, might have influenced the results. Although we applied the match method and adjustment for numerous covariates to control confounding factors, it was impossible to eliminate all confounding factors, particularly indications. Third, all data were collected from the Taiwan NHIRD, and hence, the study population limited the generalization of the results to other countries with different ethnic distribution. Fourth, the results showed associations between the long-term use of drugs and cancer risk but not causation.

Moreover, we did not set a threshold for statistical significance at $0.05/45{,}368 \approx 1.10 \times 10^{-6}$ for multiple testing correction, given the large number of statistical tests and the highly selected patients—patients with cancer and long-term users of medications instead of the general population. Had we set the significance level at 1.10×10^{-6}, there would not have been enough significant associations to be useful or practical to users. Therefore, we offered in the web-based system a filter by *P* value, allowing users to choose a *P* value based on their own need for research. Moreover, considering that there might have been a small number of these highly selected patients, especially after we grouped by drug class, cancer type, age, and gender, we provided users with detailed information of sample sizes on the web-based system, showing the numbers of case and control patients either exposed or not exposed to the study medications.

Conclusion

This comprehensive retrospective study not only provides an overview of associations of cancer risk with 6 commonly prescribed groups of medications but also helps to narrow the gap in the currently insufficient research on the long-term safety of these medications. With all the quantified data visualized, the system is expected to further facilitate research on cancer risk and prevention. Since our findings have proposed only associations between cancers and long-term use of medications, further clinical trials and meta-analyses are required to assess and confirm their causality. This web-based system could potentially serve as a stepping-stone to exploring and consulting associations between long-term use of drugs and cancer risk.

XSL•FO
RenderX

Acknowledgments

This research is sponsored in part by the Ministry of Science and Technology (grant number: MOST 109-2222-E-038-002-MY2), the Ministry of Education (grant number: MOE 109-6604-001-400), and Taipei Medical University (grant number: TMU107-AE1-B18).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table.
[DOCX File , 17 KB - publichealth_v7i2e21401_app1.docx]

References

1. Diederichs C, Berger K, Bartels DB. The measurement of multiple chronic diseases--a systematic review on existing multimorbidity indices. J Gerontol A Biol Sci Med Sci 2011 Mar;66(3):301-311. [doi: 10.1093/gerona/glq208] [Medline: 21112963]
2. Jacobs E, Newton C, Thun M, Gapstur SM. Long-term use of cholesterol-lowering drugs and cancer incidence in a large United States cohort. Cancer Res 2011 Mar 01;71(5):1763-1771 [FREE Full text] [doi: 10.1158/0008-5472.CAN-10-2953] [Medline: 21343395]
3. Suissa S, Hutchinson T, Brophy J, Kezouh A. ACE-inhibitor use and the long-term risk of renal failure in diabetes. Kidney Int 2006 Mar;69(5):913-919 [FREE Full text] [doi: 10.1038/sj.ki.5000159] [Medline: 16518351]
4. Jessen F, Kaduszkiewicz H, Daerr M, Bickel H, Pentzek M, Riedel-Heller S, et al. Anticholinergic drug use and risk for dementia: target for dementia prevention. Eur Arch Psychiatry Clin Neurosci 2010 Nov;260(Suppl 2):S111-S115. [doi: 10.1007/s00406-010-0156-4] [Medline: 20960005]
5. Islam MM, Iqbal U, Walther B, Atique S, Dubey NK, Nguyen P, et al. Benzodiazepine Use and Risk of Dementia in the Elderly Population: A Systematic Review and Meta-Analysis. Neuroepidemiology 2016;47(3-4):181-191 [FREE Full text] [doi: 10.1159/000454881] [Medline: 28013304]
6. Benner J, Glynn R, Mogun H, Neumann P, Weinstein M, Avorn J. Long-term persistence in use of statin therapy in elderly patients. ACC Curr J Rev 2002 Nov;11(6):36. [doi: 10.1016/s1062-1458(02)00914-5]
7. Moysich KB, Beehler GP, Zirpoli G, Choi J, Baker JA. Use of Common Medications and Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev 2008 Jul 15;17(7):1564-1595. [doi: 10.1158/1055-9965.epi-07-2828]
8. Karlstad O, Starup-Linde J, Vestergaard P, Hjellvik V, Bazelier M, Schmidt M, et al. Use of insulin and insulin analogs and risk of cancer - systematic review and meta-analysis of observational studies. Curr Drug Saf 2013 Nov;8(5):333-348 [FREE Full text] [doi: 10.2174/15680266113136660067] [Medline: 24215311]
9. Singh S, Singh H, Singh PP, Murad MH, Limburg PJ. Antidiabetic Medications and the Risk of Colorectal Cancer in Patients with Diabetes Mellitus: A Systematic Review and Meta-analysis. Cancer Epidemiol Biomarkers Prev 2013 Sep 16;22(12):2258-2268. [doi: 10.1158/1055-9965.epi-13-0429]
10. Gandini S, Palli D, Spadola G, Bendinelli B, Cocorocchio E, Stanganelli I, et al. Anti-hypertensive drugs and skin cancer risk: a review of the literature and meta-analysis. Crit Rev Oncol Hematol 2018 Feb;122:1-9. [doi: 10.1016/j.critrevonc.2017.12.003] [Medline: 29458778]
11. Ni H, Rui Q, Zhu X, Yu Z, Gao R, Liu H. Antihypertensive drug use and breast cancer risk: a meta-analysis of observational studies. Oncotarget 2017 Sep 22;8(37):62545-62560 [FREE Full text] [doi: 10.18632/oncotarget.19117] [Medline: 28977968]
12. Ioannidis J, Zhou Y, Chang C, Schully S, Khoury M, Freedman A. Potential increased risk of cancer from commonly used medications: an umbrella review of meta-analyses. Ann Oncol 2014 Jan;25(1):16-23 [FREE Full text] [doi: 10.1093/annonc/mdt372] [Medline: 24310915]
13. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
14. Wang YH, Nguyen PA, Islam MM, Li YC, Yang HC. Development of Deep Learning Algorithm for Detection of Colorectal Cancer in EHR Data. 2019 Presented at: 17th World Congress on Medical and Health Informatics (MEDINFO 2019); Aug 25-30, 2019; Lyon, France.
15. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992 Jun;45(6):613-619. [doi: 10.1016/0895-4356(92)90133-8] [Medline: 1607900]
16. Rosner B. Fundamentals of Biostatistics. Toronto, ON, Canada: Nelson Education; 2015.
17. Choi N, Park B, Jeong S, Yu K, Yoon B. Nonaspirin Nonsteroidal Anti-inflammatory Drugs and Hemorrhagic Stroke Risk. Stroke 2008 Mar;39(3):845-849. [doi: 10.1161/strokeaha.107.497040]

XSL•FO
RenderX

18. Thrift AG, McNeil JJ, Forbes A, Donnan GA. Risk of primary intracerebral haemorrhage associated with aspirin and non-steroidal anti-inflammatory drugs: case-control study. BMJ 1999 Mar 20;318(7186):759-764 [FREE Full text] [doi: 10.1136/bmj.318.7186.759] [Medline: 10082697]
19. LTD and cancer risk. Taiwan Artificial Intelligence for Medicine & Health Innovations Center. URL: http://ltd-cancer.aimhi.tw/ [accessed 2021-02-09]
20. Sostres C, Gargallo CJ, Lanas A. Aspirin, cyclooxygenase inhibition and colorectal cancer. World J Gastrointest Pharmacol Ther 2014 Feb 06;5(1):40-49 [FREE Full text] [doi: 10.4292/wjgpt.v5.i1.40] [Medline: 24605250]
21. Higurashi T, Hosono K, Takahashi H, Komiya Y, Umezawa S, Sakai E, et al. Metformin for chemoprevention of metachronous colorectal adenoma or polyps in post-polypectomy patients without diabetes: a multicentre double-blind, placebo-controlled, randomised phase 3 trial. Lancet Oncol 2016 Apr;17(4):475-483. [doi: 10.1016/s1470-2045(15)00565-3]
22. Singh S, Singh PP, Singh AG, Murad MH, Sanchez W. Statins are associated with a reduced risk of hepatocellular cancer: a systematic review and meta-analysis. Gastroenterology 2013 Feb;144(2):323-332. [doi: 10.1053/j.gastro.2012.10.005] [Medline: 23063971]
23. Elashoff M, Matveyenko AV, Gier B, Elashoff R, Butler PC. Pancreatitis, pancreatic, and thyroid cancer with glucagon-like peptide-1-based therapies. Gastroenterology 2011 Jul;141(1):150-156 [FREE Full text] [doi: 10.1053/j.gastro.2011.02.018] [Medline: 21334333]
24. Heron M, Anderson RN. Changes in the Leading Cause of Death: Recent Patterns in Heart Disease and Cancer Mortality. NCHS Data Brief 2016 Aug;254:1-8 [FREE Full text] [Medline: 27598767]
25. Avorn J. The $2.6 billion pill--methodologic and policy considerations. N Engl J Med 2015 May 14;372(20):1877-1879. [doi: 10.1056/NEJMp1500848] [Medline: 25970049]
26. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov 2019 Jan;18(1):41-58. [doi: 10.1038/nrd.2018.168] [Medline: 30310233]
27. Hopkins AL. Drug discovery: Predicting promiscuity. Nature 2009 Nov 12;462(7270):167-168. [doi: 10.1038/462167a] [Medline: 19907483]
28. Pandeya N, Olsen C, Thompson B, Dusingize J, Neale R, Green A, QSkin Study. Aspirin and nonsteroidal anti-inflammatory drug use and keratinocyte cancers: a large population-based cohort study of skin cancer in Australia. Br J Dermatol 2019 Oct;181(4):749-760. [doi: 10.1111/bjd.17938] [Medline: 30920641]
29. García Rodríguez LA, Soriano-Gabarró M, Bromley S, Lanas A, Cea Soriano L. New use of low-dose aspirin and risk of colorectal cancer by stage at diagnosis: a nested case-control study in UK general practice. BMC Cancer 2017 Sep 07;17(1):637 [FREE Full text] [doi: 10.1186/s12885-017-3594-9] [Medline: 28882113]
30. Simon TG, Ma Y, Ludvigsson JF, Chong DQ, Giovannucci EL, Fuchs CS, et al. Association Between Aspirin Use and Risk of Hepatocellular Carcinoma. JAMA Oncol 2018 Dec 01;4(12):1683-1690 [FREE Full text] [doi: 10.1001/jamaoncol.2018.4154] [Medline: 30286235]
31. Choi JH, Lee SH, Huh G, Chun JW, You MS, Paik WH, et al. The association between use of statin or aspirin and pancreatic ductal adenocarcinoma: A nested case-control study in a Korean nationwide cohort. Cancer Med 2019 Dec;8(17):7419-7430 [FREE Full text] [doi: 10.1002/cam4.2617] [Medline: 31637875]
32. Kamarudin MNA, Sarker MMR, Zhou J, Parhar I. Metformin in colorectal cancer: molecular mechanism, preclinical and clinical aspects. J Exp Clin Cancer Res 2019 Dec 12;38(1):491 [FREE Full text] [doi: 10.1186/s13046-019-1495-2] [Medline: 31831021]
33. Ioannidis J, Zhou Y, Chang C, Schully S, Khoury M, Freedman A. Potential increased risk of cancer from commonly used medications: an umbrella review of meta-analyses. Ann Oncol 2014 Jan;25(1):16-23 [FREE Full text] [doi: 10.1093/annonc/mdt372] [Medline: 24310915]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitors
AMPK: adenosine monophosphate–activated protein kinase
AOR: adjusted odds ratio
ARB: angiotensin II antagonist
ATC: Anatomical Therapeutic Chemical
BZD: benzodiazepine
HMG-CoA: 3-hydroxy-3-methyl-glutaryl coenzyme A
ICD-9-CM: International Classification of Disease, Ninth Revision, Clinical Modification
NHI: National Health Insurance
NHIRD: National Health Insurance Research Database
NSAID: nonsteroidal anti-inflammatory drug
PHP: Hypertext Preprocessor

XSL•FO
RenderX

Edited by G Eysenbach; submitted 13.06.20; peer-reviewed by YL Cheong, E Omaki; comments to author 22.09.20; revised version received 29.09.20; accepted 17.01.21; published 15.02.21.

Please cite as:
Yang HC, Islam MM, Nguyen PAA, Wang CH, Poly TN, Huang CW, Li YCJ
Development of a Web-Based System for Exploring Cancer Risk With Long-term Use of Drugs: Logistic Regression Approach
JMIR Public Health Surveill 2021;7(2):e21401
URL: http://publichealth.jmir.org/2021/2/e21401/
doi:10.2196/21401
PMID:33587043

©Hsuan-Chia Yang, Md Mohaimenul Islam, Phung Anh Alex Nguyen, Ching-Huan Wang, Tahmina Nasrin Poly, Chih-Wei Huang, Yu-Chuan Jack Li. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 15.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis

Katja Reuter[1,2*], PhD; Melissa L Wilson[3*], PhD; Meghan Moran[4], PhD; NamQuyen Le[5], MPH; Praveen Angyan[5], MS; Anuja Majmundar[6], MA, MBA, PhD; Elsi M Kaiser[7*], MS, PhD; Jennifer B Unger[2*], PhD

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, United States

[2]Institute for Health Promotion and Disease Prevention Research, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[3]Division of Disease Prevention, Policy and Global Health, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[4]Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[6]Economic and Health Policy Research, American Cancer Society, Washington, DC, United States

[7]Department of Linguistics, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Katja Reuter, PhD
Department of Public Health & Preventive Medicine
SUNY Upstate Medical University
766 Irving Avenue
Syracuse, NY, 13210
United States
Phone: 1 315 464 1520
Email: reuterk@upstate.edu

Related Article:

Correction of: https://publichealth.jmir.org/2021/2/e24429/

(JMIR Public Health Surveill 2021;7(2):e28131) doi:10.2196/28131

In "General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis" (JMIR Public Health Surveill 2021;7(2):e24429) the authors noted errors in the affiliations of two authors.

For authors Katja Reuter and Jennifer B Unger, the affiliation:

> *Institute for Health Promotion and Disease Prevention Research, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, California, United States*

was inadvertently removed and replaced with the affiliation:

> *Economic and Health Policy Research, American Cancer Society, Washington, DC, United States*

In the originally published paper, the full list of authors and affiliations read as follows:

Katja Reuter[1,2], PhD; Melissa L Wilson[3*], PhD; Meghan Moran[4], PhD; NamQuyen Le[5], MPH; Praveen Angyan[5], MS; Anuja Majmundar[2], MA, MBA, PhD; Elsi M Kaiser[6], MS, PhD; Jennifer B Unger[2*], PhD*

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, United States

[2]Economic and Health Policy Research, American Cancer Society, Washington, DC, United States

[3]Division of Disease Prevention, Policy and Global Health, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

XSL•FO
RenderX

[4]Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[6]Department of Linguistics, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

The corrected list of authors and affiliations is as follows:

Katja Reuter[1,2*], PhD; Melissa L Wilson[3*], PhD; Meghan Moran[4], PhD; NamQuyen Le[5], MPH; Praveen Angyan[5], MS; Anuja Majmundar[6], MA, MBA, PhD; Elsi M Kaiser[7], MS, PhD; Jennifer B Unger[2*], PhD

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY, United States

[2]Institute for Health Promotion and Disease Prevention Research, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, California, United States

[3]Division of Disease Prevention, Policy and Global Health, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[4]Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Southern California Clinical and Translational Science Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States

[6]Economic and Health Policy Research, American Cancer Society, Washington, DC, United States

[7]Department of Linguistics, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

The correction will appear in the online version of the paper on the JMIR Publications website on February 24, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 22.02.21; this is a non–peer-reviewed article;accepted 23.02.21; published 24.02.21.

Please cite as:
Reuter K, Wilson ML, Moran M, Le N, Angyan P, Majmundar A, Kaiser EM, Unger JB
Correction: General Audience Engagement With Antismoking Public Health Messages Across Multiple Social Media Sites: Comparative Analysis
JMIR Public Health Surveill 2021;7(2):e28131
URL: https://publichealth.jmir.org/2021/2/e28131
doi:10.2196/28131
PMID:33626018

©Katja Reuter, Melissa L Wilson, Meghan Moran, NamQuyen Le, Praveen Angyan, Anuja Majmundar, Elsi M Kaiser, Jennifer B Unger. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 24.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Conducting Virtual, Synchronous Focus Groups Among Black Sexual Minority Men: Qualitative Study

Derek T Dangerfield II[1,2*], PhD; Charleen Wylie[1*]; Janeane N Anderson[3*], MPH, PhD

[1]Johns Hopkins School of Nursing, Baltimore, MD, United States

[2]Us Helping Us, People Into Living, Inc, Washington DC, MD, United States

[3]Department of Health Promotion and Disease Prevention, College of Nursing, University of Tennessee Health Science Center, Memphis, TN, United States

[*]all authors contributed equally

Corresponding Author:
Derek T Dangerfield II, PhD
Johns Hopkins School of Nursing
855 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 4109553757
Email: ddanger2@jhu.edu

Abstract

Background: Focus groups are useful to support HIV prevention research among US subpopulations, such as Black gay, Black bisexual, and other Black sexual minority men (BSMM). Virtual synchronous focus groups provide an electronic means to obtain qualitative data and are convenient to implement; however, the protocols and acceptability for conducting virtual synchronous focus groups in HIV prevention research among BSMM are lacking.

Objective: This paper describes the protocols and acceptability of conducting virtual synchronous focus groups in HIV prevention research among BSMM

Methods: Data for this study came from 8 virtual synchronous focus groups examined in 2 studies of HIV-negative BSMM in US cities, stratified by age (N=39): 2 groups of BSMM ages 18-24 years, 5 groups of BSMM ages 25-34 years, and 1 group of BSMM 35 years and older. Virtual synchronous focus groups were conducted via Zoom, and participants were asked to complete an electronic satisfaction survey distributed to their email via Qualtrics.

Results: The age of participants ranged from 18 to 44 years (mean 28.3, SD 6.0). All participants “strongly agreed” or “agreed” that they were satisfied participating in an online focus group. Only 17% (5/30) preferred providing written informed consent versus oral consent. Regarding privacy, most (30/30,100%) reported “strongly agree” or “agree” that their information was safe to share with other participants in the group. Additionally, 97% (29/30) reported being satisfied with the incentive.

Conclusions: Conducting virtual synchronous focus groups in HIV prevention research among BSMM is feasible. However, thorough oral informed consent with multiple opportunities for questions, culturally relevant facilitation procedures, and appropriate incentives are needed for optimal focus group participation.

(JMIR Public Health Surveill 2021;7(2):e22980) doi:10.2196/22980

KEYWORDS

engagement; recruitment; sexual health; telehealth

Introduction

Disparities in HIV and negative sexual health outcomes have impacted Black gay, Black bisexual, and other Black sexual minority men (BSMM) throughout the United States for more than a decade [1,2]. From 2014 to 2018, the incidence of HIV infections have remained unchanged among BSMM generally and have increased by 12% for BSMM ages 25-34 years [3]. In 2018, BSMM accounted for 37% of new diagnoses among gay and bisexual men [3]. If infection rates remain stable and treatment and prevention activities remain ineffective, estimates suggest a 50% lifetime risk of infection among BSMM [4]. More culturally relevant, high-impact activities to reduce HIV and improve health outcomes for this group are urgently needed

XSL•FO
RenderX

[5]. The impact of prevention strategies such as pre-exposure prophylaxis (PrEP) for HIV-negative men and antiretroviral therapy for men living with HIV could be optimized if BSMM are actively engaged in designing intervention approaches [6-8]. Focus groups can be a particularly useful method for identifying community needs and designing culturally relevant intervention strategies [8-10].

Focus groups use within-group perspectives, discussions, and interactions to gather information about collective views on a topic [8,10]. Individual in-depth interviews gather detailed, personal information regarding topics that are sensitive or about which little is known [11], while focus groups leverage group dynamics to identify community views and design culturally relevant interventions [8,11]. Studies using focus groups have uncovered nuanced sociocultural experiences, such as how the combined impact of racism, homonegativity, and stigma from society, family, and health care providers create barriers to HIV prevention behaviors among BSMM [8,10,12]. Focus group data have also revealed how psychosocial factors, such as medical mistrust, inhibit participation in prevention activities, HIV treatment, and research among BSMM [8,12,13].

Although using in-person focus groups to design interventions for BSMM is popular [10], social distancing and stay-at-home orders due to COVID-19 have prohibited in-person research protocols, which could affect progress toward obtaining the high-quality data needed to design culturally relevant strategies for BSMM. Remote research protocols are needed to continue HIV prevention and health promotion activities. Virtual focus groups (ie, focus groups that are not in person) provide an electronic means to obtaining qualitative data from several participants simultaneously [14,15].

Virtual focus groups provide an online platform by which participants can respond to a series of open-ended questions similar to an in-person focus group [16]. Virtual focus groups refer to online chatrooms, discussion boards, email correspondence [14,15,17], and more recent computer-mediated, video communication modalities such as Skype (Microsoft) and Zoom (Zoom Video Communications) [16,18,19]. Conducting virtual focus groups could help researchers overcome various challenges related to study location, research costs, and obtaining the views from a range of participants within a population [20,21]. Virtual focus groups can also reduce inhibitions among some participants and allow more free-flowing discussions [20]. Some have found that the anonymous setting of virtual focus groups allows some participants to discuss sensitive topics more candidly than they would do in person [16,21]. Recent studies have briefly described the protocols and outcomes for conducting virtual focus groups for some subpopulations, such as youths [22], transgender men and women [16,23], and gay and bisexual men [21]. White and colleagues [8] have also briefly described in-person focus group research conducted with BSMM in HIV prevention research. However, little is known regarding the protocols for conducting virtual synchronous focus groups or the acceptability of conducting this research among BSMM.

Describing the protocols and exploring the feasibility and acceptability of conducting virtual synchronous focus group research among BSMM is crucial to obtaining quality data and conducting ethical research [8,24]. It is well established that medical and research mistrust persists among BSMM [8,12,13] which could limit optimal research using virtual synchronous focus groups for this population. The intersectional experience of being both a sexual and racial minority in the United States also inhibits optimal research participation among BSMM [8,25]. Obtaining optimal research participation from BSMM requires careful consideration, as few researchers have had substantial success in reaching or engaging BSMM in HIV research [6,8,26,27]. Although some have described the general use of Zoom videoconferencing for qualitative data collection [16,18,19,28], the literature describing ethical research protocols using virtual synchronous focus groups among BSMM in HIV prevention research is lacking.

To address this deficit, this paper describes the protocols and acceptability of conducting virtual synchronous focus groups in HIV prevention research among BSMM. To date, the methodological and ethical issues related to recruitment and screening, informed consent, maintaining privacy, focus group facilitation, and providing incentives for conducting virtual synchronous focus groups with BSMM have not been adequately addressed. The procedures and protocols in conducting virtual synchronous focus group studies require clarification because research designs, teams, and protocols impact study participation, group dynamics, and the quality of responses among BSMM in HIV prevention research [8,10]. Details from this study will improve researchers' ability to ethically gather in-depth, culturally relevant, and high-quality data to design interventions to reduce HIV for BSMM.

Methods

Virtual Focus Group Sample

Data for this study came from 8 virtual synchronous focus groups examined in 2 studies of HIV-negative BSMM in US cities conducted from March 2020 to June 2020 (N=39). One study explored BSMM's preferences for a multicomponent, peer-based intervention to increase HIV-risk perceptions and PrEP initiation; the other study explored ethical and culturally responsive modalities to improve PrEP research participation among BSMM. Eligibility for both studies included being at least 18 years of age (the age limit for one study was 35 years), identifying as Black or African American and as a man, reporting having sex with a male partner in the previous 12 months, self-reporting being HIV-negative, and residing in a US city. Both studies were guided by the life course theoretical framework. Groups were stratified by age, resulting in the following age groups: 18-24 years (2 groups), 25-34 years (5 groups), and 35 years and older (1 group). Each focus group lasted between 60 and 90 minutes and was conducted online via Zoom. At the end of the focus group, participants were asked to complete an electronic satisfaction survey distributed to their email through a private Qualtrics link. The survey items were designed with consideration of the domains of focus group research conduct and asked questions related to participant satisfaction with the online modality, comfort, concerns regarding privacy, and study incentives. All study procedures

XSL•FO
RenderX

were approved by the Johns Hopkins School of Medicine Institutional Review Board. Virtual synchronous focus group procedures were conducted in the same manner for both studies and are described in the following section.

Recruitment and Screening

Participants were recruited from a combination of active and passive recruitment strategies. Passive recruitment included sharing institutional review board (IRB)–approved study fliers and study information on social media websites (eg, Facebook, Twitter) and Craigslist. Active recruitment included reaching out to participants from existing studies, contacting local community-based organizations, and using incentivized referrals from participants within the study. Interested volunteers could contact the study via phone or text messages and were provided additional study-related information by a trained research assistant using an IRB-approved screening script. The same research assistants screened participants for both studies. After questions were answered, research assistants screened volunteers for eligibility and provided eligible volunteers the details related to the time and log-in details for the virtual synchronous focus group.

Informed Consent

After eligibility was determined, research assistants emailed eligible volunteers a copy of the IRB-approved oral consent form detailing the nature of the study, explaining the potential risks involved in study participation, and indicating that their participation was voluntary. Volunteers were provided an opportunity to ask questions about the study and protocol to the research assistants and the principal investigator (PI) prior to, during, and after the study. After all scheduled study participants joined the virtual focus group, the PI reviewed study-related activities as described in the oral consent form and gave everyone the opportunity to ask additional questions verbally or within the private chat function in Zoom. The PI then prompted each individual participant to provide verbal informed consent. Once the audio recording began, all participants were again asked to individually confirm that they were providing verbal informed consent detailing the nature and risks of the study, that their participation was voluntary, and that they could discontinue at any time without consequence. The study team documented the verbal informed consent for each participant in writing.

Establishing and Maintaining Privacy

Once we determined their eligibility, participants were emailed a copy of the oral informed consent document and instructions informing them of the virtual format, that they should be in a private location, and to use headphones to protect their and others' information. They were also notified that they might be asked to show the group their location via the reverse camera function on their mobile phone or internet-connected device (eg, computer, tablet) to confirm that they were alone. The research team (ie, the research assistants and the PI) informed participants that if their location was compromised by another individual entering their space they would be removed from the virtual focus group.

Participants were provided a password-protected Zoom link and instructed not to share the link with anyone. Once all participants joined the virtual focus group, the facilitator asked each person to verbally confirm that he was in a private location to share and receive information. Participants (including facilitators) were asked to confirm their privacy by showing everyone their space on camera. This was not done if participants were alone in their cars, had headphones on, or were noted walking around their space with their phone or tablet prior to the focus group meeting with no one visible in the background. All participants were also asked if they believed the study environment (eg, facilitators, other participants) to be safe enough to share their views.

Once the facilitator and participants confirmed the group's safety and privacy, the facilitator locked the meeting using Zoom's "lock meeting" function and informed the participants that the meeting was locked. Participants were then asked to change their Zoom screen name to a pseudonym (eg, name of their favorite color) to limit the risk of a participant's given name being audio recorded or exposed in a computer screen shot.

Virtual Focus Group Facilitation

Semistructured focus group guides were designed in consultation with key informants, local community-based organization leadership and staff members who had strong ties to the target population, and investigators with expertise in HIV prevention, qualitative research, and health communications. Participants who accessed a virtual synchronous focus group using an internet-connected mobile device (eg, cellular phone, tablet) were instructed to charge their mobile device while in the focus group to maintain connection to the meeting. Given the relative newness of virtual synchronous focus groups as a data collection modality among this population, focus groups in both studies were limited to 5 participants to reduce the risk of potential privacy breaches and to increase the ease of facilitation. Focus groups were recorded using a digital audio recorder placed near the facilitators' computer to ease participants' concerns of being video recorded.

Two experienced facilitators conducted the virtual groups. One facilitator, a self-identified BSMM with experience conducting qualitative research among the population, led the discussion, managed the group, and recorded field notes. The other facilitator scheduled the groups, recorded field notes, observed group dynamics, and provided technical support for participants who had difficulties connecting to the meeting (eg, mistaken password, confirming time and attendance). Focus groups were also conducted on weekends to accommodate participant schedules. Prior to the discussion, the facilitator initiated casual conversations with the participants virtually to increase their comfort and build rapport prior to the formal discussion. Participants were provided an opportunity to ask additional questions about the study or procedures and to debrief with the facilitators regarding their attitudes toward the nature of the study and the online modality before and after the meeting, which was documented in field notes. Each group began with the facilitator discussing the purpose and ground rules for discussion (eg, one person speaks at a time, respect each other's comments, maintain privacy). For both studies, participants

XSL•FO
RenderX

were asked targeted questions related to ethical research conduct and how researchers could better engage with BSMM during focus groups.

Incentives

Participants were compensated with a US $80 electronic Amazon gift card for one study and a US $75 electronic Amazon gift card for another study. For both studies, participants who referred other eligible volunteers were compensated an additional US $40 electronic gift card for each eligible volunteer they referred up to 2 referrals. Gift cards were scheduled for dissemination within 14 business days of focus group participation and delivered directly to the participants' email on file.

Analysis of Research Protocols and Participant Satisfaction

Field notes that were documented by the research team were reviewed and organized through a process of abductive analysis, and the notes were closely analyzed with consideration to relevant frameworks for qualitative and HIV prevention research methodology [8,29,30]. Specifically, notes from each group were independently reviewed by the authorship team, and then themes related to the procedures and pragmatic issues of the groups were discussed by the research team. Age-related differences were also considered during analysis. Themes were identified through reflexive debriefing whereby the research team outlined and agreed upon pertinent, salient domains for virtual focus group conduct [8,31]. Themes were also considered relative to participant responses to the satisfaction survey.

Results

Table 1 reports the responses from the satisfaction survey regarding virtual synchronous focus group participation among BSMM. Of the 39 who participated in the groups, 30 (77%) completed the survey. The age of participants ranged from 18 to 44 years (mean 28.3, SD 6.0). Regarding virtual focus group participation, 86% (26/30) reported that they "strongly agreed" that they were satisfied participating in a focus group online, while the remaining 14% (4/30) reported "agree." Most (23/30, 77%) reported not preferring in-person focus group participation.

XSL•FO
RenderX

Table 1. Satisfaction of virtual synchronous focus group participation among Black sexual minority men (N=30).

Participants' age and responses	Value
Age (years), range	18-44
Age (years), mean (SD)	28.3 (6.0)
I was satisfied with participating in a focus group online, n (%)	
Strongly agree	26 (87)
Agree	4 (13)
I understood the purpose of the study, n (%)	
Strongly agree	26 (87)
Agree	4 (13)
I would have preferred to provide written informed consent than provide verbal informed consent, n (%)	
Strongly agree	3 (10.0)
Agree	2 (7)
Neither agree nor disagree	11 (37)
Disagree	4 (13)
Strongly disagree	10 (33)
I would have preferred to participate in the focus group in person, n (%)	
Strongly agree	3 (10)
Agree	4 (13)
Neither agree nor disagree	11 (37)
Disagree	6 (20)
Strongly disagree	6 (20.0)
In the future, I would like to participate in other online focus groups, n (%)	
Strongly agree	19 (63)
Agree	10 (33)
Neither agree nor disagree	1 (3)
It is more feasible for me to participate in an online focus group than in an in-person focus group[a], n (%)	
Strongly agree	13 (45)
Agree	9 (31)
Neither agree nor disagree	4 (14)
Disagree	3 (10)
I felt that my information was safe to share with other participants in the group, n (%)	
Strongly agree	20 (67)
Agree	10 (33)
I believe my information will be kept confidential by the research team, n (%)	
Strongly agree	19 (63)
Agree	10 (33)
Neither agree nor disagree	1 (3)
I believe the other participants were in a private space, n (%)	
Strongly agree	18 (60)
Agree	11 (37)
Neither agree nor disagree	1 (3)
I believe my information will be kept confidential by the other people who participated in the focus group, n (%)	
Strongly agree	17 (57)

XSL•FO
RenderX

Participants' age and responses	Value
Agree	10 (33)
Neither agree nor disagree	3 (10)
I was satisfied with the incentive I received for participating in the study[a], n (%)	
Strongly agree	13 (43)
Agree	15 (50)
Disagree	1 (3)

[a]Due to missing data, some responses are less than the total sample.

The following section describes the themes from the research team's debriefing of participant responses.

Motivations and Barriers to Focus Group Participation

Most participants were between ages 25 and 34 years. During screening, research assistants noted that younger participants lacked private locations to participate in a sexual health focus group focused on BSMM. Approximately one-third of participants explicitly mentioned interest in study participation because members of the investigative team were Black and the PI was a BSMM. For instance, one participant, age 30 years, said, "That's why I wanted to do this, because you will understand what we're saying better than 'them' and you need this information." During focus groups, several others reported having expectations that the research team would better understand their perspectives and needs as BSMM, and they were comfortable sharing more personal information with the research team members than they typically would do in research studies with predominately White investigators. In several focus groups, participants expressed feeling as if non-Black researchers "don't really care about us."

Informed Consent

Of those surveyed, 33% (10/30) reported "strongly disagree" to a preference of providing written informed consent versus the oral consent they provided; 13% (4/30) reported "disagree," and 37% (11/30) reported "neither agree nor disagree." Field notes documented how participants across age groups asked questions only related to the nature of the audio recording (whether their faces would be recorded via Zoom) and how long after their participation they would receive their incentive.

Establishing and Maintaining Privacy

All participants reported that they "strongly agree" or "agree" that their information was safe to share with other participants in the group; 63% (19/30) "strongly agreed" that they believed their information would be kept confidential by the research team; 33% (10/30) reported "agree." Most (18/30, 60%) strongly agreed that they believed other participants were in a private space while participating in the group, and 37% (11/30) "agreed." Of note, to maintain privacy, 2 participants between ages 25 and 34 years participated in the groups in their cars. Only 1 person (from the 25-34 year age group) was removed from the virtual focus group due to a combination of technical difficulties and a suspicious location that prompted the other participants to express concerns privately in the chat box.

Virtual Focus Group Facilitation

The team noted that the domains of the focus group guide were maintained despite the virtual nature of the research in both studies. The facilitator had to remind participants across age groups to speak up to ensure that the audio recorder could capture the conversation. Since background noises distracted the audio recording and since excited participants would occasionally speak over each other, participants were also reminded to speak one at a time and to mute themselves if they were not speaking.

The study team had no record or impression that the virtual modality limited participants' sharing their views. Across age groups, participants adequately responded to focus group questions and referred to each other by the pseudonym in the Zoom chat, respecting each other's privacy for the audio recording. Postinterview debriefing with participants revealed that groups with men 34 years old and under indicated increased comfort in participating in the focus group due to the small size (5 participants or less), although participants mentioned they would also prefer a small size in person.

Incentives

Among the participants, 45% (14/30) reported strongly agreeing that they were satisfied with the incentive, and 52% (16/30) reported "agree." However, due to administrative barriers, some groups did not receive their incentive for approximately 30 days after their participation, which caused participants to continue to reach out to the investigative team and make formal complaints. Two participants reached out to the IRB with concerns that the research team was taking advantage of their participation and did not believe the investigative team would compensate them for their time. This caused the research team to identify ways to immediately compensate participants after their virtual research conduct.

Discussion

This paper outlines the protocols for conducting virtual synchronous focus groups with BSMM for HIV prevention research and provides quantitative and qualitative feedback on acceptability from participants. Overall, conducting virtual synchronous focus groups in HIV prevention research among BSMM is feasible. However, careful consideration and attention to providing informed consent, ensuring privacy, facilitating groups, and promptly providing incentives is necessary for optimal focus group participation. Focus group facilitators must be explicitly trained to thoroughly explain study goals and

XSL•FO
RenderX

research protocols, ensure privacy, and manage virtual synchronous groups with BSMM.

Conducting virtual synchronous focus groups could be a useful modality to recruit and engage BSMM who are otherwise "hard to reach," such as professional men who are unable to attend research offices during the workday and low-resourced men who may not have transportation to travel to research facilities [32]. As many studies of BSMM in HIV prevention research oversample low-resourced participants, virtual synchronous focus groups could facilitate data collection among a more representative sample. However, we did not document participants' socioeconomic status in this study and were not able to quantify the relative yield of recruitment for virtual research participation compared to in-person activities; others have found no substantial difference in recruitment yield for virtual versus in-person activities [19]. Moreover, COVID-19 impacted study recruitment, as clinics that were typically helpful were only accepting scheduled patients for in-person visits (not research staff) and staff members that could have referred patients to the studies were limited.

Providing IRB-approved oral informed consent was feasible and may be more acceptable among BSMM than obtaining written informed consent for virtual, synchronous focus group participation. This could be due to the convenience of not having to provide electronic or written signatures. Acceptability could also be due to the detailed nature of our informed consent process that provided several opportunities for study participants to ask questions and gain clarification about study-related expectations. However, it is important that the research team uniformly document oral informed consents. Although most participants reported being satisfied with the informed consent process, more information is still needed about the differences, if any, in BSMM's comprehension of study protocols when informed consent is obtained orally versus in written form. Providing multiple opportunities for participants to ask questions could build trust in the research and is imperative for ethical research conduct with this population.

Smaller focus groups may be optimal for data collection among this population [8]. Data from the present analysis suggest that virtual synchronous groups among BSMM should be limited to 5 individuals to maximize participant comfort and privacy. Although some suggest that online focus groups should be between 8 and 12 participants [11,20], limiting group size provides an opportunity for rapport building between the focus group facilitator and participants as well as among participants. Limiting the size of the group also reduces the likelihood of compromising participants' locations and lessens the risk of privacy breaches. Procedures to maintain participant safety during virtual synchronous focus groups should be described prior to study participation, and the facilitator should be prepared to remove participants who are unable to maintain privacy standards during online discussions.

Having a culturally congruent and culturally competent research team could be an integral part of participant sharing and satisfaction across focus groups [8]. Other studies have found that having culturally congruent research teams reduces research mistrust and fosters optimal participation among BSMM [8,10,33]. Although this study did not measure this directly, having culturally congruent teams could also reduce social desirability bias among BSMM [8,10,34]. Optimizing the utility of cultural congruence includes several factors, including an aesthetic component, personal disclosures, and providing extensive details related to the purpose and importance of the research and participation. Specifically, research teams can "look like" participants and share demographic characteristics, yet maintain professionalism [8,10]. Team members can also share personal information regarding their relationships to the community (eg, being a BSMM or not, living in similar communities or not), and passionately explain why their participation in the research study is important. These activities could build rapport and trust as well as reinforce the importance of maintaining safety standards. Still, research teams should be explicitly trained to optimize virtual synchronous focus group participation among BSMM regardless of demographic characteristics by explaining all study procedures, building trust, and establishing privacy procedures.

Incentivizing research participation is also important for this population. As medical and research mistrust persists for BSMM, immediately incentivizing participants after their study participation is crucial to limiting mistrust and skepticism despite the presence of a culturally congruent team. Due to COVID-19, incentive distribution was more delayed than that in usual in-person activities that permit the immediate disbursement of cash or gift cards. Ultimately, administrative regulations on virtual incentives were updated, and we were able to promptly disburse claim codes electronically upon completion of research participation.

Limitations should be acknowledged. The parent studies included convenience samples of BSMM who were recruited in part through snowball sampling, which limits the application of current methods on a more diverse sample of BSMM. We also conducted this study among HIV-negative BSMM; it is unclear whether these procedures would be equally acceptable among BSMM who are living with HIV. Additionally, this study included relatively small samples, and we were not able to quantify potential social desirability bias in responses.

However, few studies have detailed the important concepts in conducting virtual synchronous focus groups in HIV prevention research and among BSMM. The present considerations to maximize virtual synchronous focus group participation in HIV prevention research among BSMM align with existing recommendations for approaches in community-based research and cultural competency in clinical and research settings [8,24,35-37]. Future research should quantify the relative participation rates of virtual versus in-person focus groups and continue to explore preferences for ethical research study conduct in HIV prevention research among BSMM. Future research should also test the relative impact of larger versus smaller group sizes on group dynamics and participant responses. It is important that researchers take careful consideration of research conduct with this population and remind participants that they are a part of the process of reducing HIV and promoting community public health.

XSL•FO
RenderX

Acknowledgments

This research was supported in part by a National Institute of Mental Health Grant (no. K01MH11849 to PI DTD II) and the Fordham University Research Ethics Training Institute Grant (no. R25DA031608 to PI Fisher). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Hess KL, Hu X, Lansky A, Mermin J, Hall HI. Lifetime risk of a diagnosis of HIV infection in the United States. Ann Epidemiol 2017 Apr;27(4):238-243 [FREE Full text] [doi: 10.1016/j.annepidem.2017.02.003] [Medline: 28325538]
2. Millett GA, Peterson JL, Wolitski RJ, Stall R. Greater risk for HIV infection of black men who have sex with men: a critical literature review. Am J Public Health 2006 Jun;96(6):1007-1019. [doi: 10.2105/AJPH.2005.066720] [Medline: 16670223]
3. Jeffries IW, Dailey A, Jin C, Carter J, Scales L. Trends in diagnosis of HIV Infection, linkage to medical care, and viral suppression among men who have sex with men, by race/ethnicity and age - 33 jurisdictions, United States, 2014-2018. MMWR Morb Mortal Wkly Rep 2020 Sep 25;69(38):1337-1342 [FREE Full text] [doi: 10.15585/mmwr.mm6938a1] [Medline: 32970045]
4. HIV and African American gay and bisexual men. Centers For Disease Control and Prevention. 2020 Feb 04. URL: https://www.cdc.gov/hiv/group/msm/bmsm.html [accessed 2020-02-18]
5. Nelson LE. Awakening to an intersectional reality: ending the HIV epidemic in the USA starts with reducing inequities among Black MSM. J Urban Health 2020 Oct;97(5):589-591. [doi: 10.1007/s11524-020-00487-x] [Medline: 33025307]
6. Magnus M, Franks J, Griffith S, Arnold M, Goodman K, Wheeler D, HPTN 061 Study Group. Engaging, recruiting, and retaining black men who have sex with men in research studies: don't underestimate the importance of staffing--lessons learned from HPTN 061, the BROTHERS study. J Public Health Manag Pract 2014;20(6):E1-E9 [FREE Full text] [doi: 10.1097/PHH.0000000000000025] [Medline: 24406940]
7. Watson C, Wilton L, Lucas J, Bryant L, Victorianne G, Aradhya K, et al. Development of a Black caucus within the HIV prevention trials network (HPTN): representing the perspectives of Black men who have sex with men (MSM). Int J Environ Res Public Health 2020 Feb 03;17(3) [FREE Full text] [doi: 10.3390/ijerph17030871] [Medline: 32028553]
8. White J, Dangerfield II D, Grieb S. Methodological considerations for conducting focus groups in HIV prevention research among Black men who have sex with men. Public Health Nurs 2019 Feb 13. [doi: 10.1111/phn.12592] [Medline: 30761620]
9. Wilson PA, Valera P, Martos AJ, Wittlin NM, Muñoz-Laboy MA, Parker RG. Contributions of qualitative research in informing HIV/AIDS interventions targeting Black MSM in the United States. J Sex Res 2016;53(6):642-654 [FREE Full text] [doi: 10.1080/00224499.2015.1016139] [Medline: 26241373]
10. Dangerfield Ii DT, Harawa NT, McWells C, Hilliard C, Bluthenthal RN. Exploring the preferences of a culturally congruent, peer-based HIV prevention intervention for black men who have sex with men. Sex Health 2018 Nov;15(5):424-430 [FREE Full text] [doi: 10.1071/SH18057] [Medline: 30185352]
11. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. Br Dent J 2008 Mar 22;204(6):291-295. [doi: 10.1038/bdj.2008.192]
12. Cahill S, Taylor SW, Elsesser SA, Mena L, Hickson D, Mayer KH. Stigma, medical mistrust, and perceived racism may affect PrEP awareness and uptake in black compared to white gay and bisexual men in Jackson, Mississippi and Boston, Massachusetts. AIDS Care 2017 Nov;29(11):1351-1358 [FREE Full text] [doi: 10.1080/09540121.2017.1300633] [Medline: 28286983]
13. Dale SK, Bogart LM, Wagner GJ, Galvan FH, Klein DJ. Medical mistrust is related to lower longitudinal medication adherence among African-American males with HIV. J Health Psychol 2016 Jul;21(7):1311-1321 [FREE Full text] [doi: 10.1177/1359105314551950] [Medline: 25293970]
14. Adler CL, Zarchin YR. The "virtual focus group": using the Internet to reach pregnant women on home bed rest. J Obstet Gynecol Neonatal Nurs 2002;31(4):418-427. [doi: 10.1111/j.1552-6909.2002.tb00064.x] [Medline: 12146931]
15. Sweet C. Designing and conducting virtual focus groups. Qualitative Mrkt Res: An Int J 2001 Sep;4(3):130-135. [doi: 10.1108/13522750110393035]
16. Wirtz AL, Cooney EE, Chaudhry A, Reisner SL, American Cohort To Study HIV Acquisition Among Transgender Women. Computer-mediated communication to facilitate synchronous online focus group discussions: feasibility study for qualitative HIV research among transgender women across the United States. J Med Internet Res 2019 Mar 29;21(3):e12569 [FREE Full text] [doi: 10.2196/12569] [Medline: 30924782]
17. Murray PJ. Using virtual focus groups in qualitative research. Qual Health Res 2016 Jul;7(4):542-549. [doi: 10.1177/104973239700700408]

XSL•FO
RenderX

18. Dodds S, Hess A. Adapting research methodology during COVID-19: lessons for transformative service research. JOSM 2020 Jul 17 (forthcoming) [FREE Full text] [doi: 10.1108/josm-05-2020-0153]
19. Rupert DJ, Poehlman JA, Hayes JJ, Ray SE, Moultrie RR. Virtual versus in-person focus groups: comparison of costs, recruitment, and participant logistics. J Med Internet Res 2017 Mar 22;19(3):e80 [FREE Full text] [doi: 10.2196/jmir.6980] [Medline: 28330832]
20. Stewart DW, Shamdasani P. Online focus groups. Journal of Advertising 2016 Nov 16;46(1):48-60. [doi: 10.1080/00913367.2016.1252288]
21. Woodyatt CR, Finneran CA, Stephenson R. In-person versus online focus group discussions: a comparative analysis of data quality. Qual Health Res 2016 May;26(6):741-749. [doi: 10.1177/1049732316631510] [Medline: 26935719]
22. Doornwaard SM, den Boer F, Vanwesenbeeck I, van Nijnatten CHCJ, ter Bogt TFM, van den Eijnden RJJM. Dutch adolescents' motives, perceptions, and reflections toward sex-related internet use: results of a web-based focus-group study. The Journal of Sex Research 2016 Dec 16;54(8):1038-1050. [doi: 10.1080/00224499.2016.1255873] [Medline: 27982701]
23. Miner MH, Bockting WO, Romine RS, Raman S. Conducting internet research with the transgender population: reaching broad samples and collecting valid data. Soc Sci Comput Rev 2012 May 01;30(2):202-211 [FREE Full text] [doi: 10.1177/0894439311404795] [Medline: 24031157]
24. Davidson P, Halcomb E, Gholizadeh L. Focus groups in health research. In: Research Methods in Health: Foundations for Evidence-Based Practice. Melbourne, Vic: Oxford University Press; 2013:54-72.
25. Quinn K, Bowleg L, Dickson-Gomez J. "The fear of being Black plus the fear of being gay": The effects of intersectional stigma on PrEP use among young Black gay, bisexual, and other men who have sex with men. Soc Sci Med 2019 Jul;232:86-93 [FREE Full text] [doi: 10.1016/j.socscimed.2019.04.042] [Medline: 31075752]
26. Hickson DA, Truong NL, Smith-Bankhead N, Sturdevant N, Duncan DT, Schnorr J, et al. Rationale, design and methods of the ecological study of sexual behaviors and HIV/STI among African American men who have sex with men in the Southeastern United States (the MARI study). PLoS One 2015;10(12):e0143823 [FREE Full text] [doi: 10.1371/journal.pone.0143823] [Medline: 26700018]
27. Wheeler DP. Methodological issues in conducting community-based health and social services research among urban Black and African American LGBT populations. Journal of Gay & Lesbian Social Services 2003 Apr;15(1-2):65-78. [doi: 10.1300/j041v15n01_05]
28. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. International Journal of Qualitative Methods 2019 Sep 11;18:160940691987459. [doi: 10.1177/1609406919874596]
29. Timmermans S, Tavory I. Theory construction in qualitative research. Sociological Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
30. Winder TJA. "Shouting it out": religion and the development of Black gay identities. Qual Sociol 2015 Oct 13;38(4):375-394. [doi: 10.1007/s11133-015-9316-1]
31. Berger R. Now I see it, now I don't: researcher's position and reflexivity in qualitative research. Qualitative Research 2013 Jan 03;15(2):219-234. [doi: 10.1177/1468794112468475]
32. Zwaanswijk M, van Dulmen S. Advantages of asynchronous online focus groups and face-to-face focus groups as perceived by child, adolescent and adult participants: a survey study. BMC Res Notes 2014 Oct 24;7:756 [FREE Full text] [doi: 10.1186/1756-0500-7-756] [Medline: 25341440]
33. Dangerfield DT, Cooper J, Heidari O, Allen S, Winder T, Lucas G. Nursing and health care preferences among opioid and stimulant using black sexual minority men: an exploratory study. J Assoc Nurses AIDS Care 2020 Aug 17. [doi: 10.1097/JNC.0000000000000201] [Medline: 32809988]
34. Quinn K, Dickson-Gomez J, Zarwell M, Pearson B, Lewis M. "A gay man and a doctor are just like, a recipe for destruction": how racism and homonegativity in healthcare settings influence prep uptake among young Black MSM. AIDS Behav 2018 Dec;23(7):1951-1963. [doi: 10.1007/s10461-018-2375-z] [Medline: 30565092]
35. Levy ME, Wilton L, Phillips G, Glick SN, Kuo I, Brewer RA, et al. Understanding structural barriers to accessing HIV testing and prevention services among black men who have sex with men (BMSM) in the United States. AIDS Behav 2014 May;18(5):972-996 [FREE Full text] [doi: 10.1007/s10461-014-0719-x] [Medline: 24531769]
36. Lucas JP, King G, Watkins P, Watson CC, Hutchinson CS, Rogers C, et al. The Utilization of Good Participatory Practice (GPP) during the Planning and Implementation of a PrEP Study among Black MSM. AIDS Research and Human Retroviruses 2014 Oct;30(S1):A52-A52. [doi: 10.1089/aid.2014.5091.abstract]
37. Polanco FR, Dominguez DC, Grady C, Stoll P, Ramos C, Mican JM, et al. Conducting HIV research in racial and ethnic minority communities: building a successful interdisciplinary research team. J Assoc Nurses AIDS Care 2011;22(5):388-396 [FREE Full text] [doi: 10.1016/j.jana.2010.10.008] [Medline: 21277228]

Abbreviations

BSMM: Black sexual minority men
IRB: institutional review board

XSL•FO
RenderX

PI: principal investigator
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 28.07.20; peer-reviewed by V Minichiello, J Torres, T Spence; comments to author 11.12.20; revised version received 23.12.20; accepted 09.01.21; published 08.02.21.

Please cite as:
Dangerfield II DT, Wylie C, Anderson JN
Conducting Virtual, Synchronous Focus Groups Among Black Sexual Minority Men: Qualitative Study
JMIR Public Health Surveill 2021;7(2):e22980
URL: http://publichealth.jmir.org/2021/2/e22980/
doi:10.2196/22980
PMID:33427671

©Derek T Dangerfield II, Charleen Wylie, Janeane N Anderson. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 08.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Fostering Engagement With Health and Housing Innovation: Development of Participant Personas in a Social Housing Cohort

Andrew James Williams[1,2], PhD; Tamaryn Menneer[2,3], PhD; Mansi Sidana[4], BSc; Tim Walker[5], PhD; Kath Maguire[2], PhD; Markus Mueller[3,4], Dr rer nat; Cheryl Paterson[2], MSc; Michael Leyshon[5], PhD; Catherine Leyshon[5], PhD; Emma Seymour[2], DipM; Zoë Howard[2], BA; Emma Bland[2], MSc; Karyn Morrissey[2], PhD; Timothy J Taylor[2], PhD

[1]School of Medicine, University of St Andrews, St Andrews, Fife, United Kingdom
[2]European Centre for Environment and Human Health, University of Exeter, Truro, Cornwall, United Kingdom
[3]Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall, United Kingdom
[4]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom
[5]Centre for Geography and Environmental Science, University of Exeter, Penryn, Cornwall, United Kingdom

Corresponding Author:
Andrew James Williams, PhD
School of Medicine
University of St Andrews
Room 219, Medical and Biological Sciences Building
North Haugh
St Andrews, Fife, KY16 9TF
United Kingdom
Phone: 44 7906019236
Email: ajw45@st-andrews.ac.uk

Abstract

Background: Personas, based on customer or population data, are widely used to inform design decisions in the commercial sector. The variety of methods available means that personas can be produced from projects of different types and scale.

Objective: This study aims to experiment with the use of personas that bring together data from a survey, household air measurements and electricity usage sensors, and an interview within a research and innovation project, with the aim of supporting eHealth and eWell-being product, process, and service development through broadening the engagement with and understanding of the data about the local community.

Methods: The project participants were social housing residents (adults only) living in central Cornwall, a rural unitary authority in the United Kingdom. A total of 329 households were recruited between September 2017 and November 2018, with 235 (71.4%) providing complete baseline survey data on demographics, socioeconomic position, household composition, home environment, technology ownership, pet ownership, smoking, social cohesion, volunteering, caring, mental well-being, physical and mental health–related quality of life, and activity. K-prototype cluster analysis was used to identify 8 clusters among the baseline survey responses. The sensor and interview data were subsequently analyzed by cluster and the insights from all 3 data sources were brought together to produce the personas, known as the Smartline Archetypes.

Results: The Smartline Archetypes proved to be an engaging way of presenting data, accessible to a broader group of stakeholders than those who accessed the raw anonymized data, thereby providing a vehicle for greater research engagement, innovation, and impact.

Conclusions: Through the adoption of a tool widely used in practice, research projects could generate greater policy and practical impact, while also becoming more transparent and open to the public.

(JMIR Public Health Surveill 2021;7(2):e25037) doi:10.2196/25037

KEYWORDS

user-centered design; community; social network analysis; United Kingdom; mobile phone

XSL•FO
RenderX

Introduction

History of Personas

In 1999, software developer Alan Cooper [1] published the book "The Inmates are Running the Asylum" in which he advocated "user-centered design." To focus the design of software or any other product on the intended user, Cooper suggested using "personas" [1-5]. Since then, personas have been applied in a wide variety of fields where systems, services, or products are being designed for human use. Such applications include health service design [3,6,7], eHealth services [2,6,8-13], and health behavior change [14-17].

There is a long history of typology in the social sciences, whether seeking to identify types of individuals, organizations, or societies. Both Plato and Aristotle considered there to be forms that were not specific to any person or entity but representing some fundamental collective characteristics, which are seen as the origins of the concept of archetypes [18,19]. Typologies and categorical groups have been useful in the development of the understanding of various aspects of society, such as politics, history, and development [20,21]. Jung developed the earlier ideas around archetypes in the field of psychology as innate and universal primordial ideas (prototypes), which were useful for interpreting behaviors and actions [22]. Ernest Dichter later applied Jung's archetypes to advertising and marketing [23].

Possibly, the most well-known use of a persona in health care services in the United Kingdom was Torbay and South Devon Health and Care National Health Service Trust's "Mrs Smith," a persona of an older woman created to support the integration of health and social care services [24-26]. Most recently, methods for persona development have begun to be applied in research projects to support the understanding of the complex system of social determinants of health [27-29]. In addition, the expansion in the forms and amounts of data collected make it necessary for data producers and data holders to present data in formats that are more accessible while maintaining participant confidentiality.

Persona Development Methods

Two significant challenges in developing personas are avoiding harmful stereotypes and achieving a balance between making the personas relatable so that they are engaging but avoiding being so specific that they do not relate to a large enough group of people (customer base) [5,30]. The process for developing personas typically comprises a number of steps, starting with identifying basic details of the personas, such as demographics, and subsequently adding layers of detail until a sufficiently life-like and relatable character is formed [3,16,30]. The types of details required for the persona are selected to fit the purpose for which the personas are being designed. For example, the designer of a new magazine would want details about the interests and lifestyle choices of the persona, whereas the designer of a health service would want to know the persona's health state and treatment preferences. The approaches taken to this process vary from those based purely on anecdotes or experiences and are therefore more prone to stereotyping [2,7,14,31], through to those based on the objective analysis of data reducing the risk of stereotyping within the personas [3,10,12,13,17,20,27,32,33]. Most people advocate a mixed methods approach using both objective and subjective data to avoid stereotypes and overly specific personas [3,9,10,16,17]. Recently, those who will use the eventual product, process, or service have become involved in the creation of the personas to be used in the development process [2,6,8,34].

The common quantitative methods often applied to persona development are factor analysis [33], latent class analysis [27], k-means cluster analysis [12,13,17], and hierarchical cluster analysis [3,10,32]. The different quantitative methods relate to whether the fundamental characteristics of the clusters are observable or latent hidden attributes. Hagenaars and Halman [20] and Floyd et al [4] have critiqued and compared the various methods based on their statistical properties; however, it is likely that the most appropriate methods for creating personas will depend on the specific scenario and how they will be used.

Regardless of specific methods, there is agreement about the value of the personas in terms of provoking empathy, interest, and understanding; providing grounding and personalization; and supporting better product, process, and service development [2,5,27,28,31,33]. Pruitt and Grudin [5] wrote that personas "provide a conduit for conveying a broad range of qualitative and quantitative data, and focus attention on aspects of design and use that other methods do not."

The aim of this study was to experiment with the use of personas within a research and innovation project to support product, process, and service development through broadening the engagement with and understanding of the data about the local community. In this paper, we outline the innovative mixed methods process we have applied to generating personas of social housing residents and the uses to which these have been put to date. Although the initial process was an established technique for data-driven persona development [12,13,17,33], the qualitative methods and addition of sensor data are more novel. Holden et al [3] advocated for the combination of quantitative and qualitative data "to produce richer, contextualized descriptions of personas" (p. 165), but admitted that they were only minimally able to incorporate qualitative data into their personas. Consequently, the significant incorporation of qualitative data into the final Smartline Archetypes is a novel contribution of this study. In addition, it was hoped that the personas would support participant engagement with their own data and increase the transparency of the project. Subsequently, the derivation of the personas from individual data, but representing groups of people through a fictional life-like persona, can maintain privacy while increasing accessibility to the data.

The Smartline Project

The Smartline project is a European Regional Development Fund–funded research and innovation project focused on household and community health and well-being. Its purpose is to develop the eHealth and eWell-being sector in Cornwall and the Isles of Scilly in the United Kingdom [35] through collaboration between academia and business, specifically the University of Exeter, Coastline Housing (a social housing provider), Volunteer Cornwall, and Cornwall Council [36].

XSL•FO
RenderX

Cornwall is a county in the southwest of England; it only borders one other county, with the other border being the coastline. Considered a rural county, the settlements include many dispersed small villages and towns with populations up to approximately 25,000. Previous studies describing other aspects of the project have examined the associations between health and mold [37] and social cohesion [38] among the participants.

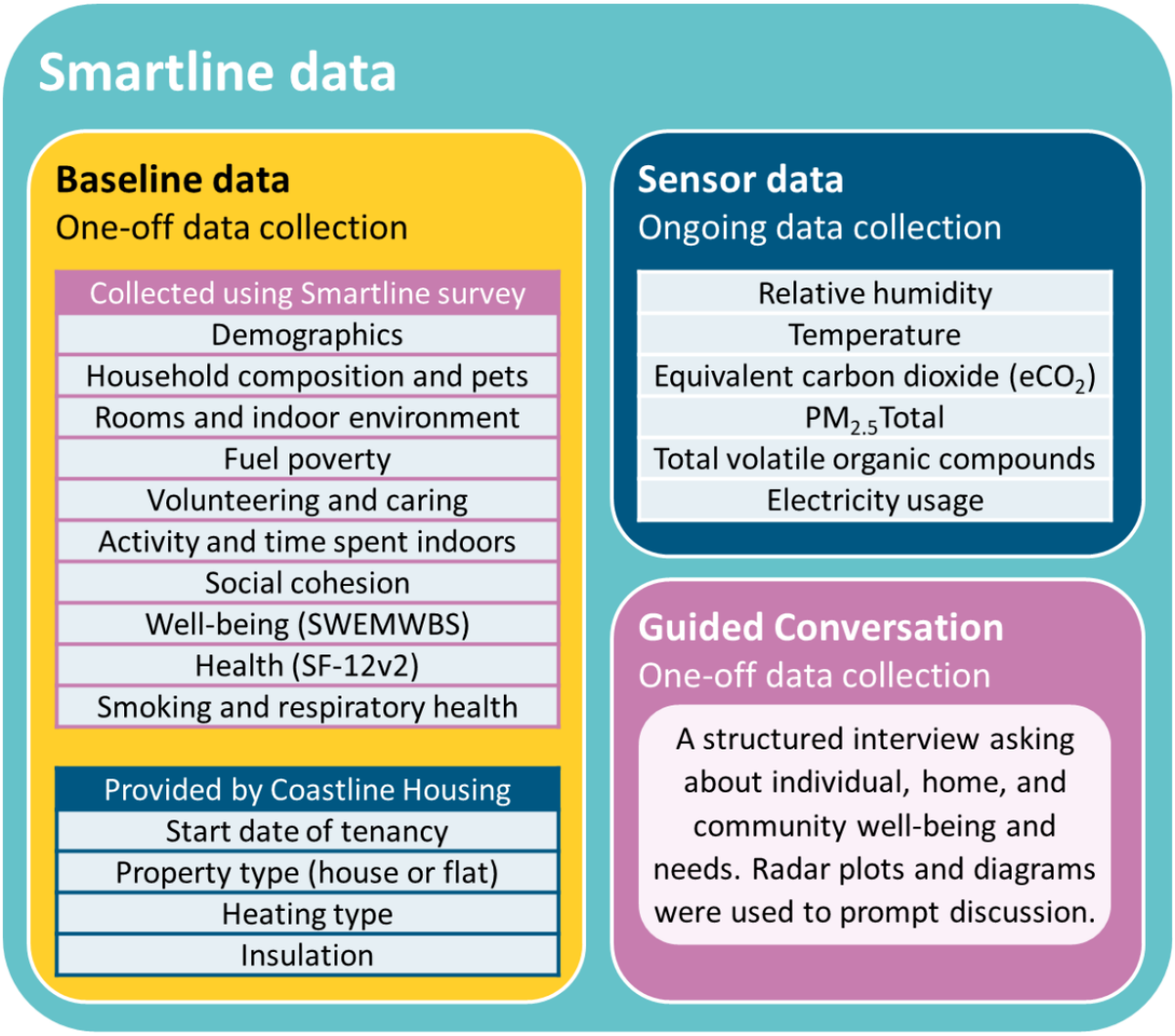
In the United Kingdom, 99.61% (2,738,980/2,718,435) of businesses are small- and medium-sized enterprises (SMEs), with fewer than 250 employees, and in Cornwall, many are microenterprises, with fewer than 10 employees [36,39,40]. Although Smartline had the opportunity to share consented and anonymized data with project partners and local enterprises, such organizations and businesses are unlikely to have the capacity or data science skills required to interact with large quantitative data sets. Therefore, it was necessary for Smartline to present data in formats that are more accessible. A Smartline Knowledge Exchange Officer (ES) had used personas in market research settings and recognized their potential to address data accessibility for SMEs in the region.

The participants were adults (older than 18 years) recruited from among Coastline Housing residents in the towns of Camborne and Redruth and the villages of Illogan and Pool. Together, these form the largest conurbation in Cornwall, with a population of 47,500 in the last census in 2011 [41,42]. Moreover, these locations were selected because they provided a high concentration of Coastline Homes needed to address the project's focus on communities and individual households. Coastline Housing undertook participant recruitment street-by-street between September 2017 and November 2018. In total, 649 households were approached; 329 were recruited into the project and completed baseline data collection (329/649, 50.7% response rate).

The project collected a variety of data (Figure 1), using a face-to-face survey, environmental and electricity usage sensors, and a structured interview called a Guided Conversation [43,44]. The project was reviewed by the University of Exeter Research Ethics Committee, and all participants provided written informed consent. All participants needed to consent to participate in the survey and to have sensors installed to join the project, but participation in the Guided Conversation was not a requirement. The surveys and Guided Conversations took place at a convenient time in the participants' homes with 2 researchers present. Sensor data measurements were recorded approximately every 3-5 min. The data were collected to stimulate innovation within the project in partnership with businesses and voluntary organizations working with the project. The personas were developed using the various data collected throughout the project to stimulate further innovation.

XSL•FO
RenderX

Figure 1. Smartline project data sets. eCO_2: equivalent carbon dioxide; $PM_{2.5}$: atmospheric particulate matter that have a diameter of less than 2.5 m; SWEMWBS: Short Warwick Edinburgh Mental Wellbeing scale.



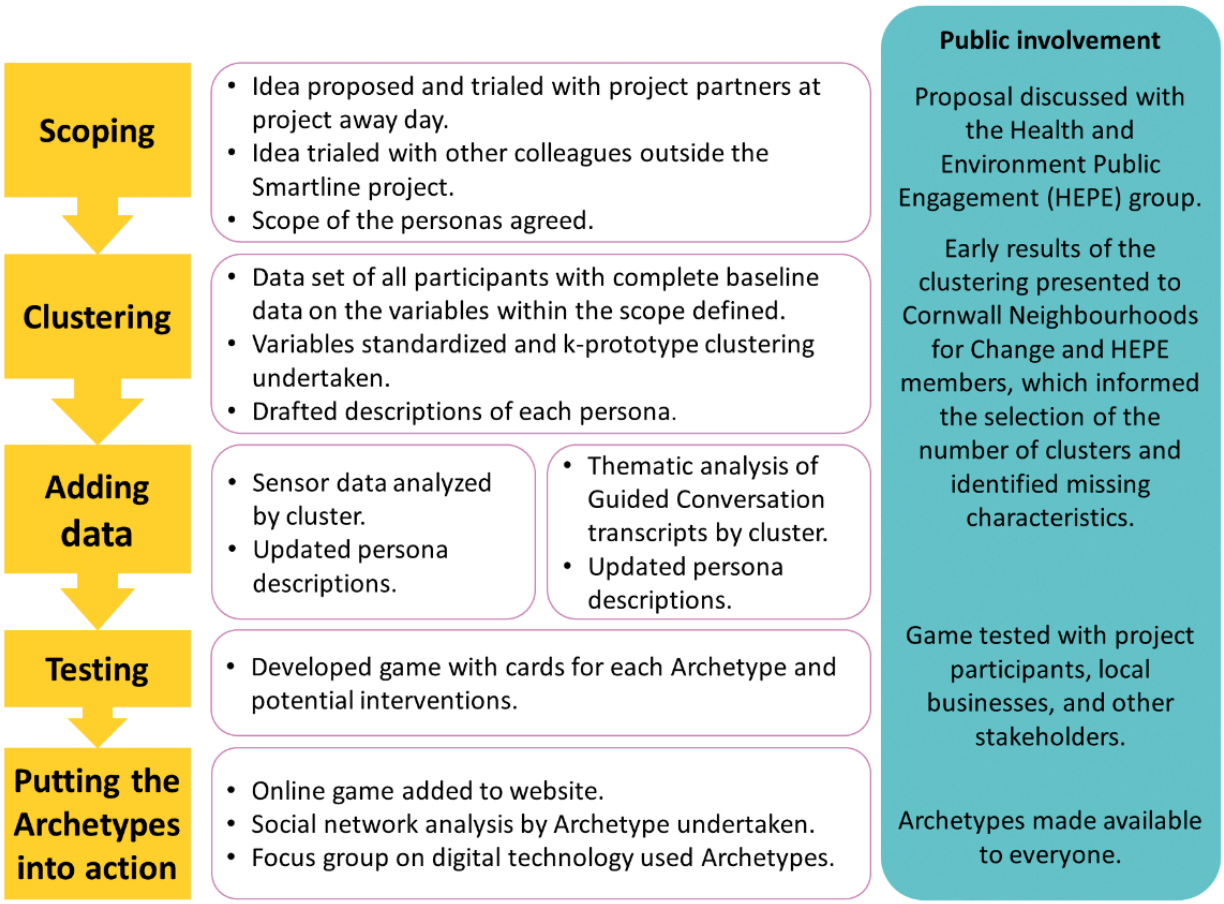
Methods

Persona Development Process

The process used to develop the personas, illustrated in Figure 2, followed the common steps of initially specifying some basic details about each persona using the baseline data and then layering on further details (the sensor and Guided Conversation data) until life-like individuals were created [3,16,30]. Smartline participants and broader public groups were involved throughout the process to ensure that the final personas were acceptable, accessible, and true to people's experience [16]. Initially, the idea was presented to a number of groups to test whether it was considered worth pursuing and to define the scope of the personas. The next step was to undertake a cluster analysis of the survey data.

XSL•FO
RenderX

Figure 2. Persona development method for the Smartline Archetypes. HEPE: Health and Environment Public Engagement group.



k-Prototype Cluster Analysis

Cluster analysis (also known as segmentation or taxonomy analysis) is a statistical method for interpreting a large data set by grouping the records into clusters. Each record has values for a set of variables (eg, gender, age, socioeconomic status). Cluster centers are randomly generated, each being a set of variable values. Records are assigned to the most similar cluster center, and cluster membership is iteratively updated to minimize the difference between records and their cluster center, based on the variable values. Records in the resulting clusters are generally more similar to one another than to records in other clusters [45]. The k-prototype approach, similar to other clustering techniques, is an unsupervised machine learning method. Developed from the k-means and k-modes methods, the k-prototype method can handle both continuous and categorical data [45,46]. The method minimizes the Euclidean distance for numerical factors, as in k-means clustering, and uses the number of mismatches between data points for the categorical variables [45].

The baseline data used for the cluster analysis included variables related to demographics, socioeconomic position, household composition, experience of home environment (comfort, mold, and fuel poverty), technology ownership, pet ownership, smoking, social cohesion, volunteering, caring, mental well-being (Short Warwick-Edinburgh Mental Wellbeing Score [47-50]), physical and mental health–related quality of life (12-item Short Form Health Survey, version 2 [51,52]), and activity (Figure 1). All variables included in the cluster analysis were quantitative data from the survey responses or the Coastline Housing data. The inclusion of these variables would allow the personas to reflect multiple aspects of the participants' lives and demonstrate the breadth of data held by the project. List-wise deletion based on the 329 selected variables left 235 (71.4%) participants on whom to conduct the cluster analysis. To account for the various scales of each variable, the data were standardized using *z*-scores before the clustering analysis. The k-prototype clustering was performed in R using the package "clustMixType" [53,54], with participants assigned to the cluster that most closely matched their characteristics.

It is necessary to specify the number of clusters to be calculated when conducting k-prototype analysis. In the literature, there are both analytical and pragmatic techniques for identifying the appropriate number of clusters [13,17]. With the intention that the personas would be accessible to the public, it was important for us to triangulate these data-driven decisions with community-focused perspectives. To those ends, both the techniques to be used and the potential granularity of the clusters were discussed with 2 groups of community partners: Health and Environment Public Engagement group and Cornwall Neighbourhoods for Change. Similarity within each cluster increases with the number of clusters. The optimum number of clusters is often chosen to be the number at which little is gained by adding more clusters. This method is known as the elbow

XSL•FO
RenderX

method, using a plot as presented in Multimedia Appendix 1. However, we found no clear analytical evidence for selecting a given number of clusters over another number. In addition, feedback on granularity from community partners and on business requirements from Smartline's Knowledge Exchange Officer (ES) suggested that a maximum of 8 clusters would be appropriate. Pragmatically, the 8 clusters were also sufficiently populated to capture a generalization across multiple people (average of 29 participants per cluster). Patterns within the summary statistics of the variables within each cluster were examined to characterize each Smartline Archetype. All the data from the survey were included in the characterization of the Smartline Archetypes, not just those variables included in the cluster analysis.

Sensor Data

Using the unique participant identifiers, the data from environmental and electricity usage sensors for each home were allocated into each of the 8 clusters. The mean sensor data readings were calculated for each household over all the readings taken in 2019. In line with the choice to use the term archetype, we anticipated that the variation within the sensor outcomes of each Smartline Archetype would be of interest, for example, to compare a high- and low-electricity user of the same Smartline Archetype. Hence, the Smartline Archetypes were integrated into the project data-sharing platform.

Qualitative Analysis

A subsample of 62 semistructured qualitative interviews (known as "Guided Conversations") were conducted. Participants were selected via nonprobability sampling out of a total sample of 329 participants. The interviews lasted for an average of 45 min and were conducted face to face by 2 researchers between November 2017 and May 2018. The responses were recorded directly onto a script by the assigned note taker and transcribed to a database after the interview. Participants were not paid, and interviews took place during a time that was most convenient for the participant. The purpose of these interviews was to identify well-being priorities and then develop an achievable action plan with the participant. They were structured around the 3 themes of well-being, home, and community under which there were a series of prompts. These themes and prompts were arrived at through a co-design process involving all project partners. The interview guide was piloted with 4 voluntary and community sector organizations and 5 Coastline Housing tenants and adjustments made.

Using unique study identifiers, the transcripts of each interview were allocated to the clusters. There was an uneven spread of interview data across Smartline Archetypes: #1, 6 interviews; #2, 3 interviews; #3, 3 interviews; # 4, 4 interviews; # 5, 6 interviews; #6, 4 interviews; #7, 10 interviews; and #8, 5 interviews. Due to the list-wise deletion of survey records, not everyone who was interviewed was allocated to a Smartline Archetype.

An interdisciplinary team of 10 researchers conducted a 5-step collaborative data analysis exercise with the interview transcripts. Using multiple coders increases the rigor in a qualitative analysis by drawing upon diverse perspectives and counteracting individual biases in the coding process as interpretations and assumptions are placed in the plain view of the group [55,56]. This method also allowed us to reasonably manage the large data set [55]. In this study, we chose to conduct the coding manually for the following two reasons. First, many of the interdisciplinary teams were unfamiliar with qualitative analysis software, and therefore, time-intensive training would be required [57]. Second, the marking up, sorting, and reorganizing of transcripts was deemed a manageable task given the 10-strong team of researchers.

Thematic analysis of interview transcripts involved a systematic 5-step process (Textbox 1). Through this exercise, the team produced a codebook that was transparently documented and justified the analytical decisions [55,58]. The outputs were additions or adaptations to each Smartline Archetype description and a 3-point list of headline descriptors. The team was split into 5 pairs, each of whom coded 2 Smartline Archetypes.

XSL•FO
RenderX

Textbox 1. Five-step thematic analysis process.

Step 1: Data familiarization and identification of significant topics

- Each pair familiarized themselves with their Smartline Archetypes transcripts and examined the graphs, which illustrated the satisfaction scores from the radar plots for each of the interview topics. The graphs facilitated quick identification of the highest and lowest scoring topics within and between each Smartline Archetype. The output of this iterative process was the identification of a significant topic or topics for each of the Smartline Archetypes.

Step 2: Open coding and subtheme development

- Open coding is the process of identifying discrete concepts and patterns in the data [59]. The team employed this process on the significant topics, identified in step 1, for each Smartline Archetype.

Step 3: Axial coding and theme identification and triangulation

- Axial coding is the dynamic and creative process of identifying connections between patterns in the data [57,59]. The team used this process in reference to the Smartline Archetype characteristics produced by the cluster analysis and the open codes. This iterative process enabled points of triangulation to be identified between the quantitatively derived characteristics and the interview data. The output from this step was the identification of a 3-point list of headline descriptors for each Smartline Archetype.

Step 4: Pull exemplar quotes from transcripts

- Quotes that exemplified the significant theme were then pulled from transcripts and added to the code book.

Step 5: Write summary sentence

- The final step was to write sentences that summarized the themes identified and insert them into the Smartline Archetype description.

Testing

To test if the Smartline Archetypes were acceptable, accessible, and true to people's experience [16], we produced a "serious game," that is, one used for more than just entertainment [60]. Each Smartline Archetype was allocated a name and a cartoon image, presented as "Top Trumps" cards. The game, which involved matching attributes to characters, was played at community events involving project participants and events attended by businesses. The feedback from participants supported the use of the Smartline Archetypes, and most people found at least one Archetype that they could relate to themselves or a neighbor. The cards also prompted conversations with participants around the support, services, or products that might be useful to that Smartline Archetype. Providing people with a character that is similar but distinct from themselves has previously been used to prompt reflection and potential behavior change by Wyatt et al [61] and Brown et al [62].

Putting the Archetypes Into Action

To date, the Archetypes have been used in three different ways. First, an updated version of the persona card game was turned into a "game" that can be played on the Smartline website [63]. Second, the personas have also been used to facilitate focus groups with participants to gather views about behaviors and attitudes toward digital technology [64]. Third, the Smartline Archetypes have also been explored as part of a social network analysis of the participants (Stevens et al, unpublished data, 2021).

Results

The Smartline Archetypes

The 8 clusters identified by the k-prototype analysis of the baseline survey data are summarized in Table 1. Two-thirds of Smartline participants were female, and their ethnic diversity reflects that of Cornwall, with only 3.9% (10/256) from an ethnic minority. The Smartline Archetypes reflected these demographics. However, public engagement with community partners identified that it was important to include some diversity among the Smartline Archetypes. Therefore, the 4 Archetypes with the lowest likelihood of being female were designated male, and the Archetype with the highest proportion of ethnic minority participants was presented as being from an ethnic minority (Archetype #6).

XSL•FO
RenderX

Table 1. Summary characteristics of the 8 clusters identified by the k-prototype analysis of the baseline survey data.

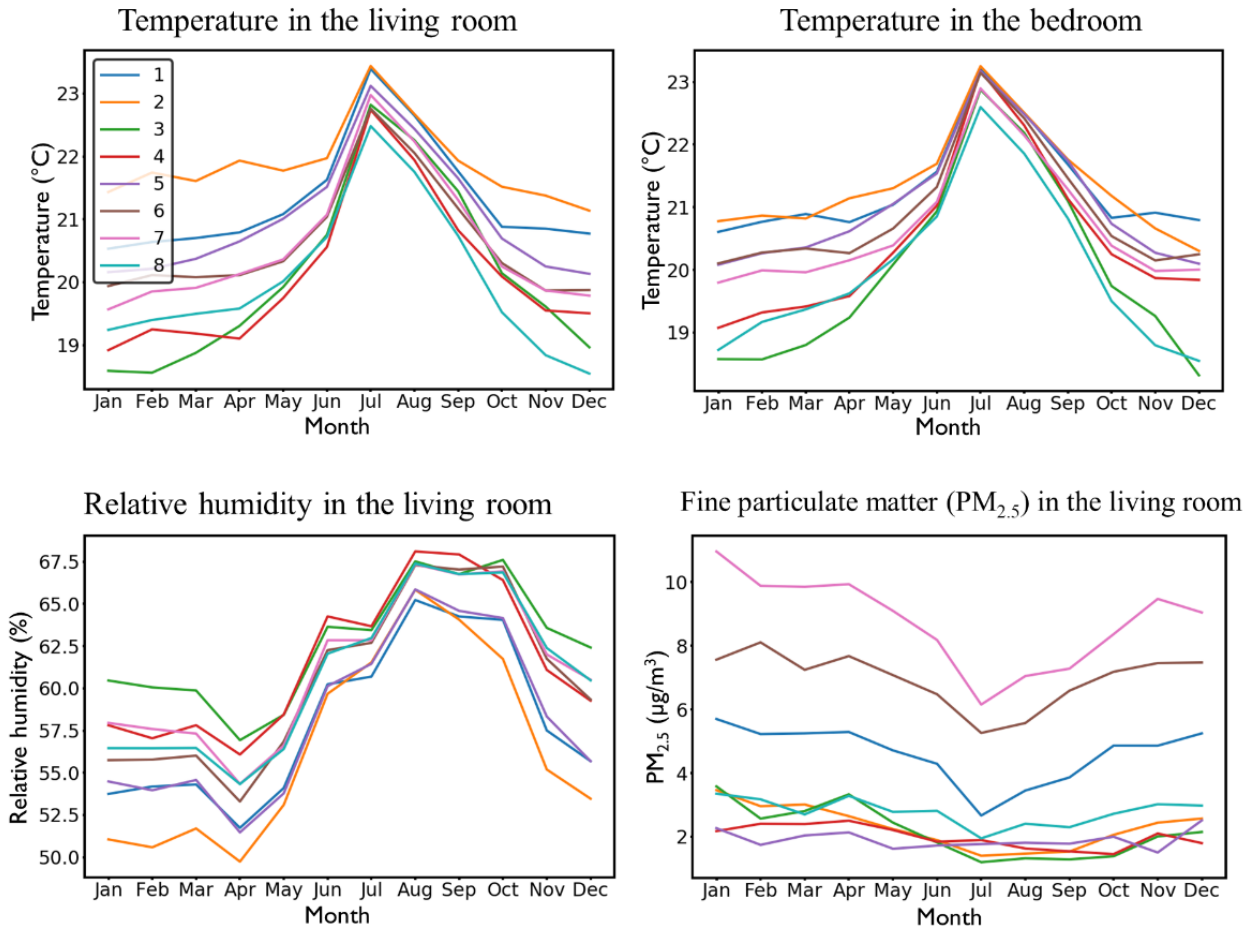
Characteristic	Cluster							
	1	2	3	4	5	6	7	8
Participants, n (%)	24 (10.2)	28 (11.9)	18 (7.7)	31 (13.2)	28 (11.9)	23 (9.8)	54 (23.0)	29 (12.3)
Female, n (%)	16 (67)	17 (601)	14 (78)	23 (74)	17 (61)	15 (65)	38 (70)	21 (72)
Age (years), median	61.0	67.5	52.0	34.0	65.5	66.0	55.0	63.0
National identity, mode	British	British	Cornish	Cornish	Cornish	British	Cornish	Cornish
Ethic minority, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)	2 (11)	1 (2)	0 (0)
Employed, n (%)	10 (42)	3 (11)	3 (17)	11 (35)	5 (18)	2 (9)	4 (7)	8 (28)
Retired, n (%)	6 (25)	18 (64)	4 (22)	0 (0)	18 (64)	13 (57)	14 (26)	14 (48)
IMD[a] 10% most deprived, n (%)	16 (67)	11 (39)	9 (50)	17 (55)	19 (68)	12 (52)	26 (48)	14 (48)
Urban, n (%)	22 (92)	27 (96)	15 (83)	28 (90)	28 (100)	21 (91)	49 (91)	24 (83)
Household size, range	1-3	1-2	1-3	2-4	1-2	1-3	1-2	1-3
Internet access	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Smart meters	No	No	No	No	No	No	No	Yes
Pets (mode)	Cats	Dogs	Cats	Cats	Cats	Dogs	Dogs	Other
Fuel poverty	No	No	Yes	No	No	No	Yes	No
Mold	No	No	Yes	Yes	No	No	Yes	Yes
Smoker	No	Yes	No	Yes	No	Yes	No	No
Volunteering	No	No	Yes	No	Yes	Yes	Yes	No
Physical health	Poor	Poor	Good	Good	Poor	Good	Poor	Good
Mental health	Average	Average	Poor	Average	Good	Good	Poor	Good
Physically active	No	No	No	Yes	Yes	Yes	No	Yes

[a]IMD: index of multiple deprivation.

Sensor data types are presented in Figure 1. Mean values were taken over 2019 for each data type and each household. Means were compared across Smartline Archetypes using a separate one-way analysis of variance (ANOVA) for each sensor data type, with the Archetype as the between-participants factor with 8 levels. Significant effects of the Archetype were investigated using the Tukey post hoc test for pairwise comparisons of Smartline Archetypes. There was no significant effect of the Archetype on relative humidity in the bedroom ($F_{7,160}$=1.382; P=.22; η^2=0.057), $PM_{2.5}$ (atmospheric particulate matter that have a diameter of less than 2.5 µm: $F_{7,115}$=1.263; P=.28; η^2=0.071), equivalent carbon dioxide ($F_{7,84}$=1.246; P=.29; η^2=0.094), and electricity usage ($F_{7,75}$=0.885; P=.52; η^2=0.076). There was a trend toward significance for temperature in the bedroom ($F_{7,160}$=1.932; P=.07; η^2=0.078) and for relative humidity in the living room ($F_{7,153}$=2.024; P=.06; η^2=0.085). Temperature in the living room differed across Archetypes ($F_{7,153}$=2.380; P=.02; η^2=0.098), with higher temperature in Smartline Archetype #2 “David Hartley” than Smartline Archetype #8 “Cathy Johnson” (P=.03). The sensor data did not reveal many additional insights about the Smartline Archetypes. However, 4 examples of the sensor data are shown in Figure 3, including those measures with the greatest differences across Archetypes. To illustrate the potential of the sensor data, monthly means were plotted. The temperature sensor data appeared to be consistent with participants who reported issues with temperature in the survey, whereas those Archetypes with higher $PM_{2.5}$ did not seem to be consistent with those living near roads, smoking, or keeping their windows closed. It was clear that there was more variation in the internal environment in winter than in summer.

XSL•FO
RenderX

Figure 3. Smartline sensor data monthly means for the archetypes across 1 year. $PM_{2.5}$: atmospheric particulate matter that have a diameter of less than 2.5 µm.



The qualitative analysis of the Guided Conversations was consistent with the findings of the quantitative cluster analysis. Through the qualitative analysis, it was possible to add depth and explain the features from the survey data. Only within Archetype #7 “Sarah Jones” did triangulating the qualitative and quantitative data prove challenging. This was the largest of the Smartline Archetypes, with more than 20% (54/235, 23.0%) of the participants included in the clustering analysis and 10 Guided Conversation transcripts. Although the quantitative approach clustered these individuals, the Guided Conversation data revealed a variety of circumstances within this Archetype. The people in Archetype #7 experienced a number of complicated circumstances around health, finances, and caring responsibilities for family members. This scenario demonstrates that although the reported data can be similar, there can be significant differences in experience that could be missed without the richness of qualitative data or user engagement. The final descriptions of each of the Smartline Archetypes bringing together the baseline survey data, household sensor, and Guided Conversation data are provided in Table 2.

XSL•FO
RenderX

Table 2. The Smartline Archetypes (the percentages are the percentage of the 235 participants in each archetype).

Archetype	Description
Archetype #1—"Jack Brown," Male, 59 years old, 10.2% (n=24)	
	• Resourceful—Competent—Capable • Jack works for a building merchant and lives with his wife and cat in one of the 10% most deprived neighborhoods in England. They have access to the internet and own quite a few pieces of technology. His health and well-being are about average and he is fairly active. He does not think that mold is affecting their health, despite having mold in the bathroom and his wife having respiratory symptoms. They help out their neighbors. He takes pride in being self-reliant but also knows whom to ask for help and is willing to do this. He feels competent and confident in completing household and "Do-it-Yourself" tasks, even those tasks that he does not necessarily enjoy.
Archetype #2—"David Hartley," Male, 65 years old, 11.9% (n=28)	
	• Poor physical health—Insular—Inhibited aspirations • David retired early owing to poor physical health. He lives alone and is a moderate smoker. He spends most of his time indoors and does not do any volunteering. He does not own much technology and does not have access to the internet. He has chronic obstructive pulmonary disease and spends most of his time sitting. His mental health is better than his physical health, which affects many aspects of his life, from exercise and recreation to occupation and learning, and their associated aspirations. He is somewhat insular by choice, although he is happy with where he lives. He does not have any mold in the home.
Archetype #3—"Mandy Green," Female, 55 years old, 7.7% (n=18)	
	• Frequent walker—Well connected—Strong sense of belonging • Mandy lives with her daughter who is at college and some cats in a 3-bedroom house in a rural area. She is a self-employed cleaner. She owns a tablet computer and a smart watch. To save money on heating, Mandy only heats certain rooms and avoids opening windows. Mandy spends a lot of time indoors at the weekends but has never smoked. Her physical health is good, but her mental health is poor. Mandy has a strong sense of belonging to her community. She takes walks frequently, including to get to her cleaning jobs. However, reports of crime in the area undermine her sense of safety.
Archetype #4—"Jennie Fryer," Female, 37 years old, 13.2% (n=31)	
	• Driven—Resourceful—Informed • Jennie lives with her partner and infant child in a 3-bedroom home. While bringing up their child, Jennie is undertaking a National Vocational Qualification course at college. She and her partner smoke, but they both try not to smoke indoors. She owns a smartphone and keeps fish. They have mold in the bathroom and bedrooms, which she is concerned is affecting the family's health; however, they only heat specific rooms. Jennie wants to work, but the jobs available locally do not fit with her family commitments, skills, and training. She is resourceful and well connected, knowing how to obtain information if she has an issue. She is worried about the lack of parking in the neighborhood.
Archetype #5—"Fred Jones," Male, 65 years old, 11.9% (n=28)	
	• Self-reliant—Happy and active—Reflective • Fred is a retired taxi driver living with his wife and a dog in a 1-bedroom house in one of the 10% most deprived areas in England. He owns a smart watch and a laptop. He feels that their home is adequately heated but uses a dehumidifier as they have some mold. Fred and his wife are fairly self-reliant, happy, and active, but reflective as they worry for the future of community and family with regard to community spirit and jobs. He volunteers formally as well as helps out neighbors, as he has a strong sense of social cohesion.
Archetype #6—"Raj Singh," Male, 60 years old, 9.8% (n=23)	
	• Active community member—Disappointed by local services—Resilient but in pain • Raj lives with his adult son and is looking for work in a factory. Although they heat all the rooms in their home, he does not think the home is adequately heated. However, they do not experience much mold. He and his son have quite a few technological devices. Raj smokes and gets short of breath, but otherwise his mental and physical health and well-being are fairly good. He helps out in his community, as he is interested in the standard of services available. Raj is more focused on the community and the services available than his indoor environment.
Archetype #7—"Sarah Jones," Female, 50 years old, 23.0% (n=54)	
	• Financial difficulties—Isolated—Caring responsibilities and disability • Sarah used to work, but she cannot work now because of a long-term condition. She has some caring responsibilities for an older parent. Sarah has concerns about smart technology and privacy, so she only has a simple mobile phone, which makes it difficult for her to access some services that require internet access. Her home smells moldy, and she is worried about how this is affecting her health. But Sarah does not go outside much and avoids opening window to save heat. She does some volunteering at a charity shop but reports low social cohesion. Sarah is frustrated by her declining mental and physical health, and her limited finances.

XSL•FO
RenderX

Archetype	Description
Archetype #8—"Cathy Johnson," Female, 61 years old, 12.3% (n=29)	
	• Content on own or with their small group of friends—Frequent walker—Enjoys mentally stimulating activities • Cathy is retired and lives in a village with her adult daughter. Cathy has upgraded to smart gas and electricity meters. Her daughter keeps a snake. Cathy dries her clothes indoors and is not concerned about the cost of heating. She does not do any volunteering but has caring responsibilities toward her daughter. Cathy's hobbies are crosswords, reading, and knitting. She is fairly insular with a close-knit group of friends. Her physical and mental health are good and have been improving. She is fairly busy at home so does not have much interest in the community.

Putting the Archetypes Into Action

Having tested the Smartline Archetypes using the "Top Trumps" style cards with a number of audiences, including the project participants themselves, the Archetypes have been put into action. The web-based version of the card game was only launched in summer 2020; therefore, it is still being evaluated. Within the digital technology focus group, participants were asked to decide whether a given Archetype would like to engage with technology and, if so, what kind. Participants' comments suggested that they felt more able to talk about themselves than the Archetypes, but the Archetypes did provide a conversation facilitator and allowed participants to avoid a more personal perspective as desired. The social networks (ego networks) of the Smartline participants who participated in this study are illustrated in Multimedia Appendix 2, with the Smartline Archetype of the participants denoted by color. An ANOVA analysis of the social network ties by Smartline Archetype identified a statistically significant difference in the number of ties reported by Archetype, with Archetype #5, "Fred Jones," reporting an average of 12 ties, while the others reported around 4-7 ties. Such information could be useful in community development or spreading health messages, through the identification of those who might spread messages well, or those who are disconnected and might need targeted messages. Various and ongoing engagements of broader stakeholders with the Smartline Archetypes have continued to confirm their validity, and the Archetypes have proved to be an engaging tool for discussions about the project and data. A number of small and microenterprises who had not engaged with the project data before engaged with the Smartline Archetypes, learning about the project participants and the prompting ideas related to their business. The number of Smartline Archetypes and the wealth of information known about each one means that specific Archetypes or specific details can be selected depending on the topic, product, process, or service being discussed.

Discussion

Principal Findings and Implications

The process of developing personas to inform product, process, and service development has been widely adopted across multiple sectors including health care [1-3,6-17,24-26]. Qualitative and quantitative research methods are being used in the process of developing personas, but their recognition as a research tool is more recent [2,3,7,9,10,12-14,16,17,20,27,31-33]. Within the Smartline Research and Innovation Project, a mixed methods process was developed to create personas from survey, household sensor, and interview data. The Smartline Archetypes were created to facilitate innovation by making the project data more accessible, particularly to small and medium-sized enterprises working in sectors related to eHealth and eWell-being.

The process used to develop the Smartline Archetypes employed existing research methods, some of which, such as k-prototype cluster analysis, had previously been applied in persona development, whereas the qualitative approach and incorporation of environmental sensor data were novel. Holden et al [3] reported that they were able to "demonstrate the value of using largely qualitative data from a multiyear study but also identify the challenges of prolonged analysis and the difficulty of incorporating a rich and heterogeneous set of findings into a single design." Therefore, the high level of triangulation we found between our data sources and the relatively rapid analytical methods applied to the qualitative data are significant developments. Overall, the approach was truly multidisciplinary, with contributions from epidemiology, health service research, mathematics, geography, and community engagement coming together into a product that reflects more than the sum of its parts [65]. Subsequently, it has been possible to apply the Smartline Archetypes in multiple ways with the project participants themselves and other stakeholders.

Data are crucial to research but can also be highly controversial, particularly with the new types and volumes of data that are becoming available. Calls for open science to increase transparency and accessibility of research meet the challenges of maintaining the duty of confidentiality regarding the data the public trusts to share with us [66]. Developing and maintaining trust in how participant data will be used is quite rightly recognized as fundamental to health research using data. Being transparent about how personal patient data are going to be used links to calls for great statistical literacy [67], which is supported by engaging communities in designing dissemination tools. The Smartline Archetypes provided an engaging opportunity to anonymously present the data collected by the study back to the participants and other stakeholders, overcoming some of the barriers to engaging with the data, such as statistical literacy.

Limitations and Areas for Development

Despite these valuable uses identified for the Smartline Archetypes, we also identified a number of weaknesses or challenges in their development. The need to specify the number of clusters to be created by the k-prototype method might limit the use of such clustering methods. Although it is possible to base the number of clusters on the data, it might also be necessary to be pragmatic and specify a certain number of clusters, which would affect the validity of the methods applied.

XSL•FO
RenderX

Clustering analyses make it possible to consider any number clusters and categorize the performance of the clustering accordingly (eg, via "elbow method"). This can accommodate or challenge prespecified requirements (eg, required minimum or maximum number of classes or clusters).

The development of personas could be based on stereotypical views of individuals or groups [5,30]. Basing our personas on the mathematical analysis of the survey data, the basic characteristics were derived using some objective criteria. Even in these circumstances, adding further embellishments to the personas could be influenced by unconscious biases or stereotypes. This influence could have occurred during the thematic analysis; however, by involving community partners and a team of researchers in this process, we hope that this risk has been minimized. This approach to persona development could also challenge stereotypes. For example, in this study, could some of the Smartline Archetypes be people who do not reflect stereotypes of social housing residents?

Starting with the survey data meant that the clusters identified emphasized the biases in the data set in terms of gender, age, ethnicity, etc. Community partners underlined that the data on which the Smartline Archetypes were based did not reflect the whole community, just those approached and willing to participate in the study. Subsequently, some diversity was added to the Smartline Archetypes, which might limit their validity. More extensive testing and validation of the Smartline Archetypes with the research participants and other stakeholders would be valuable but needs to be balanced against the risk of individual biases shaping the Archetypes. Archetype #7 "Sarah Jones" revealed a particular challenge to the use of quantitative methods alone to derive personas. Although the k-prototype methods grouped the people in this Smartline Archetype as being similar, the qualitative methods revealed significant differences in their circumstances. As the Smartline Archetype with the largest number of interview transcripts, the variation might simply reflect the larger volume of qualitative data or might reflect that objective data cannot adequately capture human experience and similar quantitative data might hide important differences between people. It is worth noting that all variables were equally weighted in the clustering process, but another approach could be to use different weightings to dictate the importance of certain characteristics over others. More theoretically, there is a need to consider whether the personas reflect collective fundamental but observable characteristics (archetypes) within which variation might be of interest or latent, hidden, or primordial ideas as in prototypes [20,27]. This distinction in the type of persona will depend upon the uses to which the personas will be put but might be an important distinction when comparing personas between studies or populations.

Conclusions

Personas are a widely adopted tool that could prove useful in research, especially in using research to inform policy, practice, and business engagement. Methods are available to bring together various types of data into personas, and the resulting personas are recognized for being useful in communicating complex data [5]. The most appropriate methods to produce personas will depend on the specific application and data available, meaning that this approach is adaptable to a range of projects and disciplines. Unlike previous research, Smartline personas were created by layering quantitative survey, household sensor, and qualitative interview data, providing a novel multifaceted perspective. Personas were used within the Smartline project to maintain participant privacy while also increasing data accessibility. Therefore, the participants themselves were better able to engage with their own data and the project, and stakeholders from multiple sectors could use the project to inform innovation. Subsequently, personas represent an opportunity for broader engagement with research and greater policy and practice impact.

Acknowledgments

The authors would like to acknowledge the contribution of the whole Smartline team including all the partners for collecting the data and making suggestions for improving the content of this paper. The Smartline project is receiving up to £3,780,374 (US $5,134,657) of funding from the England European Regional Development Fund as part of the European Structural and Investment Funds Growth Programme 2014-2020. The Ministry of Housing, Communities, and Local Government (and in London, the intermediate body Greater London Authority) is the Managing Authority for European Regional Development Fund. Established by the European Union, the European Regional Development Fund helps local areas stimulate their economic development by investing in projects that will support innovation, businesses, create jobs, and local community regenerations [68]. Additional funding of £25,000 (US $33,962) was from the South West Academic Health Science Network.

Authors' Contributions

AW, ML, CL, EB, K Morrissey, and TT devised the Smartline project, with AW overseeing this study and drafting the manuscript. EB and ES had the initial idea for the archetypes, and ZH developed graphics for each archetype. All the authors were part of the data collection team. TM, MS, and MM conducted the k-prototype analysis, with TM and MM also conducting the sensor analysis. TW, ML, and CL oversaw the qualitative data collection and analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Plot of total sum of squared distances between cluster members and their respective cluster center.
[PDF File (Adobe PDF File), 198 KB - publichealth_v7i2e25037_app1.pdf]

Multimedia Appendix 2
Smartline social network analysis showing ego Smartline Archetype.
[PDF File (Adobe PDF File), 231 KB - publichealth_v7i2e25037_app2.pdf]

References

1. Cooper A. The inmates are running the asylum. Indianapolis: Sams Publishing; 1999.
2. Bhattacharyya O, Mossman K, Gustafsson L, Schneider EC. Using Human-Centered Design to Build a Digital Health Advisor for Patients With Complex Needs: Persona and Prototype Development. J Med Internet Res 2019 May 09;21(5):e10318 [FREE Full text] [doi: 10.2196/10318] [Medline: 31094334]
3. Holden RJ, Kulanthaivel A, Purkayastha S, Goggins KM, Kripalani S. Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure. Int J Med Inform 2017 Dec;108:158-167 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.006] [Medline: 29132622]
4. Floyd IR, Cameron Jones M, Twidale MB. Resolving incommensurable debates: a preliminary identification of persona kinds, attributes, and characteristics. Artifact 2008 Apr;2(1):12-26. [doi: 10.1080/17493460802276836]
5. Pruitt J, Grudin J. Personas: practice and theory. In: Proceedings of the 2003 conference on Designing for user experiences. New York: Association for Computing Machinery; 2003 Jun Presented at: DUX03: Designing the User Experience; June, 2003; San Francisco California p. 1-15. [doi: 10.1145/997078.997089]
6. Donald M, Beanlands H, Straus S, Ronksley P, Tam-Tham H, Finlay J, et al. Preferences for a self-management e-health tool for patients with chronic kidney disease: results of a patient-oriented consensus workshop. CMAJ Open 2019;7(4):E713-E720 [FREE Full text] [doi: 10.9778/cmajo.20190081] [Medline: 31822502]
7. Price M, Bellwood P, Hill TT, Fletcher S. Team Mapping: A Novel Method to Help Community Primary Healthcare Practices Transition to Team-Based Care. Healthc Q 2020 Jan;22(4):33-39. [doi: 10.12927/hcq.2020.26086] [Medline: 32073389]
8. Fico G, Martinez-Millana A, Leuteritz J, Fioravanti A, Beltrán-Jaunsarás ME, Traver V, et al. User Centered Design to Improve Information Exchange in Diabetes Care Through eHealth : Results from a Small Scale Exploratory Study. J Med Syst 2019 Nov 18;44(1):2. [doi: 10.1007/s10916-019-1472-5] [Medline: 31741069]
9. Haldane V, Koh JJK, Srivastava A, Teo KWQ, Tan YG, Cheng RX, et al. User Preferences and Persona Design for an mHealth Intervention to Support Adherence to Cardiovascular Disease Medication in Singapore: A Multi-Method Study. JMIR Mhealth Uhealth 2019 May 28;7(5):e10465 [FREE Full text] [doi: 10.2196/10465] [Medline: 31140445]
10. Reeder B, Zaslavksy O, Wilamowska KM, Demiris G, Thompson HJ. Modeling the oldest old: personas to design technology-based solutions for older adults. AMIA Annu Symp Proc 2011;2011:1166-1175 [FREE Full text] [Medline: 22195177]
11. Schäfer K, Rasche P, Bröhl C, Theis S, Barton L, Brandl C, et al. Survey-based personas for a target-group-specific consideration of elderly end users of information and communication systems in the German health-care sector. Int J Med Inform 2019 Dec;132:103924 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.07.003] [Medline: 31569006]
12. Tanenbaum ML, Adams RN, Iturralde E, Hanes SJ, Barley RC, Naranjo D, et al. From Wary Wearers to d-Embracers: Personas of Readiness to Use Diabetes Devices. J Diabetes Sci Technol 2018 Nov;12(6):1101-1107 [FREE Full text] [doi: 10.1177/1932296818793756] [Medline: 30132692]
13. Tanenbaum ML, Adams RN, Lanning MS, Hanes SJ, Agustin BI, Naranjo D, et al. Using Cluster Analysis to Understand Clinician Readiness to Promote Continuous Glucose Monitoring Adoption. J Diabetes Sci Technol 2018 Nov;12(6):1108-1115 [FREE Full text] [doi: 10.1177/1932296818786486] [Medline: 29991281]
14. Clack L, Stühlinger M, Meier M, Wolfensberger A, Sax H. User-centred participatory design of visual cues for isolation precautions. Antimicrob Resist Infect Control 2019;8:179 [FREE Full text] [doi: 10.1186/s13756-019-0629-9] [Medline: 31827775]
15. Silverman BG, Holmes J, Kimmel S, Branas C, Ivins D, Weaver R, et al. Modeling emotion and behavior in animated personas to facilitate human behavior change: the case of the HEART-SENSE game. Health Care Manag Sci 2001 Sep;4(3):213-228. [doi: 10.1023/a:1011448916375] [Medline: 11519847]
16. Vilvens HL, Vaughn LM, Southworth H, Denny SA, Gittelman MA. Personalising Safe Sleep Messaging for Infant Caregivers in the United States. Health Soc Care Community 2020 May;28(3):891-902. [doi: 10.1111/hsc.12920] [Medline: 31847057]
17. Vosbergen S, Mulder-Wiggers JMR, Lacroix JP, Kemps HMC, Kraaijenhagen RA, Jaspers MWM, et al. Using personas to tailor educational messages to the preferences of coronary heart disease patients. J Biomed Inform 2015 Feb;53:100-112 [FREE Full text] [doi: 10.1016/j.jbi.2014.09.004] [Medline: 25239261]

XSL•FO
RenderX

18. Kraut R. Plato. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Stanford, CA: The Metaphysics Research Lab; 2017.
19. Shields C. Aristotle. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Stanford, CA: The Metaphysics Research Lab; 2015.
20. Hagenaars J, Halman L. Searching for ideal types: the potentialities of latent class analysis. European Sociological Review 1989;5(1):81-96. [doi: 10.1093/oxfordjournals.esr.a036509]
21. McKinney JC. Constructive Typology and Social Theory. New York: Appleton; 1966.
22. Jung C, Gerhard A, Hull R. Collected works of C. G. Jung, Volume 9 (Part 1) Archetypes and the collective unconscious. Princeton: Princeton University Press; 2014.
23. Stern BB. The Importance of Being Ernest: Commemorating Dichter's Contribution to Advertising Research. J. Adv. Res 2004;44(2):165-169. [doi: 10.1017/s0021849904040127]
24. Thistlethwaite P. Integrating health and social care in Torbay: improving care for Mrs Smith. London: The King's Fund; 2011. URL: https://www.kingsfund.org.uk/sites/default/files/integrating-health-social-care-torbay-case-study-kings-fund-march-2011.pdf [accessed 2021-01-11]
25. Imison C, Poteliakhoff E, Thompson J. Older people and emergency bed use: exploring variation. London: The King's Fund; 2012. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/older-people-and-emergency-bed-use-aug-2012.pdf [accessed 2021-01-11]
26. NHS Torbay and South Devon. The Torbay Model – "Mrs Smith". Torquay: Torbay and South Devon NHS Foundation Trust; 2021. URL: https://www.torbayandsouthdevon.nhs.uk/about-us/our-vision-of-health-and-care/the-torbay-model-mrs-smith/ [accessed 2021-01-11]
27. Weden MM, Bird CE, Escarce JJ, Lurie N. Neighborhood archetypes for population health research: is there no place like home? Health Place 2011 Jan;17(1):289-299 [FREE Full text] [doi: 10.1016/j.healthplace.2010.11.002] [Medline: 21168356]
28. Jais C, Hignett S, Halsall W, Kelly D, Cook M, Hogervorst E. Chris and Sally's House: Adapting a home for people living with dementia (innovative practice). Dementia (London) 2019 Nov 07:1471301219887040. [doi: 10.1177/1471301219887040] [Medline: 31698949]
29. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Stockholm: Institute for Future Studies Contract No 2007:14.
30. Pruitt J, Adlin T. The persona lifecycle: keeping people in mind throughout product design. Amsterdam: Elsevier; 2006.
31. Canham SL, Mahmood A. The use of personas in gerontological education. Gerontol Geriatr Educ 2019;40(4):468-479. [doi: 10.1080/02701960.2018.1506775] [Medline: 30124379]
32. Zagallo P, McCourt J, Idsardi R, Smith MK, Urban-Lurain M, Andrews TC, et al. Through the Eyes of Faculty: Using Personas as a Tool for Learner-Centered Professional Development. CBE Life Sci Educ 2019 Dec;18(4):ar62 [FREE Full text] [doi: 10.1187/cbe.19-06-0114] [Medline: 31755820]
33. McGinn J, Kotamraju N. Data-driven persona development. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2008 Presented at: CHI Conference on Human Factors in Computing Systems; April, 2008; Florence Italy.
34. Bonnardel N, Pichot N. Enhancing collaborative creativity with virtual dynamic personas. Appl Ergon 2020 Jan;82:102949. [doi: 10.1016/j.apergo.2019.102949] [Medline: 31563797]
35. University of Exeter. Smartline. Truro: University of Exeter; 2020. URL: https://www.smartline.org.uk/ [accessed 2021-01-11]
36. Cornwall & Isles of Scilly Integrated Teritorial Investment Strategy. Cornwall: Cornwall & Isles of Scilly Growth Programme; 2016. URL: https://www.cioslep.com/assets/file/Cornwall%20and%20IOS%20ITI%20Strategy.pdf [accessed 2021-01-11]
37. Moses L, Morrissey K, Sharpe RA, Taylor T. Exposure to Indoor Mouldy Odour Increases the Risk of Asthma in Older Adults Living in Social Housing. Int J Environ Res Public Health 2019 Jul 22;16(14) [FREE Full text] [doi: 10.3390/ijerph16142600] [Medline: 31336583]
38. Williams AJ, Maguire K, Morrissey K, Taylor T, Wyatt K. Social cohesion, mental wellbeing and health-related quality of life among a cohort of social housing residents in Cornwall: a cross sectional study. BMC Public Health 2020 Jun 22;20(1):985 [FREE Full text] [doi: 10.1186/s12889-020-09078-6] [Medline: 32571296]
39. Office for National Statistics. UK business: activity, size and location. London: Office for National Statistics; 2020. URL: https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/datasets/ukbusinessactivitysizeandlocation [accessed 2021-01-11]
40. European Commission. SME definition. Brussels: European Commission URL: https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en [accessed 2021-01-11]
41. Cornwall Council. Camborne, Pool and Redruth Community Network Area profile. Truro: Cornwall Council; 2017. URL: https://www.cornwall.gov.uk/media/28425272/camborne-pool-redruth-cna-profile.pdf [accessed 2021-01-11]
42. Cornwall Council. Camborne, Pool, Illogan & Redruth Framework. Truro: Cornwall Council; 2017 Mar. URL: https://www.cornwall.gov.uk/media/26748822/cpir-tf-2017-pages.pdf [accessed 2021-01-11]
43. Kaczynski D, Wood L, Harding A. Using radar charts with qualitative evaluation: Techniques to assess change in blended learning. Active Learning in Higher Education 2008 Mar;9(1):23-41. [doi: 10.1177/1469787407086743]

XSL•FO
RenderX

44. Shrestha G. Radar Charts: A Tool to Demonstrate Gendered Share of Resources. Gender, Technology and Development 2002;6(2):197-213. [doi: 10.1177/097185240200600202]
45. Huang Z. Clustering large data sets with mixed numeric and categorical values. In: Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference. 1997 Presented at: The First Pacific-Asia Conference on Knowledge Discovery and Data Mining URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.9984
46. Madhuri R, Murty M, Murthy JVR, Reddy PVGD, Satapathy S. Cluster analysis on different data sets using K-modes and K-prototype algorithms. ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol II; 2014 Presented at: 48th Annual Convention of Computer Society of India; 13th –15th December 2013; Vishakapatnam. [doi: 10.1007/978-3-319-03095-1_15]
47. Ng Fat L, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Qual Life Res 2017 May 16;26(5):1129-1144 [FREE Full text] [doi: 10.1007/s11136-016-1454-8] [Medline: 27853963]
48. Stewart-Brown S. The Warwick-Edinburgh Mental Wellbeing Scales - WEMWBS. Coventry: Warwick Medical School; 2020. URL: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs [accessed 2021-01-11]
49. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009;7:15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
50. Stewart-Brown S, Platt S, Tennant A, Maheswaran H, Parkinson J, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. Journal of Epidemiology & Community Health 2011 Sep 13;65(Suppl 2):A38-A39. [doi: 10.1136/jech.2011.143586.86]
51. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992 Jul 18;305(6846):160-164 [FREE Full text] [doi: 10.1136/bmj.305.6846.160] [Medline: 1285753]
52. Maruish M. User's manual for the SF-12v2 health survey. 3rd ed. Lincoln, RI: QualityMetric Incorporated; 2012.
53. Huang Z. Extensions to the K-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery 1998;2(3):283-304. [doi: 10.1023/A:1009769707641]
54. Szepannek G. clustMixType: User-Friendly Clustering of Mixed-Type Data in R. The R Journal 2018;10(2):200-208. [doi: 10.32614/rj-2018-048]
55. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook Development for Team-Based Qualitative Analysis. CAM Journal 1998;10(2):31-36. [doi: 10.1177/1525822X980100020301]
56. Richards KAR, Hemphill MA. A Practical Guide to Collaborative Qualitative Data Analysis. Journal of Teaching in Physical Education 2018 Apr;37(2):225-231. [doi: 10.1123/jtpe.2017-0084]
57. Basit T. Manual or electronic? The role of coding in qualitative data analysis. Educational Research 2003;45(2):143-154. [doi: 10.1080/0013188032000133548]
58. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 4th edition. California: SAGE Publications Ltd; 2015.
59. Saldaña J. The coding manual for qualitative researchers. 3rd edition. Los Angeles: SAGE; 2016.
60. Susi T, Johannesson M, Backlund P. Serious games - an overview (Technical Report HS- IKI -TR-07-001). Sweden: School of Humanities and Informatics, University of Skövde; 2007. URL: https://www.researchgate.net/publication/220017759_Serious_Games_-_An_Overview [accessed 2021-01-11]
61. Wyatt KM, Lloyd JJ, Abraham C, Creanor S, Dean S, Densham E, et al. The Healthy Lifestyles Programme (HeLP), a novel school-based intervention to prevent obesity in school children: study protocol for a randomised controlled trial. Trials 2013 Apr 04;14:95 [FREE Full text] [doi: 10.1186/1745-6215-14-95] [Medline: 23556434]
62. Brown K, Eernstman N, Huke AR, Reding N. The drama of resilience: learning, doing, and sharing for sustainability. Ecology and Society 2017;22(2):8. [doi: 10.5751/ES-09145-220208]
63. Williams AJ. Guess who?. Truro: Smartline Project; 2020 May 26. URL: https://smartline-exeter.squarespace.com/main-content-area/guess-who [accessed 2021-01-11]
64. Buckingham S, Walker T, Morrissey K. P94 The feasibility and acceptability of digital technology for health and wellbeing in social housing communities in Cornwall: a qualitative scoping study. Journal of Epidemiology and Community Health 2020;74(Suppl 1). [doi: 10.1136/jech-2020-SSMabstracts.186]
65. Phoenix C, Osborne NJ, Redshaw C, Moran R, Stahl-Timmins W, Depledge MH, et al. Paradigmatic approaches to studying environment and human health: (Forgotten) implications for interdisciplinary research. Environmental Science & Policy 2013 Jan;25:218-228. [doi: 10.1016/j.envsci.2012.10.015]
66. Ford E, Boyd A, Bowles JKF, Havard A, Aldridge RW, Curcin V, et al. Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. Learn Health Syst 2019 Jul;3(3):e10191 [FREE Full text] [doi: 10.1002/lrh2.10191] [Medline: 31317072]
67. Ashby D. Pigeonholes and mustard seeds: growing capacity to use data for society. J. R. Stat. Soc. A 2019 Aug 25;182(4):1121-1137. [doi: 10.1111/rssa.12483]

XSL•FO
RenderX

68. European Structural and Investment Fund. GOV.UK. URL: https://www.gov.uk/guidance/england-2014-to-2020-european-structural-and-investment-funds [accessed 2021-02-02]

Abbreviations

ANOVA: analysis of variance
$PM_{2.5}$: atmospheric particulate matter that have a diameter of less than 2.5 µm
SME: small and medium-sized enterprise

Edited by Y Khader; submitted 20.10.20; peer-reviewed by K Schäfer, E Vanegas, A Fioravanti; comments to author 13.11.20; revised version received 14.12.20; accepted 21.12.20; published 16.02.21.

Please cite as:
Williams AJ, Menneer T, Sidana M, Walker T, Maguire K, Mueller M, Paterson C, Leyshon M, Leyshon C, Seymour E, Howard Z, Bland E, Morrissey K, Taylor TJ
Fostering Engagement With Health and Housing Innovation: Development of Participant Personas in a Social Housing Cohort
JMIR Public Health Surveill 2021;7(2):e25037
URL: http://publichealth.jmir.org/2021/2/e25037/
doi:10.2196/25037
PMID:33591284

©Andrew James Williams, Tamaryn Menneer, Mansi Sidana, Tim Walker, Kath Maguire, Markus Mueller, Cheryl Paterson, Michael Leyshon, Catherine Leyshon, Emma Seymour, Zoë Howard, Emma Bland, Karyn Morrissey, Timothy J Taylor. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 16.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Global Infectious Diseases Epidemic Information Monitoring System: Development and Usability Study of an Effective Tool for Travel Health Management in China

Dayong Gu[1*], MD; Jianan He[2,3*], PhD; Jie Sun[2,3], BSc; Xin Shi[4], PhD; Ying Ye[2,3], BSc; Zishuai Zhang[5], BSc; Xiangjun Wang[6], PG; Qun Su[7], MSc; Wenjin Yu[7], PhD; Xiaopeng Yuan[8], PhD; Ruiling Dong[2,3], MIM

[1]Department of Laboratory Medicine, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen, China

[2]Shenzhen International Travel Health Care Center (Shenzhen Customs District Port Outpatient Clinics), Shenzhen Customs District, Shenzhen, China

[3]Shenzhen Academy of Inspection and Quarantine, Shenzhen Customs District, Shenzhen, China

[4]Business School, All Saints Campus, Manchester Metropolitan University, Manchester, United Kingdom

[5]Department of Statistics & Applied Probability, National University of Singapore, Singapore, Singapore

[6]Harbin Institute of Technology, Shenzhen, Shenzhen, China

[7]Shenzhen Datathinking Corporation, Shenzhen, China

[8]Department of Laboratory Medicine, ZhuJiang Hospital, Southern Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ruiling Dong, MIM
Shenzhen International Travel Health Care Center (Shenzhen Customs District Port Outpatient Clinics)
Shenzhen Customs District
Room 406, Shenzhen Kouan Hospital, Huanggang Kouan Shenghuo Qu
Futian District
Shenzhen
China
Phone: 86 15019457579
Email: 1047194259@qq.com

Abstract

Background: Obtaining comprehensive epidemic information for specific global infectious diseases is crucial to travel health. However, different infectious disease information websites may have different purposes, which may lead to misunderstanding by travelers and travel health staff when making accurate epidemic control and management decisions.

Objective: The objective of this study was to develop a Global Infectious Diseases Epidemic Information Monitoring System (GIDEIMS) in order to provide comprehensive and timely global epidemic information.

Methods: Distributed web crawler and cloud agent acceleration technologies were used to automatically collect epidemic information about more than 200 infectious diseases from 26 established epidemic websites and Baidu News. Natural language processing and in-depth learning technologies have been utilized to intelligently process epidemic information collected in 28 languages. Currently, the GIDEIMS presents world epidemic information using a geographical map, including date, disease name, reported cases in different countries, and the epidemic situation in China. In order to make a practical assessment of the GIDEIMS, we compared infectious disease data collected from the GIDEIMS and other websites on July 16, 2019.

Results: Compared with the Global Incident Map and Outbreak News Today, the GIDEIMS provided more comprehensive information on human infectious diseases. The GIDEIMS is currently used in the Health Quarantine Department of Shenzhen Customs District (Shenzhen, China) and was recommended to the Health Quarantine Administrative Department of the General Administration of Customs (China) and travel health–related departments.

Conclusions: The GIDEIMS is one of the most intelligent tools that contributes to safeguarding the health of travelers, controlling infectious disease epidemics, and effectively managing public health in China.

(JMIR Public Health Surveill 2021;7(2):e24204) doi:10.2196/24204

XSL•FO
RenderX

KEYWORDS

infectious disease; epidemic information; travel health; global; surveillance

Introduction

In the past, when infectious disease outbreaks have occurred in certain countries, such as the Middle East Respiratory Syndrome (MERS) outbreak in South Korea in 2015, specific health examinations had to be performed on travelers from those countries upon entering China [1]. This kind of health management work at Chinese ports of entry are undertaken by travel health officers (referred to as health quarantine officers in China) from China Customs. This is one of the key approaches to preventing and controlling the transmission of infectious diseases [2].

Along with the development of global economic integration, the number of international travelers to China is gradually increasing. For instance, Shenzhen is an international metropolis in China, with an urban population of more than 10 million [3]. In 2019, approximately 242 million passengers passed through the Port of Shenzhen [4]. Travel health officers at China Customs use a temperature monitoring system and epidemiological investigation as the main detecting tools to identify infected travelers [5]. Many travelers may not be aware of their infection status while traveling [6]. Therefore, travel health officers must pay attention to the latest epidemic information regarding certain infectious diseases to determine whether these travelers may be infected and prepare the related public health materials. In general, information concerning global infectious disease outbreaks is manually collected, which is a time-consuming and error-prone process. Travel health officers require an infectious disease information collection system that can automatically collect epidemic information from a large number of websites, extract key information, and translate it into the native language. For these reasons, the Harbin Institute of Technology (Shenzhen, China) and Datathinking Corporation (Shenzhen, China) developed the Global Infectious Diseases Epidemic Information Monitoring System (GIDEIMS) in coordination with the Central Laboratory of Health Quarantine of the Shenzhen International Travel Health Care Center (Shenzhen Customs District Port Outpatient Clinics). The GIDEIMS is currently used by the Health Quarantine Department at the Port of Shenzhen Customs, and the system was recommended to the Health Quarantine Administrative Department of the General Administration of Customs (China) and travel health–related departments. An Epidemic Information Team has been established by the General Administration of Customs (China) aimed at supporting the travel health officers and related departments in public health management. With the application of the GIDEIMS, updated and useful global infectious disease information is sent to the concerned parties daily, as well as useful information for infectious disease prevention and control, such as instructions on how to prepare public health materials, including masks and test kits. The aim of the GIDEIMS is to provide travelers and travel health staff with a helpful tool for public health management.

Methods

Selection of Websites

Web queries on infectious diseases could be one of the most accurate, cost-effective, and labour-extensive sources of syndromic surveillance [7]. Some established and/or official websites are frequently used by related staff and citizens to gain human infectious disease epidemic information. The websites used in the GIDEIMS are listed in Table 1; most infectious disease epidemic information, including information about COVID-19, can be found on these websites. Although all of these websites show excellent performance in the search and distribution of epidemic information, they are characterized by specific limitations. For instance, ProMED [8] reports on human diseases, as well as plant and animal diseases, while the Global Incident Map reports on fewer than 40 types of human infectious diseases, excluding chicken pox. Moreover, the epidemic information provided by the World Health Organization is reported by the member of states, and its distribution may be delayed. Data from different resources must be generated to obtain comprehensive information; the combined expertise of the different systems enhances performance for the early detection of infectious disease outbreaks [9]. Considering that the traditional manual search approach is time-consuming and may be inaccurate, as well as the fact that most of the existing websites are only offered in non–Chinese languages, Chinese travel health officers require a user-friendly infectious disease system that can effectively translate diverse information from multiple sources. For these purposes, we developed the GIDEIMS to provide a simple, effective, and sustainable tool for obtaining information on global human infectious disease epidemics.

XSL•FO
RenderX

Table 1. Websites used in the Global Infectious Diseases Epidemic Information Monitoring System.

Website number	Name and responsible department	Important column
1[a]	• WHO[b] [10]	• "Disease Outbreak News"
2[a]	• WHO Western Pacific Region [11]	• "Outbreaks and emergencies" in "Emergencies"
3[a]	• WHO Regional Office for the Eastern Mediterranean [12]	
4[a]	• WHO Regional Office for Europe [13]	• "Emergencies" from "Health topics"
5[a]	• WHO Regional Office for the Americas [14]	• Epidemiological Alerts and Updates
6[a]	• WHO Regional Office for Africa [15]	• Outbreaks and other emergencies
7[a]	• ECDC[c] [16]	• News & events
8[a,d]	• MOH[e] (Kingdom of Saudi Arabia) [17]	• Command and control center
9[a,d]	• Travel Health Service, Department of Health, Hong Kong, China [18]	• "Travel Health News"
10[a]	• CDC[f] (US) [19]	• "Outbreaks"
11[a,d]	• MOH (Singapore) [20]	• "Weekly Infectious Diseases Bulletin" in "disease update"
12[a]	• Global Polio Education Initiative [21]	• "This week" in "Polio Today"
13[g]	• Program for Monitoring Emerging Diseases, International Society for Infectious Diseases [22]	• Need to distinguish human or animal diseases
14[a,d]	• Centre for Health Protection, Department of Health (Hong Kong, China) [23]	• "Outbreaks"
15[a,d]	• CDC (South Korea) [24]	• "Domestic Infectious Disease Occurrence" from "Archives"
16[a,d]	• Department of Disease Control (Thailand) [25]	• Weekly Disease Forecast
17[g]	• Outbreak News Today, satellite of The Global Dispatch Inc [26]	• Recent posts
18[g]	• Global Incident Map [27]	• 50 newest events
19[a]	• Pan American Health Organization [28]	• "Epidemiological Alerts and Updates"
20[a]	• Travel health notices, Government of Canada [29]	• "Travel health notices"
21[g]	• Center for Infectious Disease Research and Policy, Academic Health Center, University of Minnesota (MN, US) [30]	• "News and Perspective"
22[a,d]	• CDC, MOH (Taiwan) [31]	• Professional and public version
23[a,d]	• Unit for Communicable Disease Prevention and Diseases Surveillance, CDC (Macau, China) [32]	• "latest news"
24[a,d]	• Chinese National Influenza Center [33]	• "Weekly Report"
25[g]	• HealthMap, Harvard University (MA, US) [34]	• "Outbreaks Near Me"
26[a,d]	• Nigeria Center for Disease Control and Prevention [35]	• "Weekly Epidemiological Report" from "Publications"

XSL•FO
RenderX

[a]Official public websites.

[b]WHO: World Health Organization.

[c]ECDC: European Centre for Disease Prevention and Control.

[d]Users pay close attention to epidemic information from this area.

[e]MOH: Ministry of Health.

[f]CDC: Centers for Disease Control and Prevention.

[g]This website gives comprehensive and timely epidemic information.

Meanwhile, several research organizations and government departments have their own information systems (eg, the Global Public Health Intelligence Network from the Public Health Agency of Canada). These websites were not used in the GIDEIMS due to their limitations or unavailability in China [36-40] (Table 2). At present, the GIDEIMS automatically collects information concerning more than 200 infectious diseases from 26 established official epidemic websites and Baidu news [41]. Web crawlers, which are used to retrieve information from websites and can be modified according to the requirements of the user, provide an advanced technique for information searching [42].

Table 2. Established and/or official public websites not used in the Global Infectious Diseases Epidemic Information Monitoring System.

Website number	Name and responsible department	Limitations
1	• Medisys, The European Union [43]	• Some information on global infectious diseases is delayed (eg, Ebola from Democratic Republic of the Congo).
2	• Biocaster, National Institute of Informatics (Japan) [7,8]	• This project ran from 2012 to 2016; currently unavailable.
3	• Epispider, CDC[a] (US) [7,9]	• Currently unavailable from the mentioned website address and CDCa (US) websites.
4	• Project Argus, Georgetown University (Washington, DC, US) and MITRE Corporation (VA, US), sponsored by the Government of the US [7,36]	• Currently unavailable from the mentioned website address.
5	• The RODS[b] Open Source Project, RODS Library, University of Pittsburgh (PA, US) [9]	• The software requires purchasing a license, and the website has not been updated since 2009.
6	• GPHIN[c], Public Health Agency of Canada [36,37]	• Currently unavailable in the websites of the Public Health Agency of Canada. The WHO[d] website contains the main information collected by the GPHIN [37].
7	• Google Inc [44]	• The same visitor is not allowed to visit the site frequently, Chinese government also has restrictions on access to Google.

[a]CDC: Centers for Disease Control and Prevention.

[b]RODS: Real-time Outbreak and Disease Surveillance.

[c]GPHIN: Global Public Health Intelligence Network.

[d]WHO: World Health Organization.

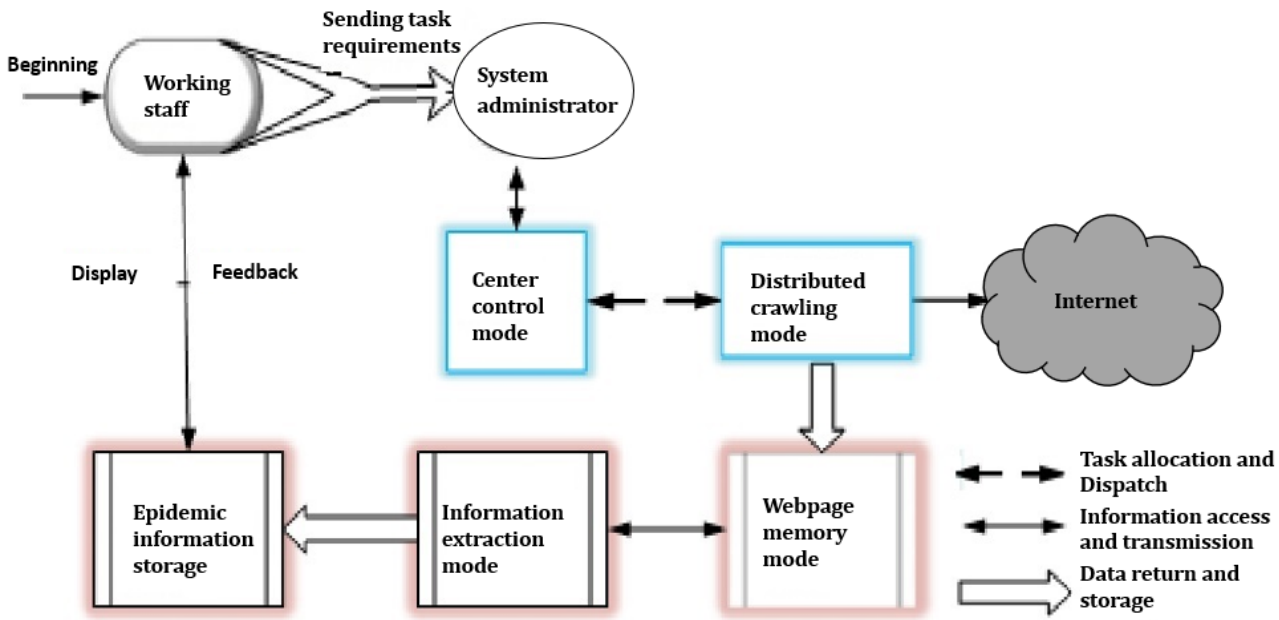
Structure

The GIDEIMS uses distributed web crawler [45,46] and cloud acceleration technologies to automatically collect epidemic information. Natural language processing and in-depth learning technologies are used to intelligently process epidemic information collected in 28 languages. The language translation service is provided by Baidu Translate, which is recognized as one of most reliable online translation services in China [47]. Cloud acceleration technology is used to improve work performance of the system.

Figure 1 shows the principle of the design of the GIDEIMS. In the figure, working staff refers to travel health officers; they provided the addresses/names of the epidemic information websites and requirements to the GIDEIMS administrator. The GIDEIMS administrator operated the center control mode according to the requirements set by the working staff. Subsequently, a series of webpage crawling, information extraction, and processing operations were performed.

XSL•FO
RenderX

Figure 1. Principle of the Global Infectious Diseases Epidemic Information Monitoring System design.



The center control mode is the core mode of the entire system. It includes the setup of the crawling strategy, assignment of tasks, and management, testing, and debugging of the system by the administrator of the GIDEIMS.

The distributed crawling mode is the execution mode of the system and the key to determining the efficiency of the entire system. In this mode, multiple crawling machines cooperate to find webpages on the internet and jointly complete the crawling task.

The main task of the webpage memory mode is to find webpages on the internet using the crawler mode and store them in a particular structure. The main task of the information extraction mode is to define extraction rules according to acquisition tasks set by the user and webpage characteristics. Moreover, this mode extracts the information available on the webpage identified by the crawler mode according to the rules and transmits the extracted results to the storage mode.

The purpose of the epidemic information storage mode is to develop a data table related to the extraction rules defined by the information extraction mode.

Results

Main Functions of the GIDEIMS

The GIDEIMS includes 7 functions: (1) GIDEIMS map, (2) latest outbreaks, (3) epidemic inquiry, (4) epidemic information entry, (5) general analysis, (6) basic setup, and (7) further functions.

GIDEIMS Map

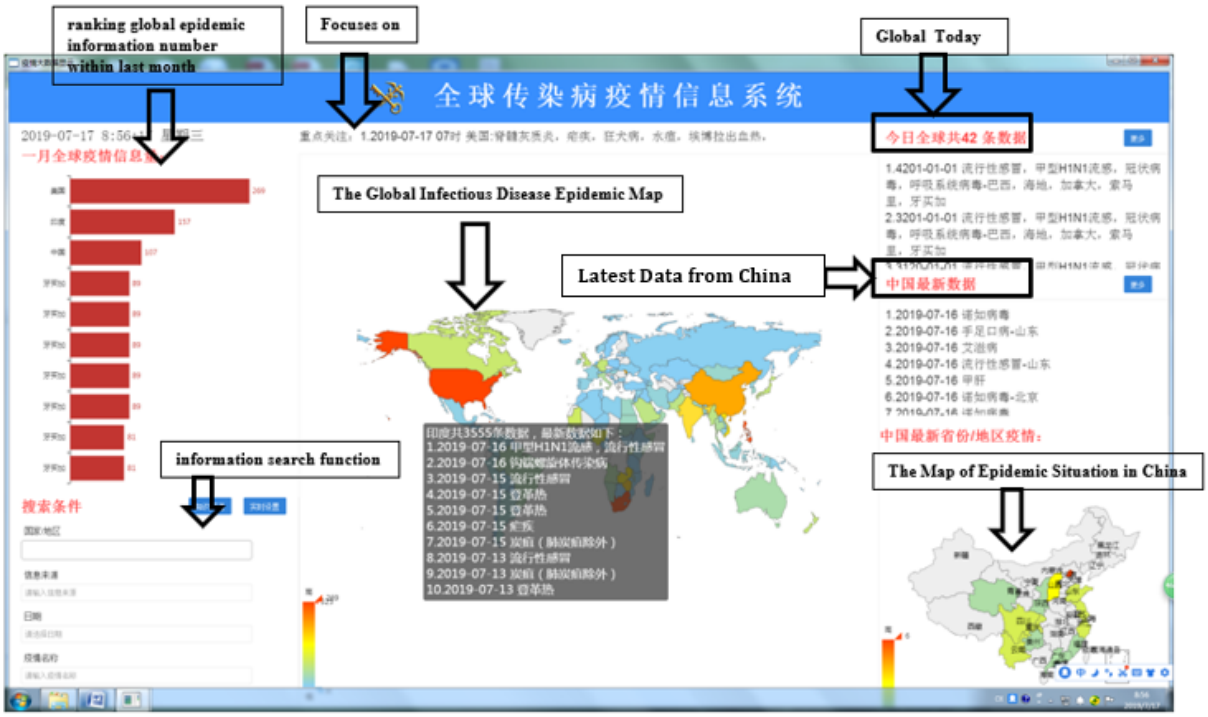
The GIDEMS map shows the latest available information worldwide and in China. Information is acquired and shown automatically by the virtual private network.

The map has two submaps: (1) a global infectious disease epidemic map (GDM), and (2) a map of the epidemic situation in China (ESC) (Figure 2). The GDM presents the epidemic information for each country using different colors and can update the epidemic information for different countries in 3 seconds. The second map, the ESC, shows the epidemic information available for different regions of China, differentiated by color. When the mouse hovers over a region on the map, a list of the captured epidemic information from this region will be automatically displayed. By clicking on the list, users can obtain the relevant detailed information. For user interest, the map shows epidemic information by countries/regions, global today (list of global epidemic information available within the last 24 hours), latest data from China (list of domestic epidemic information available within the last month), ranking of the amount of global epidemic information available within the last month (ie, the amount of epidemic information available for each country, ranked from most to least), and an information search function.

The GIDEIMS map provides information and data services with simple interfaces and convenient operation. Travelers and travel health officers can simply identify the global human infectious disease epidemic situation at a glance.

XSL•FO
RenderX

Figure 2. The Global Infectious Diseases Epidemic Information Monitoring System map.



Latest Outbreaks

In the latest outbreaks part of the system, the epidemic situation over the last 48 hours is shown, and information is updated hourly. Epidemic situations can be classified based on different data source websites and different kinds of diseases. At present, this mode is updated on an hourly basis. A large amount of data is intelligently analyzed using the background cloud server to automatically search and mine infectious disease epidemic information from official epidemic websites. The GIDEIMS can intelligently extract information including name of the epidemic disease, occurrence area, and epidemic description information from the searched webpages. Subsequently, the system automatically translates the data into Chinese and displays them to the users, facilitating the understanding of the epidemic information by users. The GIDEIMS also provides links to the original webpages and web snapshots to facilitate verification.

Epidemic Inquiry

The epidemic information collected by the GIDEIMS contains the resource titles, country/area, name of the infectious disease, reported date, crawling date, information sources, number of infected cases, number of deaths, and more. Users can utilize the general inquiry function to search the aforementioned epidemic information. The query results can be displayed in the format of lists or map reports. The outbreaks in different regions and countries can be viewed at a glance.

Users can access announcements from the health administrative department. These announcements contain the main introduction of disease, epidemic information, quarantine measures for travelers, advice for travelers heading to the area of the epidemic. Chinese travel health officers should follow the regulations of the administrative department—for example, a certain country with a severe COVID-19 epidemic situation has been regarded as a focus by the administrative department, and travelers from that country should receive COVID-19–related examinations when they enter China; hence, the prompt distribution of such announcements is very important for public health management.

The epidemic inquiry function contains an “important epidemic” section showing important infectious disease epidemic information, such as quarantine diseases that the Frontier Health and Quarantine Law (China) has regulated and other important epidemic diseases. Quarantine diseases include plague, cholera, and yellow fever, as stipulated by the state council of China [48]. In 2020, the essential epidemic diseases include plague, cholera, yellow fever, influenza, polio, Zika, dengue, chikungunya, malaria, measles, West Nile fever, and COVID-19. The list of diseases may be edited by the users in the basic setup section.

Based on the requirements set by the users, the number of infected cases and deaths, as well as epidemic information for areas of interest to the users, can also be accessed.

Other Functions

Other functions of the GIDEIMS are epidemic information entry, general analysis, basic setup, and further functions.

In the epidemic information entry function, travel health staff can manually enter epidemic information. The general analysis function reports epidemic information in a user-defined manner. Basic setup includes user management, important epidemic management, and important area management. For user management, the administrator may add, edit, or delete users.

XSL•FO
RenderX

Also, criteria such as “important epidemic” or “important area” could be adjusted based on the user’s requirements.

The GIDEIMS is continuously upgraded because of many factors, such as users’ needs, source websites changes, or the global epidemic situation of a certain disease. Users will be able to send epidemic information to administrative departments and be linked to defined social media.

Practical Case

In this section, we illustrate the GIDEIMS as a practical case to assess the system’s functionalities. The GIDEIMS map is seen by users via an independent path (ie, a virtual private network), while the other functions of the GIDEIMS are shown as websites.

Prior to using the GIDEIMS map, users install the specific application program provided by the developer in order to view the map (Figure 2). On July 16, 2019, the GIDEIMS collected 46 pieces of data compared with 10 pieces of data provided by the Global Incident Map and 6 pieces of data provided by Outbreak News Today (see Table 3). The data included 27 types of epidemic information from 23 countries/areas (10 websites). Two pieces of data were collected by both the Global Incident Map and Outbreak News Today (numbers 8 and 10 in Table 3).

XSL•FO
RenderX

Table 3. Epidemic information collected on July 16, 2019.

Data item number	Country/area	Title	Disease	Resource
1	DRC[a]	DRC: >2,500 cumulative cases of Ebola; 1,668 deaths [Chinese][b] [49]	Ebola	Baidu News
2	DRC	Fighting between DRC and Ebola has triggered the first confirmed case in Goma [Chinese][b] [50]	Ebola	Baidu News
3	DRC	DRC: Ebola virus disease [update] [Chinese Traditional][b] [51]	Ebola	Travel Health Service, Department of Health (Hong Kong, China)
4	DRC	Measles outbreak in the DRC [Chinese][b] [52]	Measles	Baidu News
5	DRC	WHO[c] will take up Ebola emergency declaration question for a fourth time [53]	Ebola	CIDRAP[d]
6	Myanmar	Swine flu death toll rises to 54 in Myanmar [54]	Influenza	Global Incident Map
7	Pakistan	54 more dengue cases surface [55]	Dengue	Global Incident Map
8	Singapore	Singapore reports 666 dengue cases last week [56,57]	Dengue	Global Incident Map; Outbreak News Today
9	Singapore	Dengue cases: 75 cases notified 16 Jul 2019 at 3 pm [58]	Dengue	National Environment Agency, Singapore
10	Philippines	Dengue in Zamboanga up 285 percent this year [59]	Dengue	Global Incident Map; Outbreak News Today
11	Cameroon	Cholera kills five in Far North region [60]	Cholera	Global Incident Map
12	India	H1N1 flu on the rise, 309 cases in Mumbai [61]	Influenza	Global Incident Map
13	India	Leptospirosis—India (03): (Maharashtra) [62]	Leptospirosis	ProMED-mail
14	India	Japanese encephalitis & other—India (17): (AS) [63]	Japanese encephalitis	ProMED-mail
15	India	Nipah-affected student to be discharged on Tuesday [64]	Nipah virus	Global Incident Map
16	India	8-year-old dies due to suspected dengue fever [65]	Dengue	Global Incident Map
17	Bangladesh	Nearly 2,800 cases in first 16 days of July [66]	Dengue	Global Incident Map
18	Canada	Canada: 1st human rabies case reported since 2012 [67]	Rabies	Outbreak News Today
19	Canada	Canada: Syphilis outbreak in Alberta [68]	Syphilis	Outbreak News Today
20	United States	Cryptosporidium in the US with Joseph Eisenberg, PhD [69]	Cryptosporidiosis	Outbreak News Today
21	United States	Anthrax in Texas update: Eight premises in three counties [70]	Anthrax	Outbreak News Today
22	United States	Florida reports 1,900 hepatitis A cases so far, Pasco County has seen the most [71]	Hepatitis A	Outbreak News Today
23	United States	Angiostrongylus cantonensis—US (04): (HI) [72]	Angiostrongyliasis	ProMED-mail
24	United States	Infant botulism—US (02): (TX) more cases [73]	Botulism	ProMED-mail
25	United States	Undiagnosed respiratory illness—US (02): (VA) fatal, retirement community [74]	Undiagnosed	ProMED-mail
26	United States	El Paso reports 3rd measles case [75]	Measles	Outbreak News Today
27	United States	Rise in Candida auris cases; New AMR plan in Wales [76]	Candida auris	CIDRAP[d]
28	Malaysia	Malaria—Malaysia: Pahang, resurgence [77]	Malaria	ProMED-mail
29	Saudi Arabia	MERS-CoV (57): Saudi Arabia (NJ,RI) WHO [78]	MERS[e]	ProMED-mail

XSL•FO
RenderX

Data item number	Country/area	Title	Disease	Resource
30	Saudi Arabia	WHO notes clusters in recent MERS cases, unveils environmental sampling guide [79]	MERS	CIDRAP
31	France	Salmonellosis—France: cured ham, alert, recall [80]	Salmonellosis	ProMED-mail
32	Syria	Brucellosis—Syria: (Quneitra) increasing incidence [Arabic][b] [81]	Brucellosis	ProMED-mail
33	Kenya	Anthrax—Kenya (12): (KU) human, cattle [82]	Anthrax	ProMED-mail
34	Taiwan	One newly diagnosed case of local dengue fever in Tainan [Chinese Traditional][b] [83]	Dengue	Taiwan CDC[f]
35	Taiwan	A new case of enterovirus complicated with severe illness [Chinese Traditional][b] [84]	Hand, foot and mouth disease	Taiwan CDC
36	Brazil, Cambodia, Pakistan, etc	Dengue fever or Chikungunya fever are prevalent in many countries in the world [Chinese Traditional][b] [85]	Dengue, chikungunya	Taiwan CDC
37	Europe	Rising European measles vaccination [86]	Measles	CIDRAP
38	Thailand	The disease control department recently revealed that the situation of dengue fever is slowing down [Thai][b] [87]	Dengue fever	The Department of Disease Control of Thailand
39	Nigeria	An update of Lassa fever outbreak in Nigeria for Week 27 [88]	Lassa fever	Nigeria CDC
40	Nigeria	An update of Cholera outbreak in Nigeria for Week 27 [89]	Cholera	Nigeria CDC
41	Namibia	H1N1 claims life in Windhoek [90]	Influenza	Global Incident Map
42	China	The reported incidence of hepatitis A and B in Gansu Province was significantly reduced [Chinese][b] [91]	Hepatitis A and B	Baidu News
43	China	Foodborne illness—China: Beijing, tap water, norovirus suspected [92]	Norovirus	ProMED-mail
44	China	Suspected Norovirus Infection in Residents of Wanke Qingqing Home District, Chaoyang, Beijing [Chinese][b] [93]	Norovirus	Baidu News
45	China	Shandong: In June, fewer people got the flu than those infected with hand, foot, and mouth disease [Chinese][b] [94]	Hand, foot, and mouth disease	Baidu News
46	Global	The latest report issued by UNAIDS: the mixed situation of countries [Chinese][b] [95]	HIV	Baidu News

[a]DRC: Democratic Republic of the Congo.

[b]Translated by Baidu Translate.

[c]WHO: World Health Organization.

[d]CIDRAP: Center for Infectious Disease Research and Policy.

[e]MERS: Middle East Respiratory Syndrome.

[f]CDC: Centres for Disease Control and Prevention.

For the other functions of the GIDEIMS, first, the user must access the system using a defined website address [96]. After entering the username and password, the user sees the latest outbreak page (Figure 3), which shows the last five pieces of epidemic information news from each website. By clicking the "more" button, the user gains access to further epidemic information. The "check" button displays the detailed information (Figure 4). Users can click the "original webpage link" to verify the information.

Users wishing to search, for example, for recent Zika epidemic information would click "Epidemic inquiry," followed by "General inquiry," enter the epidemic disease name "Zika (in Chinese)" in "Epidemic situation name," and limit the occurrence date to July 2019. The results are shown in Figure 5.

Shenzhen travel health officers use the epidemic information on a daily basis to guide their decision-making on infection disease epidemic management, such as identifying travelers

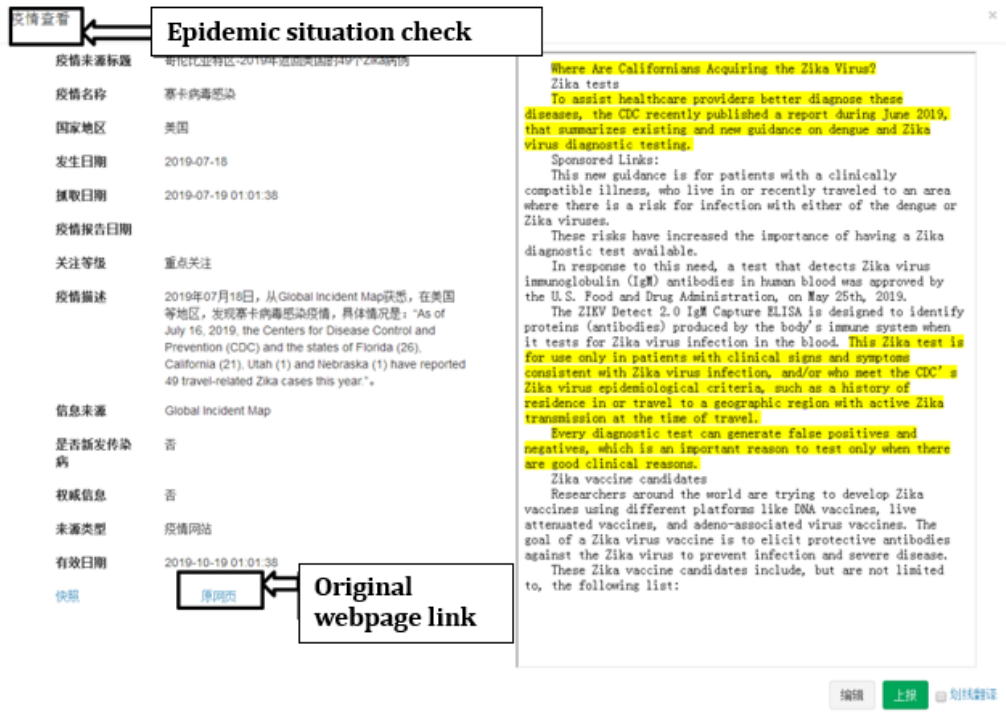
XSL•FO
RenderX

with high fever and investigating their travel history. For travelers arriving from epidemic areas of certain infectious diseases, the officers will obtain a sample and test for the infectious diseases of interest if the traveler consents.

In contrast, without the GIDEIMS, staff could spend approximately 4 hours (excluding translation time) identifying the relevant epidemic information. With the use of the GIDEIMS, staff only need to access the collected data and verify them using the links and snapshots provided by the system when necessary. Working time is shortened to less than 1 hour. The information collected by the GIDEIMS is more accessible and comprehensive than that collected through the manual method.

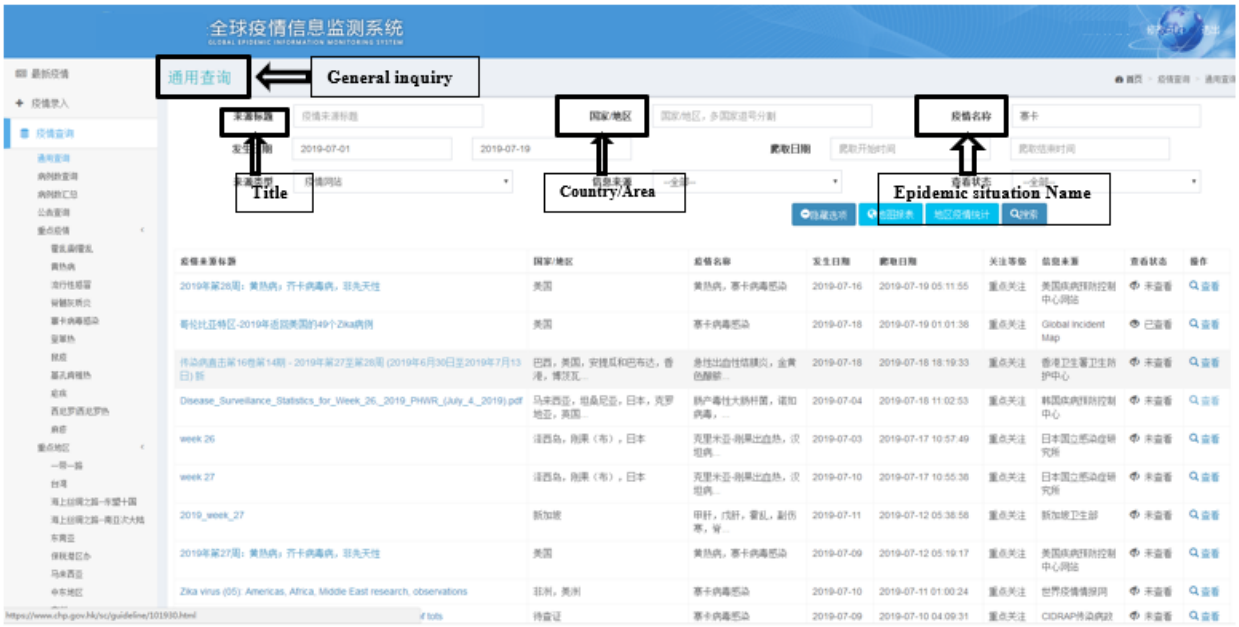
Figure 3. First page of the Global Infectious Diseases Epidemic Information Monitoring System, showing the latest outbreak.

Figure 4. Detailed information provided by the Global Infectious Diseases Epidemic Information Monitoring System.



XSL•FO
RenderX

Figure 5. Sample inquiry about the Zika epidemic situation using the general inquiry function of the Global Infectious Diseases Epidemic Information Monitoring System.



Discussion

Using the epidemic information provided by the GIDEIMS, health officers can quickly focus on travelers from certain countries and/or regions and discover infected cases as early as possible. By using the information from the GIDEIMS, staff at the Health Quarantine Center Laboratory of the Shenzhen International Travel Health Care Center (Shenzhen Customs District Port Outpatient Clinics) detected the first imported Zika cases in China [97], and thousands of suspected cases of certain infectious diseases have also been detected annually [98].Thus, while the detection work is strenuous, the GIDEIMS provides a helpful tool to effectively and sustainably identify suspected infected travelers into the Port of Shenzhen. The main advantage of the GIDEIMS is that it can automatically collect epidemic information from defined websites and translate it into the Chinese language. Although occasionally the system may duplicate information and require a manual check, most of the users of the GIDEIMS—such as the travel health officers working at Customs, travelers, researchers, and others working in the infectious disease control and prevention sector—reveal that the GIDEIMS saves time and is less labor-intensive. The first version of the GIDEIMS was developed in 2016 [99]. It is constantly upgraded to fulfill the requirements of users and adjusted according to actual situations. We constructed a visual display platform for the global infectious diseases epidemic information. The GIDEIMS is a user-friendly tool to support both travelers and travel health officers in travel health management. Meanwhile, big data obtained from the GIDEIMS may be used for infectious disease surveillance management and control.

In regard to COVID-19, information systems have largely been built to address almost every aspect of health management, including infection situation data management, remote health care system management, and syndromic surveillance [100-102]. In comparison, the GIDEIMS provides more comprehensive disease information than other COVID-19 epidemic information systems. Compared with the geographic information system for global monitoring of COVID-19 established by Johns Hopkins University [103], whose initial data were collected from the World Health Organization (WHO), US Centers for Disease Control and Prevention (CDC), China CDC, European Centre for Disease Prevention and Control (ECDC), National Health Commission of China (NHC), and DXY (a Chinese health-focused social website), the GIDEIMS could provide a full picture of the worldwide situation of the COVID-19 epidemic.

GIDEIMS breaks the barriers of language, region, time difference, and more so that it can provide enormous amounts of real-time infectious disease-related information. The system is not targeted on the individual traveler, but it provides travelers with useful information on the epidemic situation of infectious diseases.

At present, the GIDEIMS is a nonprofit application information system, where the maintenance and upgrade operations are performed by the developers free of charge. Due to the limitations of funds and human resources, the system is available in Chinese only. It is mainly provided to relevant infectious disease prevention and control departments and personnel for a free trial.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the Health and Quarantine Office of Shenzhen Customs for its advice and comments for the improvement of the GIDEIMS.

This research was supported by the National Key Research and Development Program of China (No. 2018YFC0809200, No. 2016YFF0203203); the National Natural Science Foundation of China (No. 81703271); Guangdong Science and Technology Foundation (No. 2017B020210006, No. 2016A020219005); Shenzhen Science and Technology Foundation (No. SGLH20180625171602058, CKCY20170720100145297, No. JCYJ20160427151920801, No. JCYJ20170307104024209), and the Open Project of Key Laboratory of Tropical Disease Control of the Ministry of Education (Sun Yat-sen University, No. 2019kfkt06).

The funding sources played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

All authors participated in designing the study, analyzing the results, and writing the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Jack A. Why the panic? South Korea's MERS response questioned. BMJ 2015 Jun 24;350:h3403. [doi: 10.1136/bmj.h3403] [Medline: 26108610]
2. Barbisch D, Koenig KL, Shih FFY. Is There a Case for Quarantine? Perspectives from SARS to Ebola. Disaster Medicine and Public Health Preparedness 2015 Oct;9(5):547-553. [doi: 10.1017/dmp.2015.38] [Medline: 25797363]
3. Shenzhen 2019 GDP up 6.7%. Shenzhen Government Online.: Shenzhen Daily URL: http://www.sz.gov.cn/en_szgov/news/latest/content/post_7149127.html [accessed 2020-04-21]
4. Shenzhen clears 660,000 entry-exit passengers daily in 2019. Xinhuanet. URL: http://www.xinhuanet.com/english/2020-01/15/c_138706890.htm [accessed 2020-01-15]
5. Jiwen Z, Jin X, Zhaoyin Z, Huang L, Liang S, Xu Y, et al. Early detection of Zika virus infection among travellers from areas of ongoing transmission in China. Journal of Travel Medicine 2016;23(5). [doi: 10.1093/jtm/taw047]
6. Bo L, Zhengliang L, Bo S, Zong D, Qinlan L, Qian W, et al. Sensitive HIV-1 detection in a homogeneous solution based on an electrochemical molecular beacon coupled with a nafion-graphene composite film modified screen-printed carbon electrode. Biosensors & Bioelectronics 2014 Feb 15;52C:330-336. [doi: 10.1016/j.bios.2013.09.016] [Medline: 24099877]
7. Anette H, Gustaf R, Annika L. Web queries as a source for syndromic surveillance. PLoS One 2009;4(2):e4378 [FREE Full text] [doi: 10.1371/journal.pone.0004378] [Medline: 19197389]
8. Malwina C, Lawrence C M. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. International Health 2017 May 01;9(3):177-183 [FREE Full text] [doi: 10.1093/inthealth/ihx014] [Medline: 28582558]
9. Philippe B, Laetitia V, Yann Le S, David M H, Noele P N, Abla M, et al. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. PLoS One 2014;9(3):e90536 [FREE Full text] [doi: 10.1371/journal.pone.0090536] [Medline: 24599062]
10. World Health Organization. URL: http://www.who.int/en/ [accessed 2019-07-12]
11. World Health Organization Western Pacific Region. URL: https://www.who.int/westernpacific/ [accessed 2019-07-12]
12. World Health Organization Regional Office for the Eastern Mediterranean. URL: http://www.emro.who.int/index.html [accessed 2019-07-12]
13. World Health Organization Regional Office for Europe. URL: http://www.euro.who.int/en/home [accessed 2019-07-12]
14. World Health Organization Regional Office for the Americas. URL: https://www.paho.org/hq/index.php?lang=en [accessed 2019-07-12]
15. World Health Organization Regional Office for Africa. URL: https://www.afro.who.int/ [accessed 2019-07-12]
16. European Centre for Disease Prevention and Control. URL: https://ecdc.europa.eu/en/home [accessed 2019-07-12]
17. Ministry of Health, Kingdom of Saudi Arabia. URL: https://www.moh.gov.sa [accessed 2019-07-12]
18. Travel Health Service, Department of Health, Hong Kong, China. URL: https://www.travelhealth.gov.hk [accessed 2019-07-12]
19. Centre for Disease Control and Prevention, USA. URL: https://www.cdc.gov/ [accessed 2019-07-12]
20. Ministry of Health, Singapore. URL: https://www.moh.gov.sg/ [accessed 2019-07-12]
21. Global Polio Education Initiative. URL: http://polioeradication.org/polio-today/polio-now/ [accessed 2019-07-12]
22. Program for Monitoring Emerging Diseases, International Society for Infectious Diseases. URL: http://www.promedmail.org/ [accessed 2019-07-12]

XSL•FO
RenderX

23. Centre for Health Protection, Department of Health, Hong Kong, China. URL: https://www.chp.gov.hk/ [accessed 2019-07-12]
24. Centre for Disease Control and Prevention, South Korea. URL: http://www.cdc.go.kr/CDC/eng/main.jsp [accessed 2019-07-12]
25. Department of Disease Control, Thailand. URL: http://o.ddc.moph.go.th/eng/index.php [accessed 2019-07-12]
26. Outbreak News Today, satellite of The Global Dispatch Inc. URL: http://outbreaknewstoday.com/ [accessed 2019-07-12]
27. Global incident map. URL: http://outbreaks.globalincidentmap.com/home.php [accessed 2019-07-12]
28. Pan American Health Organization. URL: http://www.paho.org/hq/ [accessed 2019-07-12]
29. Government of Canada. URL: https://travel.gc.ca/travelling/health-safety/travel-health-notices/ [accessed 2019-07-12]
30. Center for Infectious Disease Research and Policy, Academic Health Center, University of Minnesota, MN, USA. URL: http://www.cidrap.umn.edu/news-perspective [accessed 2019-07-12]
31. Centre for Disease Control and Prevention, Ministry of Health, Taiwan. URL: http://www.cdc.gov.tw [accessed 2019-07-12]
32. Unit for Communicable Disease PreventionDiseases Surveillance, CDC, Macau, China. URL: https://www.ssm.gov.mo/portal/ [accessed 2019-07-12]
33. Chinese National Influenza Center. URL: http://www.chinaivdc.cn/cnic/zyzx/ [accessed 2019-07-12]
34. HealthMap, Harvard University, MA, USA. URL: https://www.healthmap.org/ [accessed 2019-07-12]
35. Nigeria Center for Disease Control and Prevention. URL: https://ncdc.gov.ng/ [accessed 2019-07-12]
36. 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]
37. Hung Quoc N. The Use of the BioCaster Ontology for Mapping Infectious Diseases and Locations in the BioCaster Surveillance System. 2018 Jan 15 Presented at: BioLINK; January 2008; Canada URL: https://www.researchgate.net/profile/Hung_Ngo3/publication/259290195_The_Use_of_the_BioCaster_Ontology_for_Mapping_Infectious_Diseases_and_Locations_in_the_BioCaster_Surveillance_System/links/0deec52ac62fcc1b21000000.pdf [doi: 10.1038/npre.2008.2110.1]
38. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. Emerg Infect Dis 2009 May;15(5):689-695 [FREE Full text] [doi: 10.3201/eid1505.081114] [Medline: 19402953]
39. Xuemei S, Qiang C, Ming W, Lin H. The Present Status and Thought on Network-media-based Information Resources of Disease Outbreak and Surveillance. Yi Xue Xin Xi Xue Za Zhi 2013;30(4):1-5 [FREE Full text] [doi: 10.18352/bmgn-lchr.5703]
40. Office of Audit and Evaluation, Health Canada and the Public Health Agency of Canada. Evaluation of Emergency Preparedness and Response Activities 2012-13 to 2016-17, 4.4.2. Government of Canada. 2018. URL: https://www.canada.ca/en/public-health/corporate/transparency/corporate-management-reporting/evaluation/2012-2013-2016-2017-evaluation-report-emergency-preparedness-response-activities.html [accessed 2018-03-20]
41. Baidu news. URL: news.baidu.com [accessed 2021-02-07]
42. S.R. S, Sangita C. Review of web crawlers. International Journal of Knowledge & Web Intelligence 2014;5(1):49. [doi: 10.1504/IJKWI.2014.065035]
43. Medisys, The European Union. URL: http://medisys.newsbrief.eu/medisys/helsinkiedition/zh/home.html [accessed 2019-07-12]
44. Google Inc. URL: www.google.com [accessed 2019-07-12]
45. Shuai W, Guomin Z, Jian W. Reviews of Relevance Algorithm in Focused Crawler. JI SUAN JI YU XIAN DAI HUA 2009;30(2):27-30 [FREE Full text] [doi: 10.3969/j.issn.1006-2475.2013.04.007]
46. Jiang Y. Efficiency bottlenecks analysis and solution of Web crawler. Journal of Computer Applications 2008 Oct 21;28(5):1114-1116 [FREE Full text] [doi: 10.3724/SP.J.1087.2008.01114]
47. Zhongjun H. Baidu Translate: Research and Products. 2015 Presented at: Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra); 2015; Beijing. [doi: 10.18653/v1/w15-4110]
48. Standing Committee of National People's Congress of P.R.China. Frontier Health and Quarantine Law of the P.R China. China.org.cn. URL: http://www.china.org.cn/english/travel/40418.htm [accessed 2019-07-16]
49. Sina Corporation. URL: http://news.sina.com.cn/w/2019-07-16/doc-ihytcitm2491479.shtml [accessed 2019-07-16]
50. www.gangaonet.com. URL: http://www.gangaonet.com/zhusanjiao/2019/0716/116143.html [accessed 2019-07-16]
51. Travel Health Service, Department of Health, Hong Kong, China. URL: https://www.travelhealth.gov.hk/sc_chi/outbreaknews/2019/ond09July2019.html [accessed 2019-07-16]
52. cnBeta.com. URL: http://www.cnbeta.com/articles/science/868281.htm [accessed 2019-07-16]
53. Center for Infectious Disease Research and Policy. URL: http://www.cidrap.umn.edu/news-perspective/2019/07/who-will-take-ebola-emergency-declaration-question-fourth-time [accessed 2019-07-16]
54. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32264 [accessed 2019-07-16]
55. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32261 [accessed 2019-07-16]
56. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32260 [accessed 2019-07-16]

XSL•FO
RenderX

57. Outbreak News Today. URL: http://outbreaknewstoday.com/singapore-reports-666-dengue-cases-last-week-highest-weekly-total-in-3-1-2-years-76821/ [accessed 2019-07-16]
58. National Environment Agency, Singapore. URL: http://www.nea.gov.sg/ [accessed 2019-07-16]
59. Outbreak News Today. URL: http://outbreaknewstoday.com/dengue-in-zamboanga-up-285-percent-this-year-28038/ [accessed 2019-07-16]
60. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32258 [accessed 2019-07-16]
61. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32257 [accessed 2019-07-16]
62. ProMED-mail. URL: http://www.promedmail.org/ajax/getPost.php?alert_id=6569127 [accessed 2019-07-16]
63. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6570013 [accessed 2019-07-16]
64. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32296 [accessed 2019-07-16]
65. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32281 [accessed 2019-07-16]
66. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32270 [accessed 2019-07-16]
67. Outbreak News Today. URL: http://outbreaknewstoday.com/canada-1st-human-rabies-case-reported-since-2012/ [accessed 2019-07-16]
68. Outbreak News Today. URL: http://outbreaknewstoday.com/canada-syphilis-outbreak-in-alberta-89874/ [accessed 2019-07-16]
69. Outbreak News Today. URL: http://outbreaknewstoday.com/cryptosporidium-in-the-us-with-joseph-eisenberg-ph-d-80144/ [accessed 2015-07-16]
70. Outbreak News Today. URL: http://outbreaknewstoday.com/anthrax-in-texas-update-eight-premises-in-three-counties-84701/ [accessed 2019-07-16]
71. Outbreak News Today. URL: http://outbreaknewstoday.com/florida-reports-1900-hepatitis-a-cases-so-far-pasco-county-has-seen-the-most-83585/ [accessed 2019-07-15]
72. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6570547 [accessed 2019-07-15]
73. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6571003 [accessed 2019-07-15]
74. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6570174 [accessed 2019-07-15]
75. Outbreak News Today. URL: http://outbreaknewstoday.com/el-paso-reports-3rd-measles-case-48365/ [accessed 2019-07-16]
76. Center for Infectious Disease Research and Policy. URL: https://www.cidrap.umn.edu/news-perspective/2019/07/stewardship-resistance-scan-jul-16-2019 [accessed 2019-07-16]
77. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6571569 [accessed 2019-07-16]
78. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6571651 [accessed 2019-07-16]
79. Center for Infectious Disease Research and Policy. URL: http://www.cidrap.umn.edu/news-perspective/2019/07/who-notes-clusters-recent-mers-cases-unveils-environmental-sampling-guide [accessed 2019-07-16]
80. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6569722 [accessed 2019-07-16]
81. ProMED-mail. URL: http://www.promedmail.org/ajax/getPost.php?alert_id=6568527 [accessed 2019-07-16]
82. ProMED-mail. URL: https://promedmail.org/promed-post/?id=6569704 [accessed 2019-07-16]
83. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/Bulletin/Detail/U_YChDpaFYt4T2kywZjqBg?typeid=9 [accessed 2019-07-16]
84. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/Bulletin/Detail/GrqbboCXcIks-4mNSX8mvA?typeid=9 [accessed 2019-07-16]
85. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/Bulletin/Detail/qDb13SZSYwmgZ00rbAWatw?typeid=9 [accessed 2019-07-16]
86. Center for Infectious Disease Research and Policy. URL: https://www.cidrap.umn.edu/news-perspective/2019/07/news-scan-jul-16-2019 [accessed 2019-07-16]
87. The Department of Disease Control of Thailand. URL: https://ddc.moph.go.th/th/site/newsview/view/7261 [accessed 2019-07-16]
88. Nigeria Center for Disease Control and Prevention. URL: https://ncdc.gov.ng/themes/common/files/sitreps/a6a67b9036f6beb16adebfb11e65190f.pdf [accessed 2019-07-16]
89. Nigeria Center for Disease Control and Prevention. URL: https://ncdc.gov.ng/themes/common/files/sitreps/667b55aec7fb889b0b4293e79320a4ac.pdf [accessed 2019-07-16]
90. Global Incident Map. URL: http://outbreaks.globalincidentmap.com/eventdetail.php?ID=32268 [accessed 2019-07-16]
91. www.lanzhou.cn. URL: http://news.lanzhou.cn/system/2019/07/16/011794485.shtml [accessed 2019-07-16]
92. ProMED-mail. URL: http://www.promedmail.org/ajax/getPost.php?alert_id=6570349 [accessed 2019-07-16]
93. Sina Corporation. URL: http://finance.sina.com.cn/stock/relnews/cn/2019-07-16/doc-ihytcerm4014116.shtml [accessed 2019-07-16]
94. Xinhuanet. URL: http://www.sd.xinhuanet.com/news/2019-07/16/c_1124759851.htm [accessed 2019-07-16]
95. CN-Healthcare. URL: https://www.cn-healthcare.com/article/20190716/content-521492.html [accessed 2019-07-16]
96. Global Infectious Diseases Epidemic Information Monitoring System. URL: http://yq.ssdat.com:6090/quarantineSys/user/login?returnUrl=%2Fhome%2Findex [accessed 2021-02-07]

XSL•FO
RenderX

97. Chenglin D, Si-Qing L, Qiuyan Z, Mingyue X. Isolation and characterization of Zika virus imported to China using C6/36 mosquito cells. Virologica Sinica 2016;31(2):176-179 [FREE Full text] [doi: 10.1007/s12250-016-3778-5]
98. Lei S, Yang F, Yanqiu Y, Yuzhen H, Yunqing X, Dayong Gu G, et al. Epidemiological characteristics of 771 influenza cases in Shenzhen port. Zhong Guo Gong Gong Wei Sheng 2012;28(10):1352-1354 [FREE Full text] [doi: 10.11847/zgggws2012-28-10-31]
99. Xiangjun W, Chunxiao L, Muyan D, Jianan H, Dayong G, Lei S, et al. Developent of automatic collection system of global epidemic informations. Zhong Guo Guo Jing Wei Sheng Jian Yi Za Zhi 2017;40(6):431-434. [doi: 10.16408/j.1004-9770.2017.06.013]
100. Gerfalk PJ, Conboy K, Myers MD. Information systems in the age of pandemics: COVID-19 and beyond. European Journal of Information Systems 2020 Jun 14;29(3):203-207. [doi: 10.1080/0960085X.2020.1771968]
101. Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol 2020 Jun 22;33(11):2115-2127. [doi: 10.1038/S41379-020-0601-5]
102. Luo H, Lie Y, Prinzen FW. Surveillance of COVID-19 in the General Population Using an Online Questionnaire: Report From 18,161 Respondents in China. JMIR Public Health Surveillance 2020 Apr 27;6(2):e18576. [doi: 10.2196/18576]
103. Belmiro Do Nascimento J. Geographic Information Systems and COVID-19: The Johns Hopkins University DashboardJ Research Square. Research Square 2020. [doi: 10.21203/rs.3.rs-15447/v1]

Abbreviations

CDC: Centers for Disease Control and Prevention
ECDC: European Centre for Disease Prevention and Control
ESC: epidemic situation in China
GDM: global infectious disease epidemic map
GIDEIMS: Global Infectious Diseases Epidemic Information Monitoring System
MERS: Middle East Respiratory Syndrome
NHC: National Health Commission of China
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.09.20; peer-reviewed by JM Ocampo, K Hung; comments to author 03.11.20; revised version received 23.11.20; accepted 19.12.20; published 16.02.21.

Please cite as:
Gu D, He J, Sun J, Shi X, Ye Y, Zhang Z, Wang X, Su Q, Yu W, Yuan X, Dong R
The Global Infectious Diseases Epidemic Information Monitoring System: Development and Usability Study of an Effective Tool for Travel Health Management in China
JMIR Public Health Surveill 2021;7(2):e24204
URL: http://publichealth.jmir.org/2021/2/e24204/
doi:10.2196/24204
PMID:33591286

©Dayong Gu, Jianan He, Jie Sun, Xin Shi, Ying Ye, Zishuai Zhang, Xiangjun Wang, Qun Su, Wenjin Yu, Xiaopeng Yuan, Ruiling Dong. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 16.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Expert Opinions on the Most Promising Treatments and Vaccine Candidates for COVID-19: Global Cross-sectional Survey of Virus Researchers in the Early Months of the Pandemic

Bernardo Pereira Cabral[1,2*], BA, MPhil, PhD; Luiza Braga[2,3*], BA, MPhil; Fabio Mota[2*], BA, MPhil, PhD

[1]Faculty of Economics, Federal University of Bahia, Salvador, Brazil
[2]Center for Strategic Studies, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[3]Faculty of Economics, Fluminense Federal University, Niterói, Brazil
[*]all authors contributed equally

Corresponding Author:
Fabio Mota, BA, MPhil, PhD
Center for Strategic Studies
Oswaldo Cruz Foundation
Avenida Brasil, 4036
Manguinhos
Rio de Janeiro, 21040-361
Brazil
Phone: 55 2138829133
Email: fabio.mota@fiocruz.br

Abstract

Background: The COVID-19 pandemic presents a great public health challenge worldwide, especially given the urgent need to identify effective drugs and develop a vaccine in a short period of time. Globally, several drugs and vaccine candidates are in clinical trials. However, because these drugs and vaccines are still being tested, there is still no definition of which ones will succeed.

Objective: This study aimed to assess the opinions of over 1000 virus researchers with knowledge on the prevention and treatment of coronavirus-related human diseases to determine the most promising drug and vaccine candidates to address COVID-19.

Methods: We mapped the clinical trials related to COVID-19 registered at ClinicalTrials.gov. These data were used to prepare a survey questionnaire about treatments and vaccine candidates for COVID-19. In May 2020, a global survey was conducted with authors of recent scientific publications indexed in the Web of Science Core Collection related to viruses, severe acute respiratory syndrome coronavirus, coronaviruses, and COVID-19.

Results: Remdesivir, immunoglobulin from cured patients, and plasma were considered to be the most promising treatments in May 2020, while ChAdOx1 and mRNA-1273 were considered to be the most promising vaccine candidates. Almost two-thirds of the respondents (766/1219, 62.8%) believed that vaccines for COVID-19 were likely to be available in the next 18 months. Slightly fewer than 25% (289/1219, 23.7%) believed that a vaccine was feasible, but probably not within 18 months.

Conclusions: The issues addressed in this study are constantly evolving; therefore, the current state of knowledge has changed since the survey was conducted. However, for several months after the survey, the respondents' expectations were in line with recent results related to treatments and vaccine candidates for COVID-19.

(JMIR Public Health Surveill 2021;7(2):e22483) doi:10.2196/22483

KEYWORDS

COVID-19; SARS-CoV-2; vaccine; treatment; survey; public health; drug; clinical trial

XSL•FO
RenderX

Introduction

In December 2019, in Wuhan, China, a new coronavirus disease characterized by fever, cough, dyspnea, chills, muscle pain, sore throat, and loss of taste and smell [1] rapidly spread worldwide and was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [2]. By September 1, 2020, this novel coronavirus disease, COVID-19, caused by SARS-CoV-2, had resulted in more than 860,000 deaths in 216 countries, areas, or territories [3]. According to the WHO, the focus of disease management efforts to date has been on infection prevention, detecting and monitoring cases, and supportive care [4]. Currently, there is no specific treatment for COVID-19; instead, the disease is managed using a large group of experimental therapies of different types [5]. Currently, hundreds of registered clinical trials worldwide are testing the dosages and adequacy of different drugs to treat different stages of the disease [6]. Similar efforts are being made to create a vaccine [7,8]. Finding the most promising lines of treatment and vaccines to reduce the burden of COVID-19 is probably the greatest scientific and technological challenge of our time.

Despite the myriad of therapeutics and vaccine candidates being tested worldwide, it is still unclear which ones will prove successful in addressing COVID-19. Some previous studies have attempted to map the most promising treatments and vaccine candidates to fight COVID-19 [6-8]; however, these studies place little or no emphasis on experts' opinions. Our study addresses this gap by assessing the opinions of over 1000 experts involved in virus research worldwide with knowledge on the prevention and treatment of coronavirus-related human diseases. Clinical trial data were obtained from ClinicalTrials.gov, and the survey's respondents were identified from scientific publications indexed in the Web of Science (WoS) Core Collection related to viruses, severe acute respiratory syndrome coronavirus (SARS-CoV), coronaviruses, and COVID-19.

Methods

We mapped the clinical trials related to COVID-19 registered at ClinicalTrials.gov (Multimedia Appendix 1) using the Aggregate Analysis of ClinicalTrials.gov (AACT) database, maintained by the Clinical Trials Transformation Initiative [9]. The search was conducted on April 23, 2020, via the PostgreSQL AACT connection. It retrieved 406 clinical trials with coronavirus as the variable condition (diseases being tested). The records were imported into VantagePoint, version 11.0 (Search Technology Inc); we then selected only the trials related to COVID-19 by searching the condition, title, outcome measure, keywords, and Medical Subject Headings (MeSH) conditions fields for the following descriptors: *Covid-19*, *Covid19*, *Covid 19*, *Coronavirus-19*, *Coronavirus 19*, *Coronavirus 2019*, *Sars-CoV2*, *SARS-CoV-2*, *SARS COV 2*, *Novel Coronavirus*, *2019 Novel Coronavirus*, *2019-Novel coronavirus*, *2019-nCoV*, *2019nCoV*, and *Coronavirus Disease 2019*.

This search retrieved 361 clinical trials: 237 (65.7%) were interventional studies, in which biomedical interventions were assigned and tested, and 124 (34.3%) were observational studies, in which no specific intervention was assigned by the principal investigator. Considering the purpose of this study, observational studies were not included in the analysis. Of the 237 interventional studies, only those pertaining to drugs (n=173, 73.0%) and biologicals or vaccines (n=31, 17.9%) were selected. Clinical trials whose status was withdrawn (Phase 2) or suspended (Phase 1) were excluded. To select only clinical trials of vaccine candidates, we searched the studies of biologicals and vaccines for the terms *vaccine* and *vaccines* in the following fields: interventions, keywords, MeSH interventions, outcome measures, study designs, and title. At the end of these procedures, 170 clinical trials with drug interventions and 6 clinical trials with vaccine interventions were selected for analysis.

Then, these clinical trial data were used to prepare the survey questionnaire, which was structured in four parts. The first part was an introduction to the survey (purpose, respondents' role, estimated time to complete, time of availability of the questionnaire, confidentially and privacy, and consent for the use of data provided). The second part asked about the respondents' knowledge level regarding the prevention and treatment of coronavirus-related human diseases. Respondents who reported having no knowledge of these topics were disqualified from the survey and did not answer the remaining questions. The third part covered the vaccine candidates and therapeutics currently in clinical trials and required the respondents to choose the ones they believed were most likely to be successful in treating or preventing COVID-19.

The fourth part of the survey consisted of demographic questions. The results of this study were not influenced by the demographics of the respondents, which were assumed to be relatively homogeneous because the respondents were all authors of scientific publications related to viruses, coronaviruses, SARS-CoV, and COVID-19, with knowledge on prevention and treatment of coronavirus-related human diseases. Nevertheless, we asked a few demographic questions to provide a broad overview of the respondents' profiles. These questions were placed at the end of the questionnaire to prevent fatigue from answering the main questions, which could increase the number of skipped questions and the overall dropout rate.

The respondents to this survey were authors of recent scientific publications indexed in the WoS related to viruses, coronaviruses, SARS-CoV, and COVID-19. These publications were identified using the query shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Query used to identify the publications by respondents to the survey.

ti=(virus*) OR ts=(Coronavirus* OR Alphacoronavirus* OR Betacoronavirus* OR Deltacoronavirus* OR Gammacoronavirus* OR "SARS Virus*" OR "Severe Acute Respiratory Syndrome Virus*" OR "SARS-Related Coronavirus*" OR "SARS-CoV" OR "SARS-Associated Coronavirus*" OR "SARS Coronavirus*" OR "Severe acute respiratory syndrome-related coronavirus*") OR ts=("COVID-19" OR COVID19 OR "2019 novel coronavirus" OR "SARS-CoV-2" OR "2019-nCoV" OR "coronavirus disease 2019" OR "coronavirus disease-19")

Refined by: research areas (Virology OR Infectious Diseases OR Immunology OR Research Experimental Medicine OR Pharmacology Pharmacy OR General Internal Medicine)

Indexes=SCI-EXPANDED Timespan=2017-2020

The query searched all types of documents published from 2017 to May 2020 with the term *virus* in their titles or terms related to SARS-CoV, coronavirus, or COVID-19 in their title, abstract or keywords. The terms used in this query were collected in the MeSH index of the US National Library of Medicine [10], which is a controlled and hierarchically organized vocabulary used for indexing articles for PubMed [11]. We used only the Science Citation Index Expanded (SCI-EXPANDED) to identify respondents with publications in science journals. We then restricted the results to articles published in research areas related to virology, infectious diseases, immunology, experimental medicine, pharmacology, pharmacy, and general internal medicine.

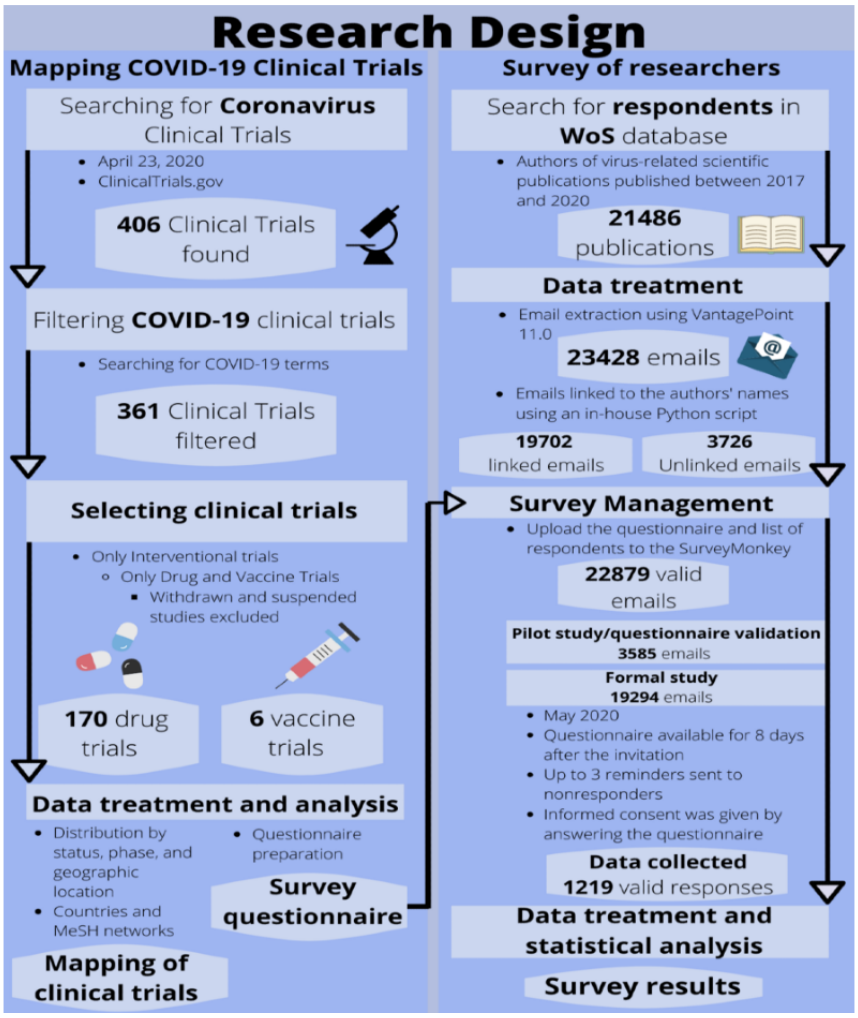
The search returned 21,486 records of publications, which were imported into the data/text mining software VantagePoint 11.0 to generate a comma-separated values (CSV) file of the authors' information (name, email, and publication title). From the information gathered, 23,428 unique emails were identified. Using an in-house Python script, we then linked approximately 84% of these emails to their account owners, which enabled personalized emails to be sent to most of the respondents. All 23,428 emails were uploaded to SurveyMonkey, the web-based survey platform that was used to build the questionnaire and to manage and conduct the web-based survey. After uploading, the number of emails was reduced to 22,879 due to previously opted-out contacts in SurveyMonkey and bounced emails.

Before the formal survey, the questionnaire was validated in a pilot study with a random sample of 10% of the 3585 emails (after uploading to SurveyMonkey) that were not linked to their account owners (unlinked emails). It is known that personalized invitation emails positively influence the response rates of web-based surveys [12]. Therefore, due to the need to obtain a substantial number of responses to be able to report statistically meaningful results, we opted to use only unlinked emails in the pilot. The results and feedback from some of the 20 respondents who took part in this pilot were used to make some changes in the original questionnaire; therefore, the data collected in the pilot were not included in the survey's results.

The pilot and the formal survey were conducted in May 2020. The invitation and reminder emails provided information about the study that was similar to that in the first part of the questionnaire itself (described above). The web-based questionnaire was available for completion on SurveyMonkey for 8 days after the email invitation was sent, and up to three reminder emails were sent to nonresponders. Figure 1 summarizes the research design.

XSL•FO
RenderX

Figure 1. Research design used in the study. MeSH: Medical Subject Headings; WoS: Web of Science.



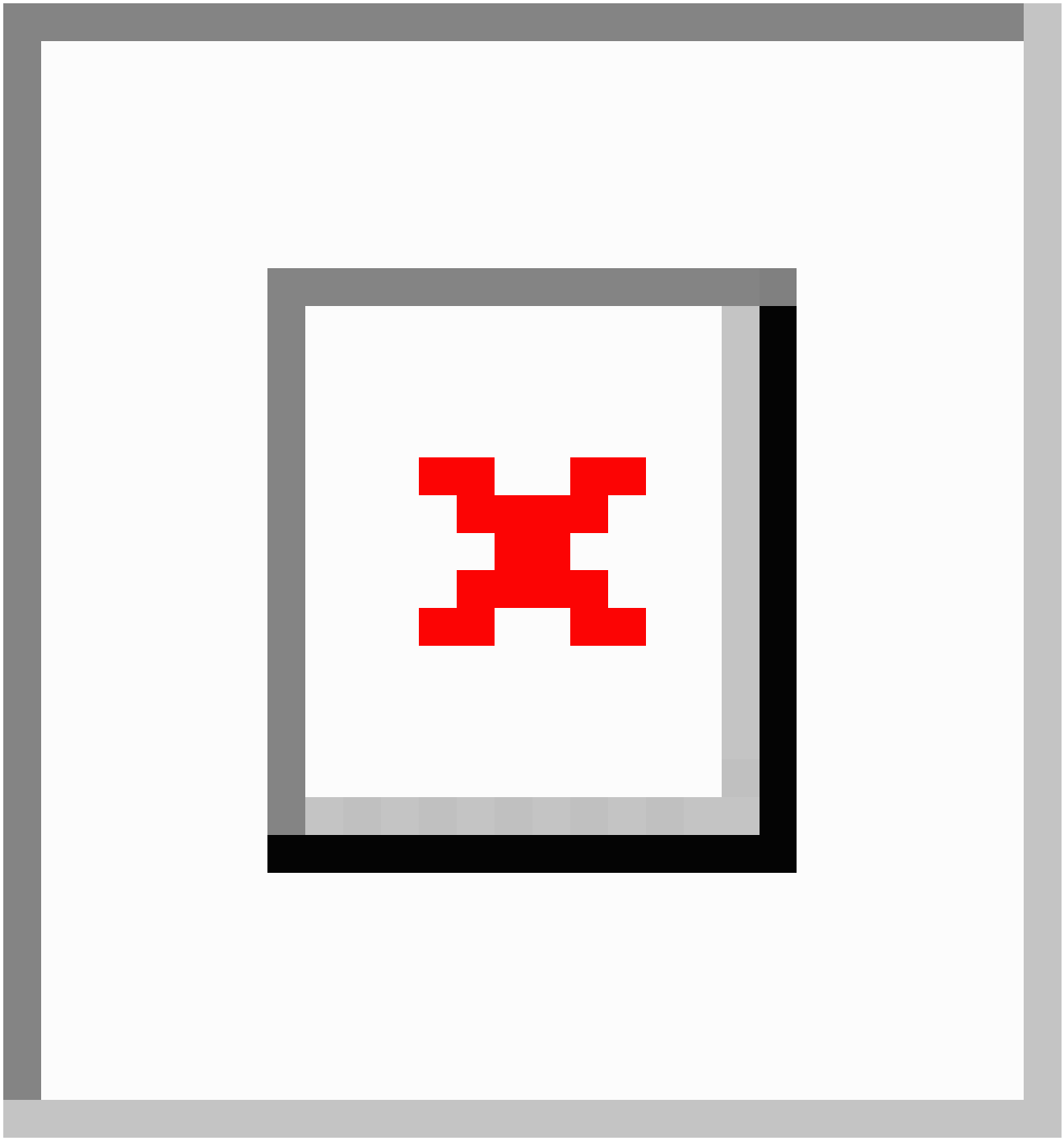
Results

A total of 1275 researchers agreed to participate in this study; however, 45 (3.5%) reported having no knowledge of the prevention or treatment of coronavirus-related human diseases and were disqualified from the survey, leaving 1230 valid responses. Of these respondents, 59.8% (735/1230) reported having good knowledge of the subject, and 40.2% (495/1230) said they had some knowledge. The sample of completed questionnaires (1050/1230, 85.4%) can be considered representative, with a 95% confidence level and a 3% margin of error. To ensure clarity in the graphic representation and description of the results, we combined the answers of these two groups of respondents. Statistically significant differences between them are reported in the descriptions of the results when pertinent. As the questions on therapeutics and vaccine candidates were mandatory in the survey, we gave the respondents an "I prefer not to answer" option. However, because these answers may reflect neutral opinions, we chose not to include them in the analysis of the results. The results by the respondents' level of knowledge, including for the answer "I prefer not to answer," are available in Multimedia Appendix 2.

Figure 2 shows the demographic characteristics of the respondents. Most of them (913/1050, 87.0%) held a doctoral degree, and 9.3% (98/1050) held a master's degree. The majority of the respondents were professors or researchers (764/1059, 72.1%), followed by physicians or clinicians (88/1059, 8.3%) and public health or health care professionals (78/1059, 7.4%). Of the respondents, 74% (780/1054) worked at universities or research organizations, 12.0% (126/1054) in government, and 9.9% (104/1054) at hospitals or similar organizations. A few respondents worked in industry (44/1054, 4.2%). Approximately 50% (493/1061, 46,5%) had more than 20 years of experience in their fields of expertise, while slightly over 30% (334/1061, 31,5%) had 10 to 20 years of practice. The respondents were similarly distributed across Europe (291/1060, 27.5%), North America (252/1060, 23.8%), Asia (224/1060, 21.1%), and South America (183/1060, 17.3%). Much smaller numbers of respondents came from Africa and Australasia and the Pacific Islands: 7.2% (76/1060) and 3.2% (34/1060), respectively.

XSL•FO
RenderX

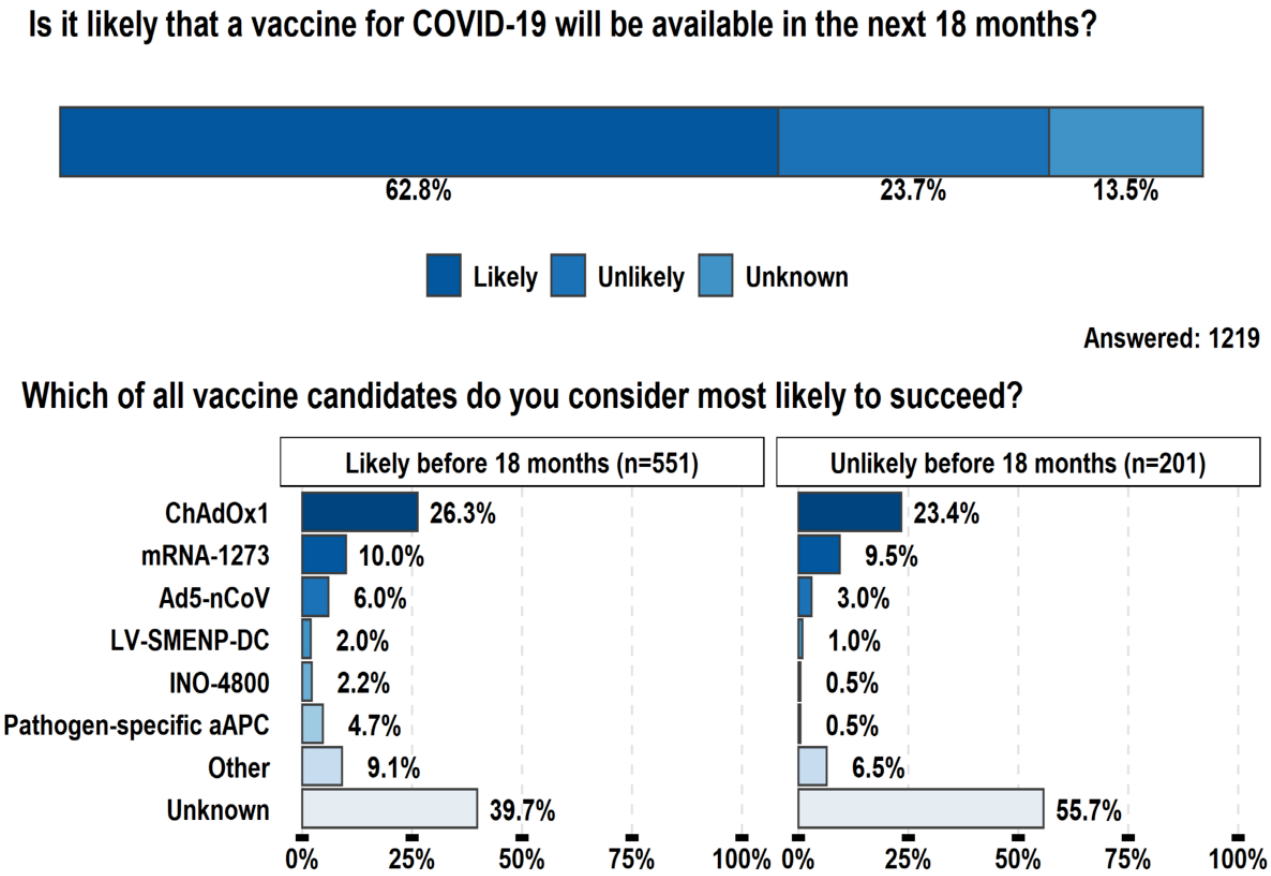
Figure 2. Demographic overview of the survey respondents.



The respondents were asked to share their opinion on whether a vaccine for COVID-19 would be available in the following 18 months. Considering that the questionnaire was administered in May 2020, this deadline was October 2021. They were given three choices: the first two were "likely" and "unlikely," which led them to a question about vaccine candidates, and the third was "unknown," which led them directly to the questions related to the treatment of COVID-19 and skipped the questions about vaccine candidates. Figure 3 shows the results. Almost two-thirds of the respondents (766/1219, 62.8%) believed that vaccines for COVID-19 were likely to be available in the next 18 months. Slightly fewer than 25% (289/1219, 23.7%) believed that a vaccine was feasible, but probably not within 18 months. A small minority (164/1219, 13.5%) did not know whether a vaccine would be available within 18 months. As we now know, several countries started administering vaccinations in December 2020, almost one year ahead of our suggested deadline.

XSL•FO
RenderX

Figure 3. Survey respondents' expectations regarding whether a vaccine for COVID-19 will be available in the next 18 months, and their choices of the most promising vaccine candidates.



On April 23, 2020, there were 6 vaccine candidates for COVID-19 in phase 1 or phase 2 clinical trials according to data collected from ClinicalTrials.gov. Accordingly, these were the only candidates offered as options for the respondents (Figure 3). The respondents who believed the development of a successful vaccine was likely, whether in the next 18 months or not, believed ChAdOx1, from the University of Oxford, to be the most promising candidate, followed by the mRNA-1273 vaccine, from Moderna. Breaking down the responses into the two groups, ChAdOx1 and mRNA-1273 were the top candidates, respectively, for 26.3% (n=145) and 10.0% (n=55) of the 551 respondents who believed that a vaccine would be available within 18 months, and 23.4% (n=47) and 9.5% (n=19) of the 201 respondents who believed that more time was needed. The respondents with self-reported good knowledge of the subject, irrespective of the time they thought the vaccine would take to be developed, had a significantly higher preference for ChAdOx1 than those with some knowledge (P=.05). Figure 3 also shows that many of the respondents reported that it was still unknown which vaccine candidate was the most likely to succeed—39.7% (219/551) of respondents who believed a vaccine was likely to be available before 18 months and 55.7% (112/201) of respondents who believed a vaccine was unlikely to be available before 18 months. A much smaller number considered that none of the 6 vaccine candidates were the most promising—9.1% (50/551) of respondents who believed a vaccine was likely to be available before 18 months and 6.5% (13/201) of respondents who believed a vaccine was unlikely to be available before 18 months.

A great number of therapeutic options for COVID-19 are currently in clinical trials; however, to offer the respondents a short list of options, we included only those that had at least two registered clinical trials by April 23, 2020. We classified these therapeutic options into eight groups of broad treatment types [5] and asked the respondents which group would be the most likely to succeed as a COVID-19 treatment in the next 6 months. The selected group led to a particular question containing several treatment options, of which the respondents were asked to select the most promising. As shown in Figure 4, antivirals and antiretrovirals were believed to be by far the most promising group, selected by 43.4% (437/1008) of the respondents who answered this question. These were followed by monoclonal antibodies (157/1008, 15.6%) and nondrug technologies (118/1008, 11.7%). All the other groups were chosen by fewer than 10.0% of the respondents, with kinase inhibitors and antibiotics being chosen by less than 1% (8/1008, 0.8%). Approximately 4.0% (38/1008) of respondents said they believed there were other broad treatment approaches that were more likely to succeed than the eight listed in the questionnaire, and 9.42% (95/1008) expressed the view that the answer to this question is still unknown. In five of the questions—when the respondents were asked to specify the vaccine, the broad treatment, the antiviral/antiretroviral, the anti-inflammatory, and the nondrug technology—a substantially higher number of

XSL•FO
RenderX

respondents with some knowledge of the subject marked "unknown" than those with good knowledge.

Considering all the eight groups together, only 9 of the 36 possible treatment options were chosen by more than 2.5% of the respondents. Remdesivir, from Gilead Sciences, was by far the top choice, being selected by 22.4% (225/1008) of the total respondents. Remdesivir was also the only treatment for which there was any statistically significant difference (P=.05) between the responses given by the two groups of respondents (with good knowledge and some degree of knowledge on the subject). The respondents with good knowledge were almost twice as likely to select remdesivir than the ones with some knowledge.

Although the group of monoclonal antibodies ranked second, it attracted a high percentage of "unknown" (79/1008, 7.8%) and "other" (31/1008, 3.1%) responses. Of the four monoclonal antibodies in this group, tocilizumab (35/1008, 3.5%) was the preferred option. In terms of individual treatments, the top two choices after remdesivir were both nondrug technologies: immunoglobulin from cured patients (64/1008, 6.3%) and plasma (38/1008, 3.8%). The other treatments selected by at least 2.5% of the respondents are interferons (28/1008, 2.8%), hydroxychloroquine (26/1008, 2.6%), favipiravir (25/1008, 2.5%), lopinavir-ritonavir (24/1008, 2.4%), and recombinant human angiotensin-converting enzyme 2 (rhACE2) (20/1008, 2%).

Figure 4. Therapeutics chosen by the respondents to be most likely to succeed in treating COVID-19. § Umifenovir (0.69%); ¢ Unknown antimalarial (0.20%); ª Mefloquine (0.10%); + Chloroquine (0.10%); $ Other antibiotic (0.10%); * Losartan (0.10%); º Captopril (0.10%); @ Bevacizumab (0.20%); £ Acalabrutinib (0.10%); # Other kinase inhibitor (0.10%); ¬ Other nondrug (0.10%).

XSL•FO
RenderX

Discussion

Principal Findings

Although experimental therapies for COVID-19 are listed in the WHO's Research and Development (R&D) Blueprint of COVID-19 experimental treatments [5] and some other recent scientific papers [6], the absence of an effective specific treatment has led to dissent among the stakeholders invested in finding a cure for COVID-19. Diverging views are found beyond medical and biomedical scientists, being found in a variety of health professionals, managers, policy makers, etc. Although this lack of consensus regarding treatment for COVID-19 was also observed among the virus-related researchers who participated in this study, a great number of them (437/1008, 43.4%) stated a preference for antivirals and antiretrovirals over the seven other broad treatment groups addressed in this study.

In the antivirals and antiretrovirals group, remdesivir was the drug that most respondents believed to be the most promising for treating patients with COVID-19. Remdesivir is a nucleoside analog drug with extensive antiviral activity, and it is effective in the treatment of the lethal Ebola and Nipah viruses [13]. Its use in patients with COVID-19 started after a high-profile publication reported on its compassionate use in a 53-patient case series [14]. This publication showed clinical improvement in 36 out of 53 patients (68%); however, it was not randomized, so other evidence was needed [6]. Another published study reported on the use of remdesivir in 236 patients in 10 hospitals in Wuhan, China, in a double-blind, placebo-controlled, multicenter randomized trial. Remdesivir was not associated with a difference in the primary outcome of time to clinical improvement. This study was stopped early because no further patients meeting the eligibility criteria were admitted to hospital in Wuhan after some weeks [15].

On April 23, 2020, five clinical trials evaluating the use of remdesivir in patients with COVID-19 were registered at ClinicalTrials.gov. Four months later, this number had increased to >40. The most recent systematic review and network meta-analysis of drug treatments for COVID-19 shows, with moderate certainty, the benefits of the use of remdesivir both for the resolution of symptoms and the duration of mechanical ventilation. However, there are still uncertainties about its effect on mortality and other important outcomes [16].

Additionally in the group of antivirals and antiretrovirals, the respondents considered interferons, favipiravir, and lopinavir-ritonavir to be promising treatments for COVID-19. Interferons are often evaluated as candidates for the treatment of emerging viral infections before specific treatments are developed because of their unspecific antiviral effects [17]. They are usually prescribed in combination with other drugs, and this was indeed the case in most of the clinical trials involving interferons that were registered by April 23, 2020. Several types of interferons are available for antiviral treatment of patients with COVID-19, including alpha [16], beta [18], and lambda [19] interferons. As for favipiravir, limited clinical experience has been reported to date [6]; however, recent experimental treatments with favipiravir in combination with interferon alfa reported significant improvements in chest imaging [20]. In the case of lopinavir-ritonavir, a recent evaluation showed no observed benefit in 99 hospitalized adult patients [21].

Nondrugs are also considered to be promising for the treatment of patients with COVID-19. Immunotherapy with IgG can increase immune response in newly infected patients, reducing the disease burden [22]. Plasma collected from recovered patients functions similarly. In a recent experiment, 5 severely ill patients in China showed improvement after receiving convalescent plasma [23]. However, despite its potential usefulness, the absence of large-scale trials and more rigorous investigations of convalescent plasma is an obstacle to its more widespread use in patients with COVID-19 [24].

Another potential drug available to treat COVID-19 is a member of the monoclonal antibody group. Tocilizumab is a humanized monoclonal antibody against the interleukin-6 receptor, which has been reported to be one of the most important cytokines involved in COVID-19 [25]. A recent publication showed clinical improvement of respiratory function in 91% of patients who were prescribed tocilizumab [26]. However, in the absence of other studies, its use should be approached with caution [6].

On April 23, 2020, hydroxychloroquine was the drug being tested in the highest number of clinical trials. It was considered to be a promising treatment for COVID-19 by 2.6% of the respondents. Hydroxychloroquine is commonly used as a treatment for malaria, systemic lupus erythematosus, and rheumatoid arthritis [6]. Despite initial reports of its success [27], recent scientific publications have not shown any significant reduction in mortality, and some have expressed caution because of its potentially harmful effects [28,29].

In the angiotensin-converting enzyme (ACE) inhibitors group, rhACE2 was another option selected by at least 2% of the respondents. The rationale for its use comes from the assumption that the SARS-CoV-2 virus uses angiotensin-converting enzyme 2 (ACE2) receptors to enter host cells [30]. Hence, exogenously supplementing rhACE2 may block the interaction of the virus with its cellular receptor [31]. However, further clinical studies are needed [30].

Although it was selected only by 0.7% of respondents, dexamethasone is the drug that has probably showed the greatest efficacy in reducing the COVID-19 burden, especially in critically ill patients. After the UK-based Randomised Evaluation of COVID-19 Therapy (RECOVERY) trial reported benefits of the use of dexamethasone [32], other trials and a recent meta-analysis showed that the use of dexamethasone and other systemic corticosteroids is associated with a 28-day reduction in patient mortality [33].

Regardless of the importance of pharmacological treatment to reduce the short- and medium-term burden of COVID-19, there is a consensus in the scientific community that large-scale vaccination will be the most effective long-term strategy to end the pandemic [7,34]. It usually takes years to develop a new vaccine; however, the need for rapid development has required a completely different paradigm [8]. Yet, even considering the amount of progress made in the few months after our survey, in September 2020, most vaccine candidates were still in the

XSL•FO
RenderX

preclinical phase, and dozens were in phase 1 or 2. Some vaccines were in phase 3. This is the case, for example, for the vaccine candidates from BioNTech/Pfizer and from the Gamaleya Research Institute, which showed good results in terms of safety and immunogenicity [35,36]. In December 2020, the BioNTech/Pfizer vaccine was the first to be made available worldwide after being approved by regulatory bodies in the United States, Europe, Canada, and the United Kingdom. All these candidates highlight the diversity of vaccine development technology platforms, including recent advances not yet used in any licensed viral vaccine, such as RNA- and DNA-based technologies [7].

Considered the most promising vaccine candidate by the respondents of this survey, ChAdOx1, from the University of Oxford, uses a more traditional nonreplicating viral vector technological platform. This type of platform offers long-term stability, induces a strong immune response, and offers high levels of protein expression. Because some vaccines are already using this platform for other diseases, there is also the advantage of potentially using the existing industrial capacity for its production [7]. A preliminary report of a phase 1/2 single-blind randomized controlled trial of ChAdOx1 showed an acceptable safety profile and homologous increase in antibody responses [37]. In January 2020, this vaccine was also approved for wide use in the United Kingdom and Mexico.

A nonreplicating viral vector platform is also being used for the Ad5-nCoV vaccine candidate, from CanSino Biologics, which was the third most highly ranked vaccine candidate in the study. Ad5-nCoV has already been indicated to be tolerable and immunogenic at 28 days postvaccination [38], and it is currently being used in the Chinese military [39]. The second most selected vaccine, mRNA-1273, developed by Moderna, uses innovative messenger RNA (mRNA) technology and was also approved for emergency use in Israel, the United States, and Canada. Because this virus uses synthetic processes and does not require culture or fermentation, it may be easier for Moderna to produce a large number of doses with its RNA-based technology [8,34].

Final Remarks

This study presents the results of a global survey of over 1000 virus specialists with knowledge on the prevention and treatment of coronavirus-related human diseases that took place in May 2020. A great number of the respondents believed that antiviral and antiretroviral therapies were the most likely treatments to succeed in the fight against COVID-19. Of the 36 options, grouped into eight broad treatment types, remdesivir and immunoglobulin from cured patients were regarded as the most likely to prove effective. Although remdesivir is currently being used in clinical practice with guidelines [40], the respondents did not expect that dexamethasone and other systemic steroids could have an important role in treatment [33].

As mentioned, the race for a COVID-19 vaccine has given rise to a large number of other vaccine candidates, with some already showing good results in terms of safety and immunogenicity and others already being used in a few countries. Of the vaccine candidates presented to the respondents, ChAdOx1 and mRNA-1273 were considered the two most promising options for COVID-19 prevention. Also, most respondents indicated that it was likely that a vaccine would be available within 18 months. This seemed very optimistic, given that a new vaccine usually takes 10 years to develop. Even the fast-tracked development of the first Ebola vaccine required 5 years [7]. However, changes in the clinical trial process were implemented worldwide in an attempt to accelerate the discovery of an entirely new vaccine to prevent COVID-19 [8]. These changes were already taking place in May 2020; this may help us understand why the respondents felt it would take such a short time to develop a vaccine, which is exactly what happened a few months later.

Due to the tremendous scientific efforts to end the COVID-19 pandemic, much has changed since we conducted this survey in May 2020. As a large amount of new information on COVID-19 is published daily, the issues addressed in this study are constantly evolving; therefore, the current state of knowledge is different from that in May 2020. Yet, comparing what was known back at the beginning of the fight against COVID-19 with what is known today, many of the results reported in this study are in line with what is currently happening. Thus, if the future confirms the expectations of the researchers who participated in this study, the discovery that will bring an end to the COVID-19 pandemic will not be very far away.

Acknowledgments

The authors thank the Center for Strategic Studies of Oswaldo Cruz Foundation for institutional support and all the respondents of the survey. We thank the two anonymous reviewers and the JMIR editor, Dr Rita Kukafka, for contributing to the improvement of this article. We also thank Dr Claúdio Damasceno for support in the initial phase of this study and Dr Nicolle Menezes for comments and suggestions that contributed to the improvement of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical trials related to COVID-19.
[DOCX File , 1380 KB - publichealth_v7i2e22483_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Statistics for the responses to the survey.
[PDF File (Adobe PDF File), 445 KB - publichealth_v7i2e22483_app2.pdf]

References

1. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). J Gen Intern Med 2020 May 4;35(5):1545-1549 [FREE Full text] [doi: 10.1007/s11606-020-05762-w] [Medline: 32133578]
2. WHO timeline - COVID-19. World Health Organization. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-05-23]
3. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-05-23]
4. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. World Health Organization. 2020 Mar 13. URL: https://www.who.int/docs/default-source/coronaviruse/clinical-management-of-novel-cov.pdf [accessed 2021-02-19]
5. COVID 19 landscape of experimental treatments. World Health Organization. URL: https://www.who.int/publications/i/item/covid-19-landscape-of-experimental-treatments [accessed 2021-02-19]
6. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): A review. JAMA 2020 May 12;323(18):1824-1836. [doi: 10.1001/jama.2020.6019] [Medline: 32282022]
7. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 May 9;19(5):305-306. [doi: 10.1038/d41573-020-00073-5] [Medline: 32273591]
8. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 vaccines at pandemic speed. N Engl J Med 2020 May 21;382(21):1969-1973. [doi: 10.1056/nejmp2005630]
9. Improving public access to aggregate content of ClinicalTrials.gov. Clinical Trials Transformation Initiative. URL: https://aact.ctti-clinicaltrials.org/ [accessed 2021-02-19]
10. MeSH. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/mesh [accessed 2021-02-19]
11. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2021-02-19]
12. Fan W, Yan Z. Factors affecting response rates of the web survey: A systematic review. Comput Hum Behav 2010 Mar;26(2):132-139. [doi: 10.1016/j.chb.2009.10.015]
13. Cao Y, Deng Q, Dai S. Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: an evaluation of the evidence. Travel Med Infect Dis 2020 Apr;35:101647 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101647] [Medline: 32247927]
14. Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate use of remdesivir for patients with severe Covid-19. N Engl J Med 2020 Jun 11;382(24):2327-2336. [doi: 10.1056/nejmoa2007016]
15. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet 2020 May;395(10236):1569-1578. [doi: 10.1016/s0140-6736(20)31022-9]
16. Gao L, Yu S, Chen Q, Duan Z, Zhou J, Mao C, et al. A randomized controlled trial of low-dose recombinant human interferons alpha-2b nasal spray to prevent acute viral respiratory infections in military recruits. Vaccine 2010 Jun 17;28(28):4445-4451 [FREE Full text] [doi: 10.1016/j.vaccine.2010.03.062] [Medline: 20394720]
17. Sallard E, Lescure F, Yazdanpanah Y, Mentre F, Peiffer-Smadja N. Type 1 interferons as a potential treatment against COVID-19. Antiviral Res 2020 Jun;178:104791 [FREE Full text] [doi: 10.1016/j.antiviral.2020.104791] [Medline: 32275914]
18. Shalhoub S. Interferon beta-1b for COVID-19. Lancet 2020 May;395(10238):1670-1671. [doi: 10.1016/s0140-6736(20)31101-6]
19. Prokunina-Olsson L, Alphonse N, Dickenson R, Durbin JE, Glenn JS, Hartmann R, et al. COVID-19 and emerging viral infections: the case for interferon lambda. J Exp Med 2020 May 04;217(5) [FREE Full text] [doi: 10.1084/jem.20200653] [Medline: 32289152]
20. Cai Q, Yang M, Liu D, Chen J, Shu D, Xia J, et al. Experimental treatment with favipiravir for COVID-19: an open-label control study. Engineering (Beijing) 2020 Oct;6(10):1192-1198 [FREE Full text] [doi: 10.1016/j.eng.2020.03.007] [Medline: 32346491]
21. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. N Engl J Med 2020 May 07;382(19):1787-1799. [doi: 10.1056/nejmoa2001282]
22. Jawhara S. Could intravenous immunoglobulin collected from recovered coronavirus patients protect against COVID-19 and strengthen the immune system of new patients? Int J Mol Sci 2020 Mar 25;21(7):2272 [FREE Full text] [doi: 10.3390/ijms21072272] [Medline: 32218340]
23. Shen C, Wang Z, Zhao F, Yang Y, Li J, Yuan J, et al. Treatment of 5 critically ill patients with COVID-19 with convalescent plasma. JAMA 2020 Apr 28;323(16):1582-1589 [FREE Full text] [doi: 10.1001/jama.2020.4783] [Medline: 32219428]
24. Roback JD, Guarner J. Convalescent plasma to treat COVID-19: possibilities and challenges. JAMA 2020 Apr 28;323(16):1561-1562. [doi: 10.1001/jama.2020.4940] [Medline: 32219429]

XSL•FO
RenderX

25. Luo P, Liu Y, Qiu L, Liu X, Liu D, Li J. Tocilizumab treatment in COVID-19: a single center experience. J Med Virol 2020 Jul 15;92(7):814-818 [FREE Full text] [doi: 10.1002/jmv.25801] [Medline: 32253759]
26. Xu X, Han M, Li T, Sun W, Wang D, Fu B, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A 2020 May 19;117(20):10970-10975 [FREE Full text] [doi: 10.1073/pnas.2005615117] [Medline: 32350134]
27. Gautret P, Lagier J, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrobial Agents 2020 Jul;56(1):105949 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105949]
28. Bonow RO, Hernandez AF, Turakhia M. Hydroxychloroquine, coronavirus disease 2019, and QT prolongation. JAMA Cardiol 2020 Sep 01;5(9):986-987. [doi: 10.1001/jamacardio.2020.1782] [Medline: 32936259]
29. Ferner RE, Aronson JK. Chloroquine and hydroxychloroquine in covid-19. BMJ 2020 Apr 08;369:m1432. [doi: 10.1136/bmj.m1432] [Medline: 32269046]
30. Roshanravan N, Ghaffari S, Hedayati M. Angiotensin converting enzyme-2 as therapeutic target in COVID-19. Diabetes Metab Syndr 2020 Jul;14(4):637-639 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.022] [Medline: 32428864]
31. Guo J, Huang Z, Lin L, Lv J. Coronavirus disease 2019 (COVID‐19) and cardiovascular disease: a viewpoint on the potential influence of angiotensin‐converting enzyme inhibitors/angiotensin receptor blockers on onset and severity of severe acute respiratory syndrome coronavirus 2 infection. JAHA 2020 Apr 09;9(7):A. [doi: 10.1161/jaha.120.016219]
32. RECOVERY Collaborative Group T. Dexamethasone in hospitalized patients with Covid-19 — preliminary report. N Engl J Med 2020 Jul 17. [doi: 10.1056/nejmoa2021436]
33. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, et al. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. JAMA 2020 Oct 06;324(13):1330-1341. [doi: 10.1001/jama.2020.17023] [Medline: 32876694]
34. Cohen J. Vaccine designers take first shots at COVID-19. Science 2020 Apr 03;368(6486):14-16. [doi: 10.1126/science.368.6486.14] [Medline: 32241928]
35. Logunov D, Dolzhikova I, Zubkova O, Tukhvatulin A, Shcheblyakov D, Dzharullaeva A, et al. Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. Lancet 2020 Sep;396(10255):887-897. [doi: 10.1016/S0140-6736(20)31866-3]
36. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. Nature 2020 Oct 12;586(7830):589-593. [doi: 10.1038/s41586-020-2639-4] [Medline: 32785213]
37. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, Oxford COVID Vaccine Trial Group. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet 2020 Aug 15;396(10249):467-478 [FREE Full text] [doi: 10.1016/S0140-6736(20)31604-4] [Medline: 32702298]
38. Zhu F, Li Y, Guan X, Hou L, Wang W, Li J, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet 2020 Jun;395(10240):1845-1854. [doi: 10.1016/S0140-6736(20)31208-3]
39. Mahase E. Covid-19: Where are we on immunity and vaccines? BMJ 2020 Aug 05;370:m3096. [doi: 10.1136/bmj.m3096] [Medline: 32759299]
40. Rochwerg B, Agarwal A, Zeng L, Leo YS, Appiah JA, Agoritsas T, et al. Remdesivir for severe covid-19: a clinical practice guideline. BMJ 2020 Jul 30;370:m2924. [doi: 10.1136/bmj.m2924] [Medline: 32732352]

Abbreviations

AACT: Aggregate Analysis of ClinicalTrials.gov
ACE: angiotensin-converting enzyme
ACE2: angiotensin-converting enzyme 2
CSV: comma-separated values
MeSH: Medical Subject Headings
R&D: research and development
RECOVERY: Randomised Evaluation of COVID-19 Therapy
SARS-CoV: severe acute respiratory syndrome coronavirus
SCI-EXPANDED: Science Citation Index Expanded
WHO: World Health Organization
WoS: Web of Science

XSL•FO
RenderX

Edited by G Eysenbach, R Kukafka; submitted 13.07.20; peer-reviewed by JM Chen, R Poluru; comments to author 29.08.20; revised version received 06.09.20; accepted 08.12.20; published 26.02.21.

Please cite as:
Cabral BP, Braga L, Mota F
Expert Opinions on the Most Promising Treatments and Vaccine Candidates for COVID-19: Global Cross-sectional Survey of Virus Researchers in the Early Months of the Pandemic
JMIR Public Health Surveill 2021;7(2):e22483
URL: https://publichealth.jmir.org/2021/2/e22483
doi:10.2196/22483
PMID:33635275

©Bernardo Pereira Cabral, Luiza Braga, Fabio Mota. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 26.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Methods for Social Media Monitoring Related to Vaccination: Systematic Scoping Review

Emilie Karafillakis[1*], MSc; Sam Martin[1*], PhD; Clarissa Simas[1], MSc; Kate Olsson[2], MPH; Judit Takacs[2,3], PhD; Sara Dada[1], MSc; Heidi Jane Larson[1,4], PhD

[1]London School of Hygiene & Tropical Medicine, Vaccine Confidence Project, London, United Kingdom

[2]European Centre for Disease Prevention and Control, Stockhom, Sweden

[3]Centre for Social Sciences, Hungarian Academy of Sciences, Budapest, Hungary

[4]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, United States

[*]these authors contributed equally

Corresponding Author:
Emilie Karafillakis, MSc
London School of Hygiene & Tropical Medicine
Vaccine Confidence Project
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 2079272831 ext 2831
Email: emilie.karafillakis@lshtm.ac.uk

Abstract

Background: Social media has changed the communication landscape, exposing individuals to an ever-growing amount of information while also allowing them to create and share content. Although vaccine skepticism is not new, social media has amplified public concerns and facilitated their spread globally. Multiple studies have been conducted to monitor vaccination discussions on social media. However, there is currently insufficient evidence on the best methods to perform social media monitoring.

Objective: The aim of this study was to identify the methods most commonly used for monitoring vaccination-related topics on different social media platforms, along with their effectiveness and limitations.

Methods: A systematic scoping review was conducted by applying a comprehensive search strategy to multiple databases in December 2018. The articles' titles, abstracts, and full texts were screened by two reviewers using inclusion and exclusion criteria. After data extraction, a descriptive analysis was performed to summarize the methods used to monitor and analyze social media, including data extraction tools; ethical considerations; search strategies; periods monitored; geolocalization of content; and sentiments, content, and reach analyses.

Results: This review identified 86 articles on social media monitoring of vaccination, most of which were published after 2015. Although 35 out of the 86 studies used manual browser search tools to collect data from social media, this was time-consuming and only allowed for the analysis of small samples compared to social media application program interfaces or automated monitoring tools. Although simple search strategies were considered less precise, only 10 out of the 86 studies used comprehensive lists of keywords (eg, with hashtags or words related to specific events or concerns). Partly due to privacy settings, geolocalization of data was extremely difficult to obtain, limiting the possibility of performing country-specific analyses. Finally, 20 out of the 86 studies performed trend or content analyses, whereas most of the studies (70%, 60/86) analyzed sentiments toward vaccination. Automated sentiment analyses, performed using leverage, supervised machine learning, or automated software, were fast and provided strong and accurate results. Most studies focused on negative (n=33) and positive (n=31) sentiments toward vaccination, and may have failed to capture the nuances and complexity of emotions around vaccination. Finally, 49 out of the 86 studies determined the reach of social media posts by looking at numbers of followers and engagement (eg, retweets, shares, likes).

Conclusions: Social media monitoring still constitutes a new means to research and understand public sentiments around vaccination. A wide range of methods are currently used by researchers. Future research should focus on evaluating these methods to offer more evidence and support the development of social media monitoring as a valuable research design.

XSL•FO
RenderX

(JMIR Public Health Surveill 2021;7(2):e17149) doi:10.2196/17149

KEYWORDS

vaccination; antivaccination movement; vaccination refusal; social media; internet; research design; review; media monitoring; social listening; infodemiology; infoveillance

Introduction

Although public questioning of vaccination is as old as vaccination itself [1], continuous advancements in the global communication landscape, associated with the rise of social media as an interactive health information ecosystem, have contributed to the unmediated spread of vaccine hesitancy [2]. This new boundless information ecosystem has shaped the nature of conversations about vaccination, with evidence showing that social media can facilitate the quick diffusion of negative sentiments and misinformation about vaccination [2-6]. Furthermore, individuals have been found to more commonly engage with negative information around vaccination than positive content [7-10]. In this context, public trust in information provided by authorities and experts can decrease [11-13], influencing vaccine decisions [14]. Recent evidence has shown that social media users tend to cluster and create so-called "echo chambers" based on their views toward vaccination [15]; however, Leask et al [16] highlight that "a patient's trust in the source of information may be more important than what is in the information," stressing the importance of reaching individuals, across all clusters, through trustworthy sources.

Social media monitoring (infoveillance) provides opportunities to listen, in real time, to online narratives about vaccines, and to detect changes in sentiments and confidence early [17]. Information gathered from social media monitoring is crucial to inform the development of targeted and audience-focused communication strategies to maintain or rebuild trust in vaccination [17,18]. However, as social media monitoring can be resource- and time-intensive, and can raise issues of confidentiality, transparency, and privacy [17,19,20], evidence of public health communities investing in such listening mechanisms remains sparse.

The aim of this scoping review was to systematically summarize the methodologies that have been used to monitor and analyze social media on vaccination using an innovative three-step model. The findings presented in this paper come from a broader European Centre for Disease Prevention and Control (ECDC) technical report [21]. The aim of the ECDC report was to provide guidance for public health agencies to monitor and engage with social media, whereas this paper primarily focuses on the academic implications of social media monitoring. The specific objectives of this scoping review were to (1) identify the methods most commonly used for monitoring different social media platforms; and (2) identify the extent to which methods have been evaluated, along with their effectiveness and limitations.

Methods

Design

Systematic scoping reviews are used to map international literature with the aim of clarifying "working definitions and conceptual boundaries of a topic or field" [22] as well as identifying how research is conducted [22-24]. Systematic scoping reviews focus on scoping larger, more complex, and heterogeneous topics than systematic literature reviews. A systematic scoping review approach was therefore adapted to fulfill the goal of summarizing study methodologies used to monitor social media content around vaccination. The methodology for this scoping review was based on the work of Arksey et al [23] and Peters et al [24].

Framing Social Media

Kaplan et al [25] define social media as "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content." They further classify social media into blogs, collaborative projects (eg, Wikipedia), social networking sites (eg, Facebook), content communities (eg, YouTube), virtual social worlds (eg, Second Life), and virtual game worlds (eg, World of Warcraft) [25].

However, social media is not merely an information tool but also represents a continuously evolving social environment directly influenced by how individuals produce and share content, and interact with each other. For the purpose of this scoping review, we consider social media as not simply a means of communication but further a space within which individuals socialize and organize. This review therefore focuses on social networking sites and content communities, and excludes online platforms that do not have social interactions as their main purpose (eg, blogs or websites with a comments section).

Search Strategy and Screening Process

The search strategy for the scoping review was developed by librarians at ECDC and researchers at the Vaccine Confidence Project (VCP), and was peer-reviewed to balance feasibility and comprehensiveness, including both social media and vaccination-related English keywords (see Multimedia Appendix 1). The search was conducted by one VCP researcher on the EMBASE database, and was adapted to search the PubMed, Scopus, MEDLINE, PsycINFO, PubPsych, Open Grey, and Web of Science databases in December 2018.

Identified articles were exported into Endnote, and duplicates were removed based on ECDC guidelines consisting of 6 rounds of deduplication looking for articles with similar author, year, and title; title, volume, and pages; author, volume, and pages; year, volume, issue, and pages; title; and author and year. The automated deduplication function in Endnote was not used, and articles were compared visually to ensure that only true

XSL•FO
RenderX

duplicates were removed. Two VCP reviewers independently screened articles by title and abstract and by full text using a set of predefined inclusion and exclusion criteria. Disagreements were resolved by discussion.

Articles were included if they described studies performed to monitor or analyze data collected from social media around vaccination. The definition of social media described above was used as one of the inclusion criteria, limiting results to social networking sites and content communities. No restrictions were made with respect to location or language, as a team of official translators was available at the ECDC.

Articles were excluded if they were published before 2000 or if they were not about human vaccines. Articles that monitored online media (eg, news, websites) but did not collect any data from social media were excluded. The following article types were also excluded: conference abstracts, editorials, commentaries, and letters to the editor.

Data Management and Analysis

Two VCP researchers extracted the following data from the included articles: country, aim, study population, period of monitoring, vaccine, social media, media monitoring methodologies (tool for data collection, keywords, exclusion criteria, geolocation), analysis (sentiment coding and analysis, reach, spread and interaction analyses, other types of analyses), results (number of posts), and evaluation and limitations.

To facilitate the description of social media monitoring methods, a three-step model of social media monitoring was developed (Figure 1), including (1) preparation, (2) data extraction, and (3) data analysis steps. The preparation phase consists of defining the purpose of social media monitoring and addressing any ethical considerations. The data extraction phase includes selecting data extraction tools and periods of monitoring, developing comprehensive search strategies, and extracting the data. Finally, the data analysis stage includes geolocation, trends, content, sentiments, and reach analyses. The findings summarized in this paper are organized according to this three-step model.

Three researchers summarized, charted, and analyzed the data. A descriptive analysis was conducted for the types of data collection tools used to gather data from social media, the keywords and search strategies used, and the various analytical methods. These researchers reviewed and compared results in the data extraction sheet, listed and identified the frequency of different methods used for social media monitoring, and identified common themes. Two researchers met to discuss the findings and interpret them together with contextual information, identifying needs for further research.

Figure 1. The three-step model of social media monitoring.

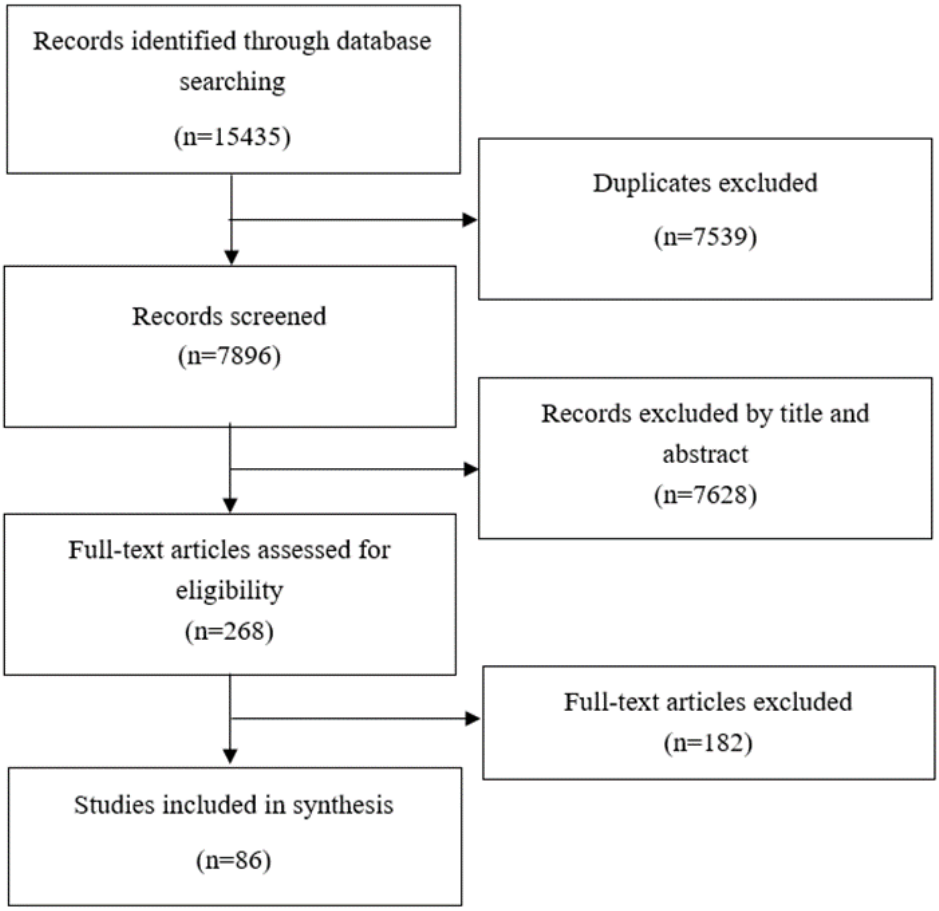
XSL•FO
RenderX

Results

Included Studies

The search strategy generated 15,435 articles, from which 7539 duplicates and 7628 irrelevant articles were excluded after screening by title and abstract (Figure 2). From the 268 articles screened by full text, 182 were excluded for the following reasons: not about social media (n=141); no data provided (n=19); conference abstracts, editorials, or letters to the editor (n=6); article not accessible (even after enquiring multiple libraries and contacting authors) (n=4); article containing data already published in another included article (n=1); and not on vaccination (n=1). At the end of the screening process, 86 articles in English, Spanish, and Italian were included for analysis. Articles in Spanish and Italian were analyzed by a researcher fluent in these two languages.

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



Study Characteristics

The first study identified on social media monitoring around vaccination was published in 2006 [26], with an increasing number of studies published yearly since then. Most studies analyzed online discourse on Twitter (n=42) [8,27-67], YouTube (n=12) [68-79], Facebook (n=11) [80-90], and online forums (eg, babytree, mothering.com, mumsnet, KaksPlus; n=9) [26,91-98]. A diversification of social media platforms can be observed in recent years, with studies of platforms such as Pinterest (n=1) [99], Weibo (n=1) [100], Reddit (n=1) [101], or Yahoo! Answers (n=2) [102,103] all published after 2015. Seven studies monitored a mix of social media platforms [104-110]. More detailed study characteristics are provided in Multimedia Appendix 2.

Social Media Monitoring Methods

A range of methods were used across the 86 included studies to monitor social media, most of which have not been evaluated in terms of their accuracy and reliability. In the following sections, these methods are described based on the three steps of social media monitoring proposed in this paper: preparation, data extraction, and data analysis.

Preparation Phase

Defining the Purpose of Social Media Monitoring

The main objective of the majority of studies in this review (55/86, 64%) was to better understand how vaccination is portrayed on social media, whether through the analysis of online discourse or sentiments, or by looking at how information is produced, shared, and engaged with [8,26,27,30,33,39,42-44,46-51,53-56,58-62,68-80,82-88,90-94,98,99,101,102,105,106].

XSL•FO
RenderX

Many studies (15/86, 17%) used social media monitoring as a way to better understand general public discussions on vaccination, assuming that online discussions are a good proxy for vaccine confidence in a country or region [29,32,34,36-38,45,57,63,65,67,81,95,96,103]. This comes with important limitations due to the lack of representativeness of social media populations. Individuals discussing vaccination on social media tend to come from specific population groups, usually younger or female groups [61]. Furthermore, data extracted from social media platforms are often not representative of the entire online discourse around vaccination on these platforms (see Data Extraction Phase section).

Other objectives included estimating correlations between online activity around vaccination and coverage or outbreak data [40,41,66,104,107,109], describing systems for monitoring vaccination online [28,52,64,97], investigating relations between news media and social media posts [100,108,110], understanding the contribution of bots or trolls to online content about vaccination [31], examining political references to vaccination on social media [35], and detecting anxiety-related adverse events following immunization [89].

Addressing Ethical Considerations

Access to social media data is becoming increasingly restricted, as some users set their profiles, conversations, or pages as "private" [111]. Yet, only 15% (13/86) of the studies included in this review were found to have been reviewed and to have received approval from an institutional ethics review board for their study [32,36-38,50,56,67,81,83,85,86,90,93]. An additional 9 studies sent protocols to institutional or ethical review boards and were deemed exempt because they only analyzed public data and social media users were not considered as "human research subjects" [8,31,44,48,60,69,70,75,80]. One study also explained that guidelines from an institutional review board were considered and applied during the study to protect social media users [92].

However, even once ethical approval had been obtained, questions about anonymity, confidentiality, and informed consent remained. For example, one study explained that anonymization of data is extremely difficult to maintain on social media, as content and quotes (whether from private or public data) can easily be traced back to users, revealing their identity [54]. The authors of another study performed on Facebook in Israel explained that although they anonymized their data, informed consent was not required as "subjects would expect to be observed by strangers" when posting messages on the internet [85]. Some studies also discussed the limitations of focusing on public data and the distorted view this creates [28,43,54,61,83,85,88,97,109].

Data Extraction Phase

Selecting Data Extraction Tools and Periods of Monitoring

Studies included in this review were found to use different monitoring tools to extract data from social media platforms. Thirty-five studies used manual browser search tools such as search bars available on Twitter, YouTube, or Facebook [26,60,68-81,83,85,86,89-100,102,104,106,109]. Due to browser and user interface limitations, studies that used manual browser search functions were time-consuming and collected small amounts of data over short periods of time. In one study, the analysis had to be limited to 30 Facebook pages [80], which affected the possibility of capturing data over different periods (ie, possibly missing trends in the number of posts around influenza seasons). Furthermore, owing to the time needed to assess Facebook pages, it was found to be impractical to analyze each complete page in detail. Another limitation comes from browser cookies and personal tracking algorithms imposed by search engines, which can influence manual search results and their listed popularity. Researchers from another study indicated that findings from both the Google and Facebook searches were dependent on the geographic location of the reviewer's browser settings [89]. Researchers using manual search engines are also restricted to the way search results are presented on different platforms; for example, Pinterest does not list its pins chronologically and does not provide exact time stamps [99]. This made using a more conventional content analysis sampling method (eg, a constructed 2-week time period) virtually impossible in this particular study.

Forty-nine studies used either social media application program interfaces (APIs) (n=24) [29,31-33,35,37-41,43,45,48,54,56-58,65-67,82,84,87,88,101,103], automated monitoring tools (n=20) [27,36,42,44,46,47,49-53,55,59,61,62,64,105,107,108,110], or a combination of both (n=5) [8,28,32,34,63] to extract data from social media platforms. The term API refers to a software intermediary that allows two apps to talk to each other [112]. APIs pull and interpret data from servers storing information for Facebook, Twitter, YouTube, Reddit, and many more platforms. It is important to note that APIs do not provide comprehensive access to all social media content, and often only pull random samples of content; for example, Twitter provides access to roughly 1% of public Tweets through its API [113]. Automatic monitoring tools refer to automated web platforms that access social media data via APIs. These automated tools come with user-friendly interfaces, which can be free (with limited access to a random sample of all posts), open source (open to development from other developers), or commercial (where access to a larger percentage of posts is allowed, which can be real-time or archival via a subscription pricing structure). Regardless of the data collection period, studies with the highest number of results and the most robust datasets consistently came from the use of social media APIs or automatic data sampling. The "Yahoo! Answers" API provided the largest sample size from a single platform over a sampling period of 5 years (16 million messages) [103] and Crimson Hexagon was the automated platform that provided the largest mixed sample size, with a mixture of 58,078 Facebook posts and 82,993 tweets over a 7.5-year period [105]. The Yahoo! Answers study found that the API data were difficult to stratify by age, gender, income, education level, or marital status, which may have limited generalizability [103]. Similarly, Smith et al [57] found that using the Twitter API to use social media discussions as a proxy for the population at large is problematic. The difficulty in finding the correct self-assigned demographic of users, and whether they are real users or automated bots, makes the findings less generalizable. Finally, a large number (n=45) of studies using either APIs or

XSL•FO
RenderX

automated software focused on Twitter due to the ease of access given by the platform to its data stream compared to other platforms, which may give a skewed perspective of social media attitudes toward vaccines [8,27-59,61-67,105,107,108,110].

Although using a mixture of tools to collect data from social media is possible, only one study used a combination of APIs and manual tools [30].

Developing Comprehensive Search Strategies

This review found a diverse range of search strategies developed to extract data from social media platforms. Simple search strategies with one to three keywords were most common. Only 10 studies used more extensive search strategies with Boolean operators to link keywords (eg, AND, OR) or truncations to identify words with different endings (eg, vaccin*) [27,29,42,43,50,51,66,101,105,108]. Although simpler search strategies were perceived as a limitation by some [8,71,73], no data were available on the accuracy of short strategies as opposed to longer and more complex search strategies. Studies that evaluated their search strategies found that the categorization of keywords into "relevant," "semirelevant," and "nonrelevant" can increase precision [43], and that keywords should reflect cultural and normative differences [108].

Across all studies, most keywords were related to vaccines (ie, synonyms of the word "vaccine" or brand names of vaccines) and vaccine-preventable diseases. Some studies also searched for adverse events claimed to be linked to vaccination by the public (eg, autism, autoimmune disorders), keywords related to specific controversies (eg, mercury, big pharma, aluminum), or the names of people involved in controversies (eg, Jenny McCarthy, Andrew Wakefield). In addition to keywords, certain studies used hashtags (eg, #vaccine, #cdcwhistleblower, #vaccineswork) [8,30,34,44,48,49,55,60,61,105], questions inputted into search engines (eg, "should I get the HPV [human papillomavirus] vaccine?") [75,102], or phrases to refer to specific events (eg, "fainting in school children after vaccine") [89].

Predefined exclusion criteria were also used to screen data and exclude irrelevant or duplicate results. The question of how to deal with data from dubious sources, trolls, or bots was raised, and although researchers in two studies decided to exclude them, two studies specifically analyzed them and acknowledged their impact on the quality and validity of their findings [31,45,57,108].

Data Analysis Phase

Analyzing Metadata, Including Geolocation

The included studies analyzed a range of metadata, from the number of posts to users' characteristics. Information about the geographical source of social media data was extremely difficult to obtain, as this information was often private, not provided by social media users, or, as one study conducted on Twitter explained, because "accurate location information can be found in only a small proportion of tweets that have coordinates stored in the metadata of the tweet" [40]. This could explain why most of the studies in this review were performed "globally" (n=41). Despite these challenges, three types of strategies were used to restrict data to certain regions or countries: using keywords in local languages; using location-specific search terms (eg, United Kingdom, Scotland); and directly identifying local or national Facebook groups, pages, or online discussion forums [50,82,86,90,91,93]. Once social media posts were collected, other tools were used to identify and analyze the precise location of data. Some studies manually screened content or collected metadata [39,56,108], whereas others used automated mechanisms and software (Carmen, Geodict, Nominatim, GeoSocial Gauge) to retrieve this information from Twitter [28,29,40,41,54,61,66]. Two studies used dictionaries of terms for geographical entities of countries (GeoNames and the US Office of Management and Budget's Metropolitan and Micropolitan Statistical Areas) to automatically identify mentions of countries or cities in social media posts or profile pages [29,61]. Some authors also explained that most bots spreading negative content about vaccination online do not report their locations, which could explain why most tweets with geolocation information available were more positive toward vaccination [61].

Analyzing Trends, Content, and Sentiments

Once data from social media were extracted from studies, different analyses were performed, ranging from detecting the number of posts available over a period of time to more detailed content analysis to identify the frequency of particular concerns or conspiracies around vaccination [8,27,46-48,53,59,60,68-71,75-77,79,82,83,90,94,99]. Several studies also performed qualitative thematic analysis [45,56,71,83,85,86,95,99,110], or language and discourse analysis [26,50,52,84,92,93,106]. Four studies compared social media posts to disease incidence or outbreak cases [40,103,105,107].

The most common type of analysis looked at sentiments expressed toward vaccination (70%, 60/86). Sentiments can be understood in a variety of ways, reflected by the range of words identified to designate sentiments toward vaccination across all studies. Most studies used the terms "negative" (n=33), "positive" (n=31), or "neutral" (n=37); however, each study defined these in a slightly different way, which could have influenced study findings and what was perceived by researchers as "negative" or "positive." Other common sentiments were anti- or provaccine, encouraging or discouraging, ambiguous, or hesitant. Only two studies provided a more comprehensive list of sentiments such as frustration, humor, sarcasm, concern, relief, or minimized risk [53,107]. One study also looked at sentiment as a "yes or no" question: "does this message indicate that someone received or intended to receive a flu vaccine?" [41]. In one study, the World Health Organization determinants of the vaccine hesitancy framework were used to design and test a list of sentiments [90].

Sentiment was determined not only by looking at social media posts or comments but also by coding links, headlines, sources, images, captions, or hashtags [27,35,42,46,60,64,99]. Coding hashtags was sometimes difficult; for example, those using the hashtag #antivaxxers were often denouncing vaccine hesitancy. Similarly, "positive" hashtags such as #provaxxers can be used in a negative context to criticize those who promote vaccination

XSL•FO
RenderX

[27]. Coding sarcasm, irony, slang, and hyperboles was also complicated and prone to subjectivity biases [27,32,34,42,44,63,101].

Sentiment analyses were performed manually (40/86, 67%) or using an automated system (19/86, 32%). When data were analyzed manually, studies used multiple coders (between 2 and 4) and assessed interrater reliability scores to ensure accuracy and reliability. Many studies also emphasized the need to provide coders with training and codebooks with precise definitions of codes [8,26,31,32,46,47,53,60,71,73,83,90,107,109]. Manual sentiment analysis was prone to limitations, particularly because it relied on subjective coding and was labor-intensive, thereby reducing the total number of posts that could be analyzed by a single person [8,43,68,107].

Sentiments were also analyzed using automated systems, with most studies using such systems performed on Twitter (16/19, 84%). Leverage or supervised machine learning was used to code sentiments by training machines to learn how to code different sentiments using a set of manually coded results (ranging between 693 and 8261 posts) [28,34,37-39,41,48,49,51,57,62,63,66,67,82]. An alternative to manually coding some results to train the machine was to use Amazon Mechanical Turk [41,51,57,66]. Other automated systems that have been used to code sentiments included Latent Dirichlet Allocation, an unsupervised machine-learning algorithm that automatically determines topics in a text [57,101]; Naive Bayes [65]; Lightside [61]; BrightView classifier from Crimson Hexagon [105]; and Topsy [29,44]. Using such programs also came with limitations, including the aptitude of machines to correctly detect sentiments around vaccination, the reliance on manual coding of some part of the data to train the system prone to researchers' biases and subjectivity, and the need for high computational and technical skills [28,44,101].

Assessing the Reach of Social Media

Overall, 49 studies measured potential social media reach and thus estimations of the number of people that see content posted on social media. Reach was determined by the number of followers a user had, as well as the number of engagements with a post (eg, retweeted, shared, saved, liked, and commented upon).

Most studies provided a short descriptive summary of the reach of social media posts, whereas some proposed more detailed or comprehensive analyses. Interactions between different social network communities were studied to understand how information can spread and be shared on social media. Studies found that analyzing retweets was useful to understand the spread of certain sentiments toward vaccination and, in the case of disease outbreaks, to detect how the spread of social media information online can impact vaccination coverage. One particular study investigated how two kinds of communities interacted with each other within conversations about health and its relation to vaccines [62]. From a retweet network of 660,892 tweets published by 269,623 users, the study compared "structural community" with another "opinion group," and used community detection algorithms and autotagging to measure the interaction, sentiment, and influence that retweets had in conversations between the two communities [62]. Similarly, another study focused on shared concerns about the HPV vaccine and assessed how international followers express similar concerns to those of the groups or individuals they follow [56].

Another study examined communication patterns revealed through retweeting, assessing the impact of various sources of information, contrasting diverse types of authoritative content (eg, health organizations and official news organizations) and grassroots campaign arguments (with the antivaccination community views serving as a prototypical example) [54]. Finally, one study looked at tweeted images, and evaluated predictive factors for determining whether an image was retweeted, including the sentiment of the image and the objects shown in the image [33].

Discussion

Principal Findings

Over 80 articles have been published on social media monitoring around vaccination. This growing academic interest, particularly since 2015, acknowledges the role of social media in influencing public confidence in vaccination, and emphasizes the need to better understand the types of information about vaccination circulating on social media and its spread within and between online social networks [114,115]. Social media monitoring still constitutes a relatively new research field, for which tools and approaches continue to evolve. A wide range of methods, varied in style and complexity, have been identified and summarized through this systematic scoping review.

In an effort to summarize media monitoring articles, we developed a three-step model for this review. The first stage, preparation, consists of defining the purpose of social media monitoring while considering any potential ethical issues. The second stage, data extraction, should include the selection of data extraction tools as well as periods of monitoring, and the development of targeted, comprehensive, and precise search strategies. Finally, the third stage, data analysis, could focus on different types of analyses: metadata and geolocation, trends, content, sentiment, or reach. The model was found to be useful in structuring methodologies for social media monitoring, and could be used in the future as a standardized protocol for performing social media monitoring. Further research could be performed to evaluate different components of the model, and propose a more detailed and complex framework for media monitoring.

Standardization of Social Media Monitoring Methods

Although the large number of articles identified via the scoping review provided sufficient evidence to summarize methods that have been used to monitor social media, almost none of the articles evaluated the precision and accuracy of their monitoring and analysis methodologies. Furthermore, researchers have not drawn on a coherent body of agreed-upon methodologies, and instead created an amalgamation of methodological choices that sets no standards for the right sample size, no recommended time period for different types of analyses per platform, or no recommendations for studying the extremes of positive or negative views (which are not always representative of the general population) [116]. There is also a lack of standardization

XSL•FO
RenderX

of which specific API tools or analytical classifiers to use for analyzing social media discourse, interaction, or trends.

There have been recent calls to better standardize social media monitoring methodologies, including practices such as search strategies, so that the quality of the data query is reflected in more accurate and precise data and study findings [117]. However, it may be apt that social media monitoring remains a flexible research design, as the nature and access to social media discourse on vaccination is continuously evolving. The fast-evolving nature of different social media platforms, the crossover of shared data, boundaries to privacy, and public policy surrounding public discourse on vaccination and disease outbreaks may also necessitate a more methodologically diverse approach to keep up with ever-changing developments. Although standardization may not be the best practice for this relatively new research design, there is a need to evaluate the different tools that will be used at each stage of social media monitoring to determine which ones offer the most precise, accurate, and representative results.

Establishing the Purpose of Social Media Monitoring

Using social media to understand prevailing issues of interest and concern in certain communities can be a useful listening tool for public health institutions, which can then use media monitoring to detect key themes or questions around vaccination circulating in the population. However, many studies included in this review explicitly discussed limitations regarding the lack of population representativeness in investigating social media content. Evidence shows that social media users often represent specific population groups in terms of age, gender, education level, or socioeconomic status [118]. For instance, users discussing vaccination online were found to be younger and female [61]. Another challenge comes from the fact that content being shared by social media users is not always representative of their personal views or feelings, with evidence showing that social media content is often more extreme or impulsive [119]. Due to these issues, social media monitoring is best seen as an alternative to surveys or qualitative interviews in obtaining data about vaccination beliefs and opinions, without assuming representativeness of total populations but rather specific interest groups.

Social media users could also be considered as a configuration of a research population group, and the field of social media monitoring could be seen as an opportunity to understand what information users are exposed to, and how information about vaccination is shared and spread online. In this way, social media monitoring would be used as a new research methodology to study a new type of population. Social media monitoring comes with representativeness challenges of its own, as access to data becomes limited due to inaccessible private content, the challenge of studying all social media platforms at once, or limitations imposed by automated software. However, social media monitoring opens the door to more dynamic research that continuously evolves and responds to a perpetually changing world.

Important Ethical Considerations

The considerable increase in the number of social media monitoring studies poses questions regarding the safe use of data available online. Even though researchers in previous studies may not have been legally compelled to obtain ethics approval, the lack of guidance on good ethical conduct when using social media information is a cause for concern. Issues of confidentiality and anonymization of data still arise, as some studies included in this review published screenshots of users' data that included users' profile names. Another issue relates to data coming from minors, which should be considered more carefully, even when publicly available [120]. Although these concerns should not unduly hinder the development of social media monitoring, they should highlight the need for guidelines to ensure ethical conduct and respect for social media users, and the importance of submitting research proposals to ethics boards.

Recent controversies with regard to the exploitation of users' data in the Facebook and Cambridge Analytica scandal, and the public outcry of users feeling unnerved being monitored and manipulated, have indeed opened up conversations and legislation around the ethics of handling user data from social media in research [121]. The overall aim of the 2018 EU General Data Protection Regulation (GDPR) is to increase people's control over their own personal data and "to protect all EU citizens from privacy and data breaches in an increasingly data-driven world" [122]. For companies, organizations, and researchers, this means obtaining consent for using and retaining customers' personal data, while granting more rights to the "data subject" to be informed and to control how their personal data are used. Such legislation may change the way future researchers must anonymize data, as well as restrict what sections of social media platforms (eg, public vs private) are available for research [123].

Accessing Data From Different Social Media Platforms: The Twitter Bias

The majority of studies that had the largest datasets, collected over longer time periods, were those with access to social media platforms' APIs or automated data collection tools. Studies that had smaller samples and used less sophisticated keyword searches were those that relied on manual data collection, and were thus constrained by time, resources, and the limitations of the browser tools used. Although some studies also discussed the time-consuming nature of manual data collection methodologies, the time required to perform searches was not commonly discussed. More research could be performed to evaluate the clear benefits and limitations of manual and automated data extraction tools, including the time required to complete searches. Studies that used the paid version of APIs via automated monitoring software seemed to have a more representative sample, as access to paid data offers access to all historical and current posts. However, there are still issues with the relative opaqueness of the paid access to Twitter, Facebook, or YouTube APIs, which do not advertise the mechanisms behind collection of data, do not inform researchers of what percentage of "all" data they are given, and may thus not provide representative data [124,125]. This also prevents researchers

XSL•FO
RenderX

from fully comparing studies over time, as the API sampling algorithm itself will change. It is arguable that such an environment presents risks and opportunities both for data collection strategies in terms of availability and data privacy issues, as well as an evolution in our fundamental understanding of how social media research fits into the rapidly changing public discourse in relation to vaccines. Finally, the financial cost associated with the use of most APIs and automated tools could constitute a barrier to those in lower resource settings [126].

One of the main reasons for the bias toward the use of Twitter in a majority of studies within this review may be because Twitter provides the most openly available API, both for free and with paid access [116]. However, studies using these freely collected tweets only have access to a small (1%) sample of all tweets, creating representativeness challenges [113]. Accessing the free Twitter API also raises issues around periodical collection due to restrictive access to intermittent collection points. This means that any data collection is limited to pockets of time that are not necessarily continuous, truncating the 1% sample into different time periods [127]. This focus on Twitter constitutes an important bias for social media monitoring research, as it fails to capture the real-time evolution of the social media environment, and the flow of users and content from platform to platform [128].

Finally, although subscription-based data analytics companies provide more comprehensive access to other APIs such as YouTube and Instagram, the data provided by these companies can be more skewed toward those of brand marketing (eg, brand strength, brand influencers, and brand trends) [129]. However, there have been growing opportunities in recent years to allow academics to work in partnership with data analytics companies to forge a better understanding of how to look at social media images and text from a social sciences and public policy perspective [130].

Social Media Analyses: Complexity of Analyzing Sentiments

Although social media data can be analyzed in various ways, ranging from analyzing trends in the number of posts identified to more complex content analyses, most studies focused on sentiment analyses. Identifying sentiments toward vaccination expressed in social media can be useful to detect changes in beliefs and possible drops in confidence. Manual and automated methods of analyzing sentiments both come with their own benefits and limitations: manual coding may be easier and requires less technical skills than automated coding, whereas it is more prone to subjectivity biases and is time-consuming, and therefore does not allow for the analysis of a large number of posts. The complexity of coding discourse, particularly those using sarcasm or irony, is apparent with both automated and manual systems [40]. Researchers should choose analytical methods based on their personal objectives and resources available. Despite the possibility of using a combination of manual and automated coding or selecting automated systems that require less technical skills, more accurate and easy-to-use automated systems should also be developed.

The development of sentiment analysis as a tool in social media monitoring raises other challenges. Most studies identified in this review used simple binary categorizations of sentiments (eg, “negative” vs “positive”). However, discussions around vaccination tend to elicit complex sentiments, closely linked to deeper, more contextual themes of trust, confidence, and risk perception [131]. Categorizing sentiments as either negative or positive therefore fails to recognize nuances that would be crucial for the development of targeted responses to rebuild trust in vaccination. If automated coding systems are to be further developed, they need to take into account the nuances in sentiments around vaccination and move beyond the use of binary variables. More complex sentiment analysis will also improve the quality of the coding of videos, images, and emojis [132-135].

Considerations for Future Research: Changing Digital Ecosystem

Following the Cambridge Analytica data misuse scandal and an increase in the amount of antivaccine content, Facebook announced a number of API changes aimed at better protecting user information between 2017 and 2019. These restrictions, along with GDPR laws, will pose restrictions on the type of data and research that can be performed on social media platforms and will require researchers to continuously adapt their methodologies [136,137]. Furthermore, platforms such as Pinterest, Facebook, and YouTube are responding to requests from public bodies to alter their content to respond to concerns about the spread of misinformation about vaccines and the presence of antivaccination content on social media [138-140]. These actions from social media platforms may change what users see but also what researchers study. Indeed, it may be that antivaccine groups move away from platforms that no longer monetize or make it easy for them to share information. Those with antivaccination views have not only been found to be using a mixture of websites and social media but also to migrate over to the dark web, where they are able to create and construct content-specific platforms from which their chosen ideologies can be shared [141]. The question of who should decide what content falls under antivaccination sentiment is also important. Social media platforms should work closely with vaccination experts to identify which posts to remove or keep, especially to avoid infringing on the public’s freedom of expression.

Study Limitations

Some limitations of this systematic scoping review should be acknowledged. Although articles in languages other than English were included for analysis, the search itself only used English keywords, which could have limited the results. Furthermore, the search strategy was comprehensive but did not include certain relevant keywords such as “infodemiology” or “infoveillance,” which should be considered in future research. Data extraction was performed by three researchers, which could have caused inconsistencies even though the same data extraction sheet was used. Finally, social media monitoring constitutes a relatively new research field, which means that many real-life, practical uses of monitoring may have been omitted as they may not have been published in publicly available peer-reviewed journals or reports. It is also important

XSL•FO
RenderX

to note that as this is a fast-moving field, a high number of articles have been published since this review was performed, particularly around the COVID-19 pandemic. Methodologies for social media monitoring are expected to continuously and rapidly evolve, and further research should be performed to regularly update this review.

Conclusion

Social media has changed the communication landscape around vaccination. The increasing use of social media by individuals to find and share information about vaccination, together with the growing volume of negative information about vaccination online, has influenced the way people assess the risks and benefits of vaccination. Social media monitoring studies have been developed with the aim of better understanding the type of information social media users are exposed to, and how this information is spread and shared across the world. This review has identified clear steps to perform social media monitoring that can be organized in three phases: (1) Preparation (defining the purpose of media monitoring, addressing ethical considerations); (2) Data extraction (selecting data extraction tools, developing comprehensive search strategies); and (3) Data analysis (geolocation, trends, content, sentiments, and reach). A wide range of tools for each of these steps have been identified in the literature but have not yet been evaluated. Therefore, to establish social media monitoring as a valuable research design, future research should aim to identify which methods are more robust and precise to extract and analyze data from social media.

Acknowledgments

This study was commissioned by the ECDC under Service Contract ECD8894. The authors would like to thank Franklin Apfel, Sabrina Cecconi, Daniel Artus, Sandra Alexiu, Maged Kamel Boulos, John Kinsman, Andrea Würz, and Marybelle Stryk for providing guidance on the methodology and reviewing the report this manuscript is based on.

Conflicts of Interest

The LSHTM research group “The Vaccine Confidence Project” has received funding for other projects from the Bill & Melinda Gates Foundation, the Center for Strategic and International Studies, Innovative Medicines Initiative (IMI), National Institute for Health Research (UK), the World Health Organization, GlaxoSmithKline (GSK), Novartis, Johnson & Johnson, and Merck. HJL has done consulting on vaccine confidence with GSK and served on the Merck Vaccine Strategic Advisory Board. EK has received support for participating in meetings and advisory roundtables organised by GSK and Merck. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Search strategy developed for the EMBASE database.
[PDF File (Adobe PDF File), 522 KB - publichealth_v7i2e17149_app1.pdf]

Multimedia Appendix 2
Summary of included articles.
[PDF File (Adobe PDF File), 622 KB - publichealth_v7i2e17149_app2.pdf]

References

1. Poland GA, Jacobson RM. The age-old struggle against the antivaccinationists. N Engl J Med 2011 Jan 13;364(2):97-99. [doi: 10.1056/NEJMp1010594] [Medline: 21226573]
2. Stahl JP, Cohen R, Denis F, Gaudelus J, Martinot A, Lery T, et al. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. Med Mal Infect 2016 May;46(3):117-122. [doi: 10.1016/j.medmal.2016.02.002] [Medline: 26987960]
3. Larson HJ, de Figueiredo A, Xiahong Z, Schulz WS, Verger P, Johnston IG, et al. The State of Vaccine Confidence 2016: Global Insights Through a 67-Country Survey. EBioMedicine 2016 Oct;12:295-301 [FREE Full text] [doi: 10.1016/j.ebiom.2016.08.042] [Medline: 27658738]
4. Larson HJ, Bertozzi S, Piot P. Redesigning the AIDS response for long-term impact. Bull World Health Organ 2011 Nov 01;89(11):846-852 [FREE Full text] [doi: 10.2471/BLT.11.087114] [Medline: 22084531]
5. Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. Soc Sci Med 2014 Jul;112:1-11 [FREE Full text] [doi: 10.1016/j.socscimed.2014.04.018] [Medline: 24788111]
6. Black S, Rappuoli R. A crisis of public confidence in vaccines. Sci Transl Med 2010 Dec 08;2(61):61mr1. [doi: 10.1126/scitranslmed.3001738] [Medline: 21148125]
7. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May;5(5) [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
8. Blankenship EB, Goff ME, Yin J, Tse ZTH, Fu K, Liang H, et al. Sentiment, contents, and retweets: a study of two vaccine-related Twitter datasets. Perm J 2018;22:17-138 [FREE Full text] [doi: 10.7812/TPP/17-138] [Medline: 29911966]

XSL•FO
RenderX

9. Basch CH, MacLean SA. A content analysis of HPV related posts on instagram. Hum Vaccin Immunother 2019;15(7-8):1476-1478 [FREE Full text] [doi: 10.1080/21645515.2018.1560774] [Medline: 30570379]
10. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
11. Konijn EA, Veldhuis J, Plaisier XS. YouTube as a research tool: three approaches. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):695-701. [doi: 10.1089/cyber.2012.0357] [Medline: 23659721]
12. Kelly J, Fealy G, Watson R. The image of you: constructing nursing identities in YouTube. J Adv Nurs 2012 Aug;68(8):1804-1813. [doi: 10.1111/j.1365-2648.2011.05872.x] [Medline: 22070735]
13. Pink S. Digital Ethnography: Principles and Practices. London, UK: SAGE Publications Ltd; 2016.
14. Brunson EK. The impact of social networks on parents' vaccination decisions. Pediatrics 2013 May;131(5):e1397-e1404. [doi: 10.1542/peds.2012-2452] [Medline: 23589813]
15. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. Nature 2020 Jun;582(7811):230-233. [doi: 10.1038/s41586-020-2281-1] [Medline: 32499650]
16. Leask J, Kinnersley P, Jackson C, Cheater F, Bedford H, Rowles G. Communicating with parents about vaccination: a framework for health professionals. BMC Pediatr 2012 Sep 21;12:154 [FREE Full text] [doi: 10.1186/1471-2431-12-154] [Medline: 22998654]
17. Bahri P, Fogd J, Morales D, Kurz X, ADVANCE consortium. Application of real-time global media monitoring and 'derived questions' for enhancing communication by regulatory bodies: the case of human papillomavirus vaccines. BMC Med 2017 May 02;15(1):91 [FREE Full text] [doi: 10.1186/s12916-017-0850-4] [Medline: 28464900]
18. Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC Public Health 2019 Oct 23;19(1):1348 [FREE Full text] [doi: 10.1186/s12889-019-7659-3] [Medline: 31640660]
19. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
20. Social media strategy development: a guide to using social media for public health communication. European Centre for Disease Prevention Control. Stockholm, Sweden: ECDC; 2016. URL: https://www.ecdc.europa.eu/en/publications-data/social-media-strategy-development-guide-using-social-media-public-health [accessed 2021-01-25]
21. Systematic scoping review on social media monitoring methods and interventions relating to vaccine hesitancy. European Centre for Disease Prevention and Control. Stockholm URL: https://www.ecdc.europa.eu/en/publications-data/systematic-scoping-review-social-media-monitoring-methods-and-interventions [accessed 2021-01-28]
22. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Peters M, Godfrey C, McInerney P, Soares C, Khalil H, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews. Adelaide, SA Australia: The Joanna Briggs Institute; 2015. URL: https://nursing.lsuhsc.edu/JBI/docs/ReviewersManuals/Scoping-.pdf [accessed 2021-01-25]
25. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
26. Skea Z. Communication about the MMR vaccine: an analysis of newspaper coverage and an internet discussion forum.: PhD thesis, University of Aberdeen; 2006. URL: https://eu03.alma.exlibrisgroup.com/view/delivery/44ABE_INST/12153284070005941 [accessed 2021-01-25]
27. Addawood A. Usage of scientific references in MMR vaccination debates on Twitter. 2018 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); 2018; Barcelona p. 971-979. [doi: 10.1109/asonam.2018.8508385]
28. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly Available Online Tool Facilitates Real-Time Monitoring Of Vaccine Conversations And Sentiments. Health Aff (Millwood) 2016 Feb;35(2):341-347. [doi: 10.1377/hlthaff.2015.1092] [Medline: 26858390]
29. Becker BFH, Larson HJ, Bonhoeffer J, van Mulligen EM, Kors JA, Sturkenboom MCJM. Evaluation of a multinational, multilingual vaccine debate on Twitter. Vaccine 2016 Dec 07;34(50):6166-6171. [doi: 10.1016/j.vaccine.2016.11.007] [Medline: 27840012]
30. Bello-Orgaz G, Hernandez-Castro J, Camacho D. Detecting discussion communities on vaccination in twitter. Future Gener Comput Syst 2017 Jan;66:125-136. [doi: 10.1016/j.future.2016.06.032]
31. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]

XSL•FO
RenderX

32. Chakraborty P, Colditz JB, Silvestre AJ, Friedman MR, Bogen KW, Primack BA, et al. Observation of public sentiment toward human papillomavirus vaccination on Twitter. Cogent Medicine 2017 Dec 12;4(1):1390853. [doi: 10.1080/2331205x.2017.1390853]
33. Chen T, Dredze M. Vaccine images on Twitter: analysis of what images are shared. J Med Internet Res 2018 Apr 03;20(4):e130 [FREE Full text] [doi: 10.2196/jmir.8221] [Medline: 29615386]
34. D'Andrea E, Ducange P, Bechini A, Renda A, Marcelloni F. Monitoring the public opinion about the vaccination topic from tweets analysis. Expert Syst Appl 2019 Feb;116:209-226. [doi: 10.1016/j.eswa.2018.09.009]
35. Dredze M, Wood-Doughty Z, Quinn SC, Broniatowski DA. Vaccine opponents' use of Twitter during the 2016 US presidential election: Implications for practice and policy. Vaccine 2017 Aug 24;35(36):4670-4672. [doi: 10.1016/j.vaccine.2017.06.066] [Medline: 28688783]
36. Du J, Tang L, Xiang Y, Zhi D, Xu J, Song H, et al. Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models. J Med Internet Res 2018 Jul 09;20(7):e236 [FREE Full text] [doi: 10.2196/jmir.9413] [Medline: 29986843]
37. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 03;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]
38. 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]
39. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
40. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
41. Huang X, Smith MC, Paul MJ, Ryzhkov D, Quinn SC, Broniatowski DA. Examining patterns of influenza vaccination in social media. In: AAAI Publications. 2017 Presented at: Thirty-First AAAI Conference on Artificial Intelligence; 2017; San Francisco.
42. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
43. Kaptein R, Boertjes E, Langley D. Analyzing discussions on Twitter: Case study on HPV vaccinations. In: de Rijke M, Kenter T, de Vries AP, Zhai C, de Jong F, Radinsky K, et al, editors. Advances in Information Retrieval: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings. Switzerland: Springer, Cham; 2014:474-480.
44. Keim-Malpass J, Mitchell EM, Sun E, Kennedy C. Using Twitter to Understand Public Perceptions Regarding the #HPV Vaccine: Opportunities for Public Health Nurses to Engage in Social Marketing. Public Health Nurs 2017 Jul;34(4):316-323. [doi: 10.1111/phn.12318] [Medline: 28261846]
45. Krittanawong C, Tunhasiriwet A, Chirapongsathorn S, Kitai T. Tweeting influenza vaccine to cardiovascular health community. Eur J Cardiovasc Nurs 2017 Dec;16(8):704-706. [doi: 10.1177/1474515117707867] [Medline: 28440668]
46. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. Am J Infect Control 2013 Jun;41(6):568-570. [doi: 10.1016/j.ajic.2012.10.016] [Medline: 23726548]
47. Mahoney LM, Tang T, Ji K, Ulrich-Schad J. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. JMIR Public Health Surveill 2015;1(1):e2 [FREE Full text] [doi: 10.2196/publichealth.3310] [Medline: 27227125]
48. Massey PM, Budenz A, Leader A, Fisher K, Klassen AC, Yom-Tov E. What Drives Health Professionals to Tweet About #HPVvaccine? Identifying Strategies for Effective Communication. Prev Chronic Dis 2018 Feb 22;15:E26 [FREE Full text] [doi: 10.5888/pcd15.170320] [Medline: 29470166]
49. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
50. McNeill A, Harris PR, Briggs P. Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. Front Public Health 2016;4:26. [doi: 10.3389/fpubh.2016.00026] [Medline: 26942174]
51. Mitra T, Counts S, Pennebaker JW. Understanding anti-vaccination attitudes in social media. In: Association for the Advancement of Artificial Intelligence. 2016 Presented at: Tenth International AAAI Conference on Web and Social Media; 2016; Cologne, Germany.
52. Numnark S, Ingsriswang S, Wichadakul D. VaccineWatch: A monitoring system of vaccine messages from social media data. 2014 Presented at: The 8th International Conference on Systems Biology; October 24-27, 2014; Qingdao, China p. 112-117. [doi: 10.1109/isb.2014.6990742]

XSL•FO
RenderX

53. Porat T, Garaizar P, Ferrero M, Jones H, Ashworth M, Vadillo MA. Content and source analysis of popular tweets following a recent case of diphtheria in Spain. Eur J Public Health 2019 Feb 01;29(1):117-122. [doi: 10.1093/eurpub/cky144] [Medline: 30084926]
54. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]
55. Sanawi J, Samani M, Taibi M. #Vaccination: Identifying influencers in the vaccination discussion on twitter through social network visualisation. Int J Bus Soc 2017;18(S4):718-726.
56. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017 Oct 05;7(10):e016869 [FREE Full text] [doi: 10.1136/bmjopen-2017-016869] [Medline: 28982821]
57. Smith MC, Dredze M, Quinn SC, Broniatowski DA. Monitoring real-time spatial public health discussions in the context of vaccine hesitancy. 2017 Presented at: SMM4H@AMIA; 2017; Washington, DC.
58. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
59. Tang L, Bie B, Zhi D. Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease. Am J Infect Control 2018 Dec;46(12):1375-1380 [FREE Full text] [doi: 10.1016/j.ajic.2018.05.019] [Medline: 29929837]
60. Teoh D, Shaikh R, Vogel RI, Zoellner T, Carson L, Kulasingam S, et al. A Cross-Sectional Review of Cervical Cancer Messages on Twitter During Cervical Cancer Awareness Month. J Low Genit Tract Dis 2018 Jan;22(1):8-12 [FREE Full text] [doi: 10.1097/LGT.0000000000000363] [Medline: 29271850]
61. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
62. Yuan X, Crooks A. Examining online vaccination discussion and communities in Twitter. New York: Association for Computing Machinery; 2018 Presented at: SMSociety '18: International Conference on Social Media and Society; July 18-20, 2018; Copenhagen, Denmark p. 197-206. [doi: 10.1145/3217804.3217912]
63. D'Andrea E, Ducange P, Marcelloni F. Monitoring negative opinion about vaccines from tweets analysis. : IEEE; 2017 Presented at: Third international conference on research in computational intelligence and communication networks (ICRCICN); 2017; Kolkata p. 186-191. [doi: 10.1109/icrcicn.2017.8234504]
64. Ninkov A, Vaughan L. A webometric analysis of the online vaccination debate. J Assoc Inf Sci Technol 2017 Mar 20;68(5):1285-1294. [doi: 10.1002/asi.23758]
65. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
66. Pananos AD, Bury TM, Wang C, Schonfeld J, Mohanty SP, Nyhan B, et al. Critical dynamics in population vaccinating behavior. Proc Natl Acad Sci U S A 2017 Dec 26;114(52):13762-13767 [FREE Full text] [doi: 10.1073/pnas.1704093114] [Medline: 29229821]
67. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong M, Dunn A. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. Stud Health Technol Inform 2015;216:761-765. [Medline: 26262154]
68. Ache KA, Wallace LS. Human papillomavirus vaccination coverage on YouTube. Am J Prev Med 2008 Oct;35(4):389-392. [doi: 10.1016/j.amepre.2008.06.029] [Medline: 18675530]
69. Basch CH, Hillyer GC, Berdnik A, Basch CE. YouTube™ videos related to human papillomavirus: the need for professional communication. Int J Adolesc Med Health 2016 Apr 09;30(1). [doi: 10.1515/ijamh-2015-0122] [Medline: 27060736]
70. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? Child Care Health Dev 2017 Jul;43(4):499-503. [doi: 10.1111/cch.12442] [Medline: 28105642]
71. Briones R, Nan X, Madden K, Waks L. When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health Commun 2012;27(5):478-485. [doi: 10.1080/10410236.2011.610258] [Medline: 22029723]
72. Cambra U, Díaz V, Herrero S. Vacunas y anti vacunas en la red social Youtube. Opción: Revista de Ciencias Humanas y Sociales 2016;32:447-465.
73. Covolo L, Ceretti E, Passeri C, Boletti M, Gelatti U. What arguments on vaccinations run through YouTube videos in Italy? A content analysis. Hum Vaccin Immunother 2017 Jul 03;13(7):1693-1699 [FREE Full text] [doi: 10.1080/21645515.2017.1306159] [Medline: 28362544]
74. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: A quantitative analysis of YouTube videos. Hum Vaccin Immunother 2018 Jul 03;14(7):1654-1659 [FREE Full text] [doi: 10.1080/21645515.2018.1454572] [Medline: 29553872]

XSL•FO
RenderX

75. Ekram S, Debiec KE, Pumper MA, Moreno MA. Content and Commentary: HPV Vaccine and YouTube. J Pediatr Adolesc Gynecol 2019 Apr;32(2):153-157. [doi: 10.1016/j.jpag.2018.11.001] [Medline: 30445163]
76. Hernández-García I, Fernández Porcel C. Characteristics of YouTube™ videos in Spanish about the vaccine against meningococcus B. Vacunas (English Edition) 2018 Jul;19(2):37-43. [doi: 10.1016/j.vacune.2018.11.002]
77. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 05;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
78. Tuells J, Martínez-Martínez PJ, Duro-Torrijos JL, Caballero P, Fraga-Freijeiro P, Navarro-López V. [Characteristics of the Videos in Spanish Posted on Youtube about Human Papillomavirus Vaccines]. Rev Esp Salud Publica 2015;89(1):107-115 [FREE Full text] [doi: 10.4321/S1135-57272015000100012] [Medline: 25946591]
79. Venkatraman A, Garg N, Kumar N. Greater freedom of speech on Web 2.0 correlates with dominance of views linking vaccines to autism. Vaccine 2015 Mar 17;33(12):1422-1425. [doi: 10.1016/j.vaccine.2015.01.078] [Medline: 25665960]
80. Buchanan R, Beckett RD. Assessment of vaccination-related information for consumers available on Facebook. Health Info Libr J 2014 Sep;31(3):227-234. [doi: 10.1111/hir.12073] [Medline: 25041499]
81. Faasse K, Chatman CJ, Martin LR. A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post. Vaccine 2016 Nov 11;34(47):5808-5814. [doi: 10.1016/j.vaccine.2016.09.029] [Medline: 27707558]
82. Furini M, Menegoni G. Public Health and Social Media: Language Analysis of Vaccine Conversations. New York: IEEE; 2018 Presented at: International workshop on social sensing (SocialSens); Apr 17, 2018; Orlando, FL p. 50-55.
83. Luisi MLR. Who Gives a “Like” About the HPV Vaccine? Kansan Parent/Guardian Perceptions and Social Media Representations. Doctoral dissertation, University of Kansas. 2017. URL: https://kuscholarworks.ku.edu/bitstream/handle/1808/26037/Luisi_ku_0099D_15291_DATA_1.pdf?sequence=1 [accessed 2021-01-25]
84. Ma J, Stahl L. A multimodal critical discourse analysis of anti-vaccination information on Facebook. Libr Inf Sci Res 2017 Oct;39(4):303-310. [doi: 10.1016/j.lisr.2017.11.005]
85. Orr D, Baram-Tsabari A. Science and Politics in the Polio Vaccination Debate on Facebook: A Mixed-Methods Approach to Public Engagement in a Science-Based Dialogue. J Microbiol Biol Educ 2018;19(1) [FREE Full text] [doi: 10.1128/jmbe.v19i1.1500] [Medline: 29904532]
86. Orr D, Baram-Tsabari A, Landsman K. Social media as a platform for health-related public debates and discussions: the Polio vaccine on Facebook. Isr J Health Policy Res 2016;5:34 [FREE Full text] [doi: 10.1186/s13584-016-0093-4] [Medline: 27843544]
87. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
88. Smith N, Graham T. Mapping the anti-vaccination movement on Facebook. Inf Commun Soc 2017 Dec 27;22(9):1310-1327. [doi: 10.1080/1369118x.2017.1418406]
89. Suragh TA, Lamprianou S, MacDonald NE, Loharikar AR, Balakrishnan MR, Benes O, et al. Cluster anxiety-related adverse events following immunization (AEFI): An assessment of reports detected in social media and those identified using an online search engine. Vaccine 2018 Sep 25;36(40):5949-5954 [FREE Full text] [doi: 10.1016/j.vaccine.2018.08.064] [Medline: 30172632]
90. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. User-driven comments on a Facebook advertisement recruiting Canadian parents in a study on immunization: content analysis. JMIR Public Health Surveill 2018 Sep 20;4(3):e10090 [FREE Full text] [doi: 10.2196/10090] [Medline: 30249585]
91. Vayreda A, Antaki C. To vaccinate or not? The disqualification of commercial sources of health advice in an online forum. Commun Med 2011;8(3):273-282. [doi: 10.1558/cam.v8i3.273] [Medline: 23264989]
92. Goh D, Chi J. Central or peripheral? Information elaboration cues on childhood vaccination in an online parenting forum. Comput Hum Behav 2017 Apr;69:181-188. [doi: 10.1016/j.chb.2016.11.066]
93. Penţa MA, Băban A. Dangerous agent or saviour? HPV vaccine representations on online discussion forums in Romania. Int J Behav Med 2014 Feb;21(1):20-28. [doi: 10.1007/s12529-013-9340-z] [Medline: 24057409]
94. Fadda M, Allam A, Schulz PJ. Arguments and sources on Italian online forums on childhood vaccinations: Results of a content analysis. Vaccine 2015 Dec 16;33(51):7152-7159. [doi: 10.1016/j.vaccine.2015.11.007] [Medline: 26592140]
95. Nicholson MS, Leask J. Lessons from an online debate about measles-mumps-rubella (MMR) immunization. Vaccine 2012 May 28;30(25):3806-3812. [doi: 10.1016/j.vaccine.2011.10.072] [Medline: 22063388]
96. Luoma‐aho V, Tirkkonen P, Vos M. Monitoring the issue arenas of the swine‐flu discussion. JCOM 2013 Jul 26;17(3):239-251. [doi: 10.1108/jcom-11-2010-0069]
97. Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, et al. "Mommy Blogs" and the vaccination exemption narrative: results from a machine-learning approach for story aggregation on parenting social media sites. JMIR Public Health Surveill 2016 Nov 22;2(2):e166 [FREE Full text] [doi: 10.2196/publichealth.6586] [Medline: 27876690]
98. Skeppstedt M, Kerren A, Stede M. Vaccine hesitancy in discussion forums: computer-assisted argument mining with topic models. Stud Health Technol Inform 2018;247:366-370. [Medline: 29677984]
99. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]

XSL•FO
RenderX

100. Chen B, Zhang JM, Jiang Z, Shao J, Jiang T, Wang Z, et al. Media and public reactions toward vaccination during the 'hepatitis B vaccine crisis' in China. Vaccine 2015 Apr 08;33(15):1780-1785. [doi: 10.1016/j.vaccine.2015.02.046] [Medline: 25731787]
101. Rivera I, Warren J, Curran J. Quantifying mood, content and dynamics of health forums. In: ACSW '16: Proceedings of the Australasian Computer Science Week Multiconference. New York, USA: Association for Computing Machinery; 2016 Presented at: Australasian Computer Science Week Multiconference; February 2016; Canberra, Australia. [doi: 10.1145/2843043.2843379]
102. Lopez C, Krauskopf E, Villota C, Burzio L, Villegas J. Cervical cancer, human papillomavirus and vaccines: assessment of the information retrieved from general knowledge websites in Chile. Public Health 2017 Jul;148:19-24. [doi: 10.1016/j.puhe.2017.02.017] [Medline: 28404529]
103. Nawa N, Kogaki S, Takahashi K, Ishida H, Baden H, Katsuragi S, et al. Analysis of public concerns about influenza vaccinations by mining a massive online question dataset in Japan. Vaccine 2016 Jun 08;34(27):3207-3213. [doi: 10.1016/j.vaccine.2016.01.008] [Medline: 26776467]
104. Seeman N, Ing A, Rizo C. Assessing and responding in real time to online anti-vaccine sentiment during a flu pandemic. Healthc Q 2010;13 Spec No:8-15. [doi: 10.12927/hcq.2010.21923] [Medline: 20959725]
105. Deiner MS, Fathy C, Kim J, Niemeyer K, Ramirez D, Ackley SF, et al. Facebook and Twitter vaccine sentiment in response to measles outbreaks. Health Informatics J 2019 Sep;25(3):1116-1132 [FREE Full text] [doi: 10.1177/1460458217740723] [Medline: 29148313]
106. Sundstrom B, Aylor E, Cartmell KB, Brandt HM, Bryant DC, Hughes Halbert C, et al. Beyond the birds and the bees: a qualitative content analysis of online HPV vaccination communication. J Commun Healthc 2018 Jun 06;11(3):205-214. [doi: 10.1080/17538068.2018.1484984]
107. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. J Med Internet Res 2015 May 26;17(5):e128 [FREE Full text] [doi: 10.2196/jmir.3863] [Medline: 26013683]
108. Jang SM, Mckeever BW, Mckeever R, Kim JK. From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada, and the UK. Health Commun 2019 Jan;34(1):110-117. [doi: 10.1080/10410236.2017.1384433] [Medline: 29028371]
109. Aquino F, Donzelli G, De Franco E, Privitera G, Lopalco PL, Carducci A. The web and public confidence in MMR vaccination in Italy. Vaccine 2017 Aug 16;35(35 Pt B):4494-4498. [doi: 10.1016/j.vaccine.2017.07.029] [Medline: 28736200]
110. Lehmann BA, Ruiter RA, Kok G. A qualitative study of the coverage of influenza vaccination on Dutch news sites and social media websites. BMC Public Health 2013 Jun 05;13:547 [FREE Full text] [doi: 10.1186/1471-2458-13-547] [Medline: 23738769]
111. Zheleva E, Getoor L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on world wide web. 2009 Presented at: 18th international conference on World wide web; April 20-24, 2009; Madrid, Spain. [doi: 10.1145/1526709.1526781]
112. Lea D, Bradbery J. Oxford Advanced Learner's Dictionary. Oxford: Oxford University Press; 2020.
113. Boyd D, Crawford K. Critical questions for big data. Inf Commun Soc 2012 Jun;15(5):662-679. [doi: 10.1080/1369118x.2012.678878]
114. Lewandowsky S, Ecker U, Cook J. Beyond misinformation: understanding and coping with the “post-truth” era. J Appl Res Mem Cogn 2017 Dec;6(4):353-369. [doi: 10.1016/j.jarmac.2017.07.008]
115. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need Social media now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
116. Cihon P, Yasseri T. A biased review of biases in Twitter studies on political collective action. Front Phys 2016 Aug 09;4(34). [doi: 10.3389/fphy.2016.00034]
117. Kim Y, Huang J, Emery S. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. J Med Internet Res 2016 Feb 26;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
118. Hargittai E. Potential biases in big data: omitted voices on social media. Social Science Computer Review 2018 Jul 30;38(1):10-24. [doi: 10.1177/0894439318788322]
119. Beninger K, Fry A, Jago N, Lepps H, Nass L, Silvester H. Research using social media; users' views. London: NatCen Social Research; 2014. URL: https://www.bl.uk/britishlibrary/~/media/bl/global/social-welfare/pdfs/non-secure/r/e/s/research-using-social-media-users-views.pdf [accessed 2021-01-25]
120. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
121. González F, Yu Y, Figueroa A, López C, Aragon C. Global Reactions to the Cambridge Analytica Scandal: An Inter-Language Social Media Study. 2019 Presented at: WWW '19: The Web Conference; May 2019; San Francisco. [doi: 10.1145/3308560.3316456]

XSL•FO
RenderX

122. GDPR – the fabric of a success story. European Commission. Brussels, Belgium; 2020. URL: https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules/gdpr-fabric-success-story_en [accessed 2021-01-25]
123. Carr I, Zimmermann K, van't Veer P, Teare H. Ethical design: At the Interface of Ethics for Big Data and the European Union's General Data Protection Regulation. European Commission. 2018. URL: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ba270ebd&appId=PPGMS [accessed 2021-01-25]
124. Driscoll K, Walker S. Big data, big questions. Working within a black box: Transparency in the collection and production of big twitter data. Int J Commun 2014;8:20 [FREE Full text]
125. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. 2013 Presented at: Seventh international AAAI conference on weblogs and social media; July 8-11, 2013; Cambridge, MA.
126. Halfpenny P, Procter R. In: Seaman J, editor. Innovations in digital research method. London: Sage; 2015.
127. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y. The dynamics of protest recruitment through an online network. Sci Rep 2011;1:197. [doi: 10.1038/srep00197] [Medline: 22355712]
128. Choi TR, Sung Y. Instagram versus Snapchat: Self-expression and privacy concern on social media. Telemat Inform 2018 Dec;35(8):2289-2298. [doi: 10.1016/j.tele.2018.09.009]
129. Tufekci Z. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Michigan, USA.
130. Gray V. Learning to look: Pulsar partners with Manchester School of Arts to understand the difference between how machines and humans look at images. Pulsar. 2018. URL: https://www.pulsarplatform.com/blog/2018/learning-to-look-pulsar-partners-with-manchester-school-of-arts-to-understand-the-difference-between-how-machines-and-humans-look-at-images/ [accessed 2021-01-25]
131. Gaspar R, Pedro C, Panagiotopoulos P, Seibt B. Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. Comput Hum Behav 2016 Mar;56:179-191. [doi: 10.1016/j.chb.2015.11.040]
132. Thelwall M. Social media analytics for YouTube comments: potential and limitations. Int J Soc Res Method 2017 Sep 21;21(3):303-316. [doi: 10.1080/13645579.2017.1381821]
133. Majumder N, Hazarika D, Gelbukh A, Cambria E, Poria S. Multimodal sentiment analysis using hierarchical fusion with context modeling. Knowl Based Syst 2018 Dec;161:124-133. [doi: 10.1016/j.knosys.2018.07.041]
134. Amarasekara I, Grant WJ. Exploring the YouTube science communication gender gap: A sentiment analysis. Public Underst Sci 2019 Jan;28(1):68-84. [doi: 10.1177/0963662518786654] [Medline: 29974815]
135. Lama Y, Chen T, Dredze M, Jamison A, Quinn SC, Broniatowski DA. Discordance between human papillomavirus Twitter images and disparities in human papillomavirus risk and disease in the United States: mixed-methods analysis. J Med Internet Res 2018 Sep 14;20(9):e10244 [FREE Full text] [doi: 10.2196/10244] [Medline: 30217792]
136. Wong JC. Facebook's Zuckerberg announces privacy overhaul: We don't have the strongest reputation. The Guardian. 2019 Apr 30. URL: https://www.theguardian.com/technology/2019/apr/30/facebook-f8-conference-privacy-mark-zuckerberg [accessed 2021-01-25]
137. Rieder B. Studying Facebook via data extraction: the Netvizz application. 2013 Presented at: 5th annual ACM web science conference; May 2013; Paris, France. [doi: 10.1145/2464464.2464475]
138. Wilson M. The tech giant fighting anti-vaxxers isn't Twitter or Facebook. It's Pinterest. Fast Company. 2019. URL: https://www.fastcompany.com/90310970/the-tech-giant-fighting-anti-vaxxers-isnt-twitter-or-facebook-its-pinterest [accessed 2021-01-25]
139. Wong JC. How Facebook and YouTube help spread anti-vaxxer propaganda. The Guardian. 2019 Feb 01. URL: https://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media [accessed 2021-01-25]
140. YouTube takes ads off 'anti-vax' video channels. BBC. 2019 Feb 25. URL: https://www.bbc.com/news/technology-47357252 [accessed 2021-01-25]
141. Allen C. Extreme far right: 'pick'n'mix' ideologies and direct messaging online make for deadly new combination. The Conversation. 2019 Mar 19. URL: https://theconversation.com/extreme-far-right-picknmix-ideologies-and-direct-messaging-online-make-for-deadly-new-combination-112331 [accessed 2021-01-25]

Abbreviations

API: application programming interface
ECDC: European Centre for Disease Prevention and Control
GDPR: General Data Protection Regulation
HPV: human papillomavirus
VCP: Vaccine Confidence Project

XSL•FO
RenderX

Edited by G Eysenbach; submitted 21.11.19; peer-reviewed by A Budenz, J Jiang; comments to author 17.09.20; revised version received 05.11.20; accepted 17.12.20; published 08.02.21.

Please cite as:
Karafillakis E, Martin S, Simas C, Olsson K, Takacs J, Dada S, Larson HJ
Methods for Social Media Monitoring Related to Vaccination: Systematic Scoping Review
JMIR Public Health Surveill 2021;7(2):e17149
URL: http://publichealth.jmir.org/2021/2/e17149/
doi:10.2196/17149
PMID:33555267

©Emilie Karafillakis, Sam Martin, Clarissa Simas, Kate Olsson, Judit Takacs, Sara Dada, Heidi Jane Larson. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 08.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Effectiveness of a Web-Based Program (POP4Teens) to Prevent Prescription Opioid Misuse Among Adolescents: Randomized Controlled Trial

Lisa A Marsch[1,2,3], PhD; Sarah K Moore[1], MSW, PhD; Michael Grabinski[2,3], BA; Sarah Y Bessen[4], MPH; Jacob Borodovsky[1], PhD; Emily Scherer[5], PhD

[1]Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States

[2]HealthSim Inc, Hanover, NH, United States

[3]Square2 Systems, Inc, Hanover, NH, United States

[4]Geisel School of Medicine at Dartmouth College, Hanover, NH, United States

[5]Department of Biomedical Data Science, Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States

Corresponding Author:
Sarah K Moore, MSW, PhD
Center for Technology and Behavioral Health
Geisel School of Medicine at Dartmouth College
46 Centerra Pkwy
Lebanon, NH, 03766
United States
Phone: 1 9176789280
Email: sarah.k.moore@dartmouth.edu

Abstract

Background: Prescription opioid (PO) use is common among adolescents in the United States. Despite recent declines from unprecedented peaks in adolescent PO use (eg, in 2012-2013), there is seemingly paradoxical evidence that PO-related consequences (eg, opioid use disorder and overdoses) are increasing. These trends and their possible consequences emphasize the importance of prevention efforts targeting PO misuse. To our knowledge, we have developed the first interactive web-based program (POP4Teens [P4T]) focused specifically on the prevention of PO misuse among adolescents.

Objective: This study aimed to evaluate the effectiveness of P4T, a web-based program designed to prevent adolescent PO misuse, in comparison with JustThinkTwice (JTT), an active control website, on PO-related attitudes, knowledge, risk perception, and intentions to use.

Methods: We conducted a web-based randomized controlled trial in 2018. A total of 406 adolescents (aged 12-17 years) were randomly assigned to either P4T or JTT. The outcome variables were attitudes, knowledge, and risk perceptions associated with PO misuse, intentions to use POs, and program feedback. Data were collected at baseline and at 1, 3, and 6 months.

Results: Both programs resulted in significant and sustained improvements in intention to use POs, increased perceived risk, impacted expectancies consistent with prevention, and improved PO refusal skills. P4T produced significantly greater increases in PO-related knowledge than JTT did, and it was reportedly easier to use and more liked. Baseline scores for youth reporting past-year medical use of POs, friends who engage in nonmedical use of POs, and/or poor mental health underscored their at-risk status compared with youth from the other groups.

Conclusions: P4T positively impacted all study variables that are known to prevent PO misuse among teens. Moreover, its web-based nature simplifies the dissemination and implementation of this novel tool designed to help meet the challenges of the evolving national opioid crisis.

Trial Registration: ClinicalTrials.gov NCT02737696; https://clinicaltrials.gov/ct2/show/NCT02737696

(JMIR Public Health Surveill 2021;7(2):e18487) doi:10.2196/18487

KEYWORDS

opioids; prevention and control; adolescent; randomized controlled trial; internet

XSL•FO
RenderX

Introduction

Background

Prescription opioid (PO) use is common among adolescents in the United States. Over the last 4 decades, the lifetime prevalence of medical (13%-20%) and nonmedical use of POs (NMUPO; 7%-13%) among high school seniors (approximately 17-18 years of age) has fluctuated yet both remain high [1]. One in every 100 adolescents aged between 12 and 17 years reported current NMUPO in 2017 [2], and this estimate may underreport the problem due to potential social desirability bias associated with self-report measures, particularly among youth [3]. Predictors of NMUPO and opioid use disorder (OUD) include early initiation of medical opioid use, and the risks of OUD are inversely correlated with age of onset [4-6]. Adolescents with mental health problems are also more likely than those without such problems to use diverted POs [7]. Adolescents with NMUPO overwhelmingly report accessing these medications from friends and family and having friends or family who are prescribed opioids increases the risk for misuse and OUD [8-10].

Compared with nonusers, adolescents who engage in NMUPO have increased odds of using alcohol and other drugs, engaging in criminal behaviors, using health care (eg, receiving emergency or inpatient treatment), receiving treatment services for mental health, and reporting poorer health [11-15]. The risk of transition to heroin use is greatest among those that begin NMUPO early in adolescence [16-19], and fatal and nonfatal overdose rates are higher for these youth compared with their peers who do not engage in NMUPO [20]. Despite recent declines from unprecedented peaks in adolescent PO use (eg, in 2012-2013), there is seemingly paradoxical evidence that PO-related consequences (eg, OUD, mortality, overdoses) are increasing [21,22]. These increases are likely associated with evidence that youth with OUD rarely receive the treatment they need [23].

These trends and their possible consequences emphasize the importance of prevention efforts targeting PO misuse. To our knowledge, we have developed the first interactive web-based program (POP4Teens [P4T]) focused specifically on the prevention of PO misuse among adolescents. This program extends our previous work to develop and evaluate science-based, computer-delivered substance misuse prevention programs for youth, including a prototype of this program we previously developed and evaluated [24-26]. The scientific evidence underpinning P4T's program content comes from research on identified risk factors for PO use among youth [27], computer-delivered interventions [28], and computer-assisted instruction technology [29].

The P4T program is unique compared with other programs, as it is the first of its kind to incorporate knowledge about risk factors for PO misuse. PO misuse is unique compared with misuse of many other drugs: the threshold to access is low, potentiating the behavior (ie, usually obtained from friends and family and at no cost) [30,31]. Relatedly, youth who are prescribed opioids often divert their medication [32,33]; the perceived risks of misuse are significantly attenuated compared with other drugs, as they are dispensed by trusted medical providers [34,35]; youth perceive that parental consequences of misuse will be minor if they are *caught* with POs relative to other drugs [36].

In addition to a focus on the specific risk factors associated with POs, the interactive, activity-oriented P4T program is rooted in the prevention science literature, which has demonstrated that effective substance abuse prevention programs educate youth about the mechanism of action by which a substance produces its pharmacological effect and additionally include various types of skills training (eg, to resist peer, family, media, and other social influences known to promote drug use, and/or general skills training around social competency and coping with stressful life situations) [37,38]. Skills training increases protective factors, reduces risk factors, and produces reductions in drug use among youth [28]. Notably, programs based on social influence and general skills training models have been shown to be effective when delivered via interactive, activity-oriented programs and not traditional, didactic instructional techniques [39].

The P4T program is also unique as it is web-delivered and mobile-friendly and therefore available for use in a variety of settings. A recent scoping review of consumer health information technology (CHIT) in the prevention of substance use [40] underscores the effectiveness of *user-centered, interactive web-mediated apps* designed to *improve information access and exchange, decision making*, and *facilitate behavior changes that promote health and well-being*. Traditional means of reaching youth are hamstrung by costs associated with teacher or clinician-delivered interventions as well as manifold issues impeding intervention delivery fidelity. As of 2018, 94% of teens reported going on web daily and almost 95% reported mobile device usage and/or ownership [41]. Compared with traditional delivery methods, CHIT interventions delivered via digital platforms boast benefits of greater appeal to youth [28], reduced costs [36], standardized content delivery [36], potential for delivery across different settings (eg, on the web, home, community organizations, and health care providers' offices) [37], and minimized training and/or clinical staff burden [42].

Published scientific literature on digital platforms for interventions targeting adolescent substance misuse is burgeoning [43,44]. Such programs have demonstrated effectiveness in 3 contexts: primary care [45], schools [44], and homes [46-48]. They target the prevention of different classes of drugs: alcohol [43,44], cannabis [45,49], cigarettes [50] and generally consist of interactive web-based activities that function to increase drug-related knowledge and shape user attitudes and normative beliefs about substance use in ways that promote abstinence or delayed onset of use [51]. The web-based context for the P4T program focuses on a new class of drugs (POs) previously unexplored in a digital format and additionally employs unique informational technologies that are effective in promoting relevant knowledge and skills. The P4T content uses information technologies to teach content to mastery via interactive and computer-assisted instruction (engaged when youth access quizzes) and enables exploration of the information and skills being taught in a video-based simulated environment [29,52,53].

XSL•FO
RenderX

Objectives

The randomized controlled trial described herein focused on the prevention of PO misuse among youth. This web-based trial is designed to evaluate the effectiveness of the web-based PO misuse prevention program P4T, developed by our research group, compared with an active control website JustThinkTwice (Drug Enforcement Administration [DEA]) to impact attitudes about knowledge and perceptions of risk associated with misuse of POs as well as intentions to use and actual use of POs. Details of the development and formative evaluation of P4T have been published elsewhere [24].

Methods

Design

Participants were assigned at random using a nonlinear random number generator based on an AES block cipher in counter mode to either the experimental, web-based P4T program or the DEA's JustThinkTwice (JTT) website (the active control condition). The study team was not blinded to randomization. However, due to the web-based nature of the trial (eg, recruitment, randomization, data collection) and limited-to-nonexistent research team contact with participants following randomization, bias was minimized.

The CONSORT (Consolidated Standards of Reporting Trials) checklist has been provided in Multimedia Appendix 1.

Participants

Adolescent participants (12-17 years old) were recruited for study participation through Facebook and Google AdWords. This age range was selected to capture youth before a notable increase in risk of hazardous consequences of nonmedical PO use observed from age 12 years to 14 years [6,54], through the end of high school or before college. To be eligible, youth needed to be willing to use the study website to complete assessments and participate in the interventions, have access to a computer, a tablet, or a mobile device with an internet connection, and be able to hear audio.

From a user perspective, once an interested youth clicked on an ad, they were directed to the study *landing page*, which provided basic information about the study aims and clickable icons to *learn more*. Upon clicking on *learn more*, interested youth were offered the opportunity to receive a call from a research assistant by entering basic contact information (ie, name, email, parent's name, email, phone number, and best time to be reached) that the research assistant used to make contact, answer additional questions, collect zip code data, email the consent form, and enable backend, automatic random assignment to a study condition. Once consent was established, the research assistant emailed the baseline assessment, which, once completed (as evidenced by a timestamp record on the program's *dashboard* menu of participant activities), prompted the research assistant to send the first electronic gift card via email to the participant as compensation for their time spent in completing the assessment. Youth received US $20 for baseline, US $30 for 1-, 3-, and 6-month follow-ups and US $30 as bonus if all assessments were completed. Participants completed all assessments on the web in a secure manner, which allowed for greater anonymity and has been shown to result in more valid data compared with data collected via in-person interviews [55,56]. Delivering assessments via computer also ensured that questionnaire administration was more consistent, data processing was eliminated, and lag time between data collection and analyses was reduced. All procedures were approved by the Dartmouth College Committee for the Protection of Human Subjects.

Procedures

The experimental program P4T is a newly developed intervention [24] designed to increase knowledge about POs and their risks and address social influences on PO use, such as developing skills to refuse offers or requests for POs. The program comprised 8 primary modules, and the information and skills emphasized in each module are delivered through 4 components: a storytelling component, an information lesson component, a skill-based video, and a quiz. The 8 module topics are (1) what are opioids? (2) misconceptions that opioids are safe and nonaddictive, (3) misconceptions that use of POs without a prescription is legal, (4) risks of PO misuse, (5) nonmedication alternatives for pain management, (6) refusing offers to misuse POs, (7) refusing requests by others for a PO prescribed to you, and (8) how to know if you or someone you may know may be addicted.

In each module, users were introduced to a different character based on real stories of youth in treatment for OUD. By clicking on the character icon, users were brought to an introductory screen displaying an exemplary quote from the youth's story regarding PO use. Here, users could choose to listen to a recording of the story, read the story transcript, or view story highlights.

The program then allowed users to move between the introductory storytelling screen and the remaining 3 components (lesson, video, and quiz) personalizing the navigational flow and contributing to the interactive nature of P4T. The lessons detailed the scientific underpinnings of the story and delivered information through an interactive carousel format. The videos provided participants with concrete methods to approach social interactions and dilemmas involving PO misuse. The videos employed an *action-reaction* model in which a behavioral scenario was shown, followed by one or more characters reflecting on the situation. Finally, the quizzes allowed program users to test their understanding of the module content. Youth were given a variable amount of time to respond to quiz questions, which are presented multiple times and worded slightly differently to increase fluency and mastery, depending on how fluent youth are with the material, for example, if they are repeatedly getting a question incorrect, they will be allowed more time to respond to the next time they see a version of the question. Thus, the quizzes further contributed to the interactive nature of the program. Participants were asked to complete the 8 modules over the course of 4 weeks (approximately 2 modules per week).

The active, educational control website JTT is a web-based educational initiative supported by the DEA, which provides science-based information about teen drug abuse prevention. The intended purpose of the website, according to a DEA press

XSL•FO
RenderX

release on the occasion of the launch of the teen website reads that it is "a website especially for our teenagers that provides young people with straightforward information on the consequences of drug use and trafficking, including health, social, legal, and international consequences. The site does not inundate the user with DEA messages; rather the facts come straight from leading physicians, scientists, and legal experts." [57]

The website is not strictly dedicated to POs, although opioids are well represented. People navigating the website will find drug information, news and media, true stories of youth who have suffered due to drug use, information about consequences of drug use, drug use facts and statistics, videos, and a brief quiz. Randomized youth were asked to use this website for 30 min twice per week for 4 weeks.

Outcome Measures

Behavioral assessments accessible by links emailed to participants were completed at baseline and follow-up time points (1, 3, and 6 months). At baseline, participants completed a form measuring basic demographic and substance use history, for example, age, gender, race, ethnicity, grade in school, average school grades, lifetime prescriptions for opioids, past-year mental health status, substance use in general, PO use among participants and their friends, and past participation in PO abuse prevention programs. At the 1-month time point only, participants completed the program feedback survey. Given the pervasive misconceptions that many youths have in their understanding of the risks associated with PO misuse [58], we primarily focused on analyses of knowledge about PO misuse as well as measures that are highly predictive of future use (attitudes, intentions, skills, and perceived risk).

Primary

Attitudes Toward PO Misuse

Participants were asked about their expectancies associated with PO misuse. Positive expectancies (eg, reduce boredom) and negative expectancies (eg, get in trouble with parents) about such use were evaluated using a 5-point measure ranging from *strongly disagree* to *strongly agree*. Mounting evidence has suggested that outcome expectancies (expected or anticipated behavioral, cognitive, and emotional harms and benefits [59]) associated with substance use play a role in the initiation and progression of drug use [60]. We modeled these questions after expectancy items developed by researchers examining potential physical, social, and psychological consequence expectancies associated with alcohol consumption among youth [61]. Changes in expectancies appear to be an important target in substance use prevention across alcohol, tobacco, and marijuana [62].

Perceived Risk of PO Misuse

Participants were asked 2 separate questions about how much they think they risk harming themselves (1) physically and (2) in other ways if they try POs for nonmedical purposes (eg, 5-point scale ranging from *no risk* to *great risk*) [59]. These questions are used annually in the Monitoring the Future Study [63], and the reliability and validity of survey questions is extensively discussed in an *occasional paper* detailing design and procedures [64]. Other studies have also included questions modeled after survey questions [65]. Research has consistently shown that alcohol and cannabis use is less likely among youth who perceive those substances and/or substance use behaviors as risky for the user [66,67], and thus, perceived harmfulness is potentially central to early prevention strategies [68].

Knowledge About PO Abuse Prevention

Participants were asked to complete a 15-item, multiple-choice measure assessing knowledge about effective PO abuse prevention (eg, how opioids work in the body). Knowledge test questions were generated by the research team based on scientific literature associated with the 8 module topics in the P4T program (eg, nonmedication alternatives for pain management).

Skill Acquisition

Youth were asked to pretend a friend offered them a PO and estimate how hard it would be to refuse the offer (5-point scale of *very easy* to *very hard*) and their ability to refuse the offer (5-point scale of *definitely would* to *definitely would not*). The same format was used to ask youth to pretend a friend asked them to share or sell POs prescribed to them [69]. These items were modeled after drug resistance scales used in previous studies with youth of the same age [70,71].

Secondary

Assessment of Usage Data

We quantified usage statistics of the web-based program by assessing the extent of module completion.

Behavioral Intentions to Use Opioids Without a Prescription

Youth were asked about their intentions to use opioids without a prescription *even once or twice over the next 12 months* using a 5-point scale ranging from *definitely not* to *definitely will*. Those who provided an affirmative response indicating any likelihood were also asked about the likelihood of using opioids without a prescription nearly every month for the next 12 months. This 2-item composite (α=.73) was modeled after a measure used in a study [72] that found changes in marijuana expectations among a large cohort of adolescents associated with subsequent future marijuana initiations.

Feedback Survey

The feedback survey was a brief 15-item tool that included 11 visual analog scale items (Multimedia Appendix 2) and 4 open-ended response items (ie, What youth liked best? Least? Suggestions? Anything else to add?). Possible values for visual analog scale scores ranged from 0 to 100 mm and were anchored by the variable of interest probed in the item (eg, how interesting was the section of the program you just completed? score range: 0=not interesting to 100=very interesting).

Data Analysis

Although random assignment was designed to account for baseline differences, groups were compared for differences in demographics and other characteristics using two-tailed *t* tests, Wilcoxon rank sum tests for nonnormal variables, and chi-square tests for categorical measures. Program feedback (collected at 1 month) was also compared via *t* tests. Primary analyses included all participants randomized to a study condition

XSL•FO
RenderX

independent of early dropout, consistent with an intent-to-treat approach to clinical trials [73]. Analyses were performed using SAS statistical software (type I error=.05). Outcome measures were evaluated in separate analyses. Data from primary outcomes were evaluated using mixed effect models, which allow for nonindependent data within individuals and are robust to incomplete data. All participants with any baseline or follow-up assessments were included in the outcome analyses. In addition to the study condition indicator, these models included, in the fixed effects portion of the model, a main effect for time as well as a 2-way interaction between study condition and time. Before evaluating the comparative effectiveness of P4T, each program's effectiveness in impacting primary outcome variables without comparison is reviewed, as neither has previously been evaluated. The key estimate from this model is the 2-way interaction effect, which indicates whether changes in outcome over time are different for the 2 conditions.

Sample Size Calculation

In our previous work [74], a computer-based prevention intervention, relative to an efficacious life skills control condition, had a large impact on knowledge (Cohen d=0.84), a moderate impact on drug refusal (Cohen d=0.43), a small impact on drug use intentions and attitudes (Cohen d=0.20), and a moderate impact on alcohol use (Cohen d=0.37). Schwinn et al [69] found a comparable effect size for reducing substance use when comparing an internet-based prevention program with an assessment-only control group (Cohen d=0.40). The planned sample size of 400 participants was based on having sufficient power to detect group differences in all outcomes, assuming 80% retention at the 6-month follow-up [73,75,76]. Assuming baseline assessments account for at least 9% of the variance in corresponding outcomes (r=0.30), power is estimated to be greater than 99% using α=.05 (2 sided) for detecting a prevention group effect of Cohen d=0.84 on the outcome measure of knowledge. Power was estimated to be 92% using α=.05 (2 sided) for detecting a prevention group effect of Cohen d=0.37 on the outcome measure of substance use and 98% using α=.05 (2-sided) for detecting a prevention group effect of Cohen d=0.43 on the outcome measure of drug refusal. For the intention to use outcomes, we determined the sample size needed to detect Cohen d=0.30, which is slightly higher than the effect observed in previous work [74], to reflect the use of an active control group in the planned study instead of a comparison condition of demonstrated effectiveness. The power to detect that effect was 80% when the sample size was 320 (after 20% attrition) at the 6-month follow-up.

Results

Overview

A nationwide sample of youth from 41 of the 50 US states was recruited between June 2017 and February 2018, and data collection was completed in September 2018. Of the 1470 youth and/or parent/guardian who signed up to gather more information about participation, 406 were consented and randomly assigned to one of the 2 websites (Figures 1 and 2).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.

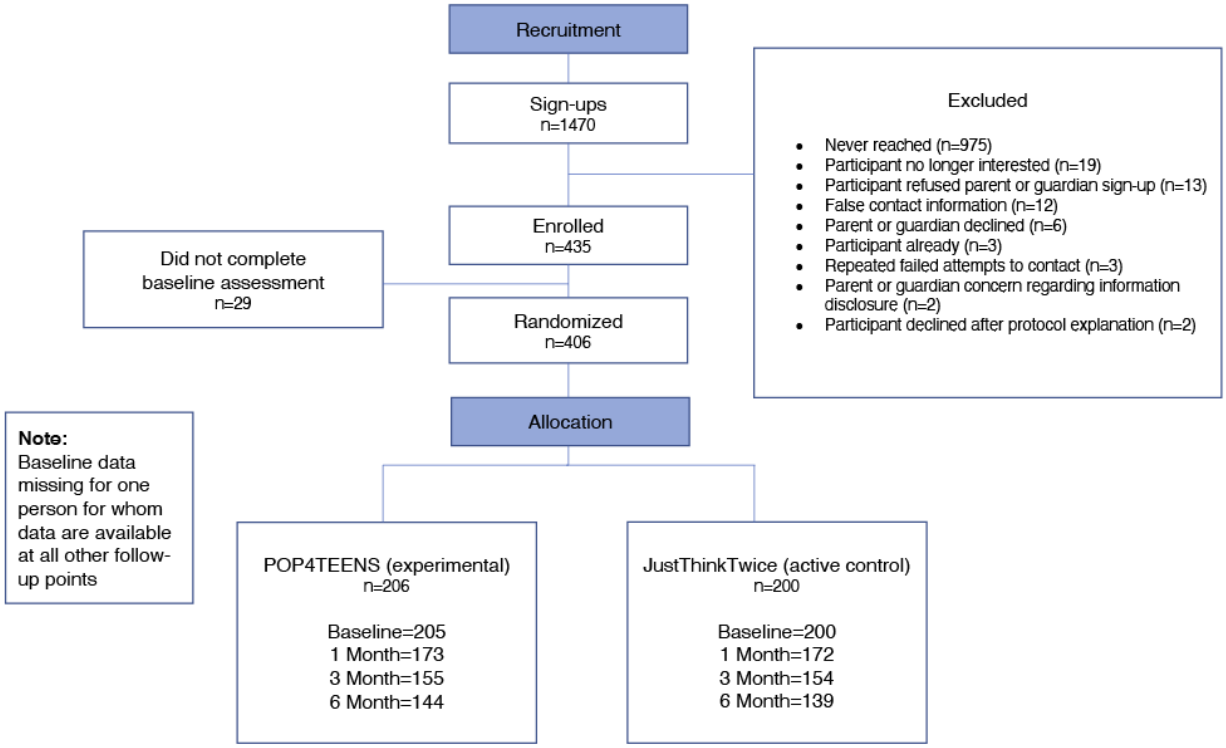
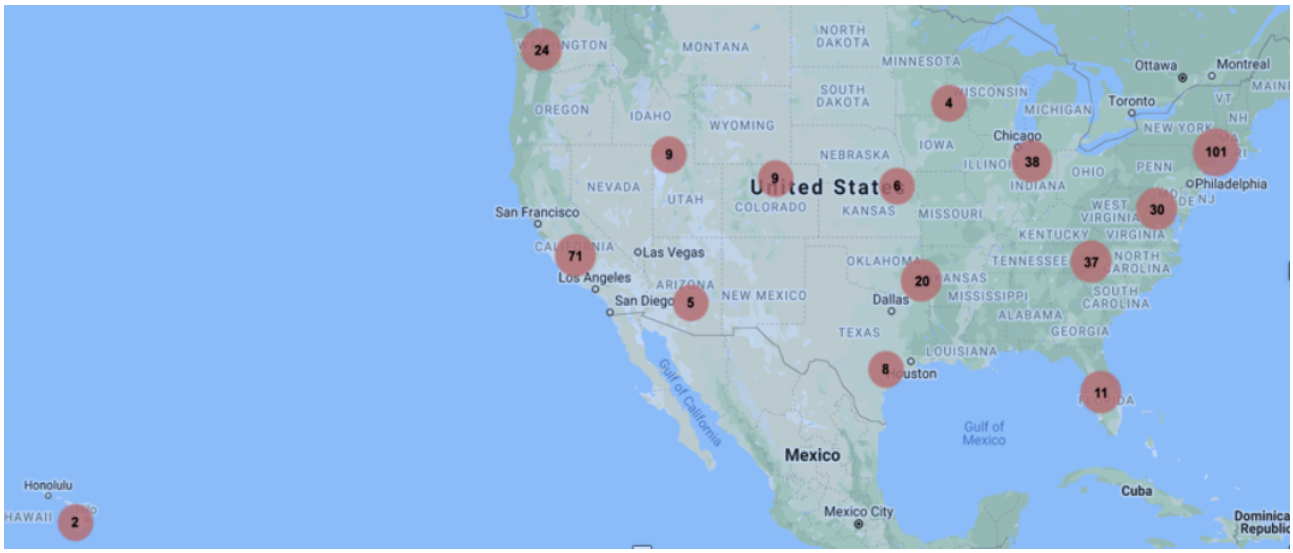


Figure 2. Geographic locations of participants. Zip code data for 379 of 406 participants (missing zip code data: n=27). In total, 41 of 50 states had at least one participant. Numbers in red circles are those of participants in a geographic area and may represent more than one state's data. Alaska (n=4) is not shown on the map.



Participant Characteristics

The sample was primarily female (285/405, 70.4%), with a mean age of 15.8 (SD 1.2) years, and high achieving with 70.9% (287/405) reporting A's for average school grade; 92.8% (376/405) reported good-to-excellent physical health, and approximately 73.6% (298/405) reported good-to-excellent mental health. More than a quarter of the sample reported lifetime exposure to medically prescribed opioids. Randomization was largely successful, as groups did not differ significantly on any baseline variable except for race: Asian

XSL•FO
RenderX

(P4T/JTT: 21%/13.5%); Black/African American (P4T/JTT: 5.9%/12%); and White (P4T/JTT: 64%/67.5%; P=.03). For additional participant characteristics, see Table 1.

XSL•FO
RenderX

Table 1. Demographic and baseline characteristics.

Demographic and baseline characteristics	Overall (N=405)	P4T[a] (n=205)	JTT[b] (n=200)	P value
Age (years), mean (SE)	15.8 (1.2)	15.7 (1.2)	15.8 (1.2)	.80
Gender, n (%)				.22
Female	285 (70.4)	148 (72.2)	137 (68.5)	
Male	113 (27.9)	54 (26.3)	59 (29.5)	
Transgender	5 (1.2)	1 (0.5)	4 (2)	
No answer	2 (0.5)	2 (0.9)	0 (0)	
Race, n (%)				.03
American Indian or Alaskan Native	3 (0.7)	1 (0.5)	2 (1)	
Asian	70 (17.3)	43 (20.9)	27 (13.5)	
Black or African American	36 (8.8)	12 (5.9)	24 (12)	
Pacific Islander	4 (0.9)	4 (1.9)	0 (0)	
White	266 (65.7)	131 (63.9)	135 (67.5)	
Other	26 (6.4)	14 (6.8)	12 (6)	
Ethnicity, n (%)				.20
Latino or Hispanic	49 (12.1)	29 (14.1)	20 (10)	
Non-Latino or non-Hispanic	356 (87.9)	176 (85.9)	180 (90)	
Grade, n (%)				.27
6th-9th (ages 13-15 years)	64 (15.8)	31 (15.1)	33 (16.5)	
10th-12th (ages 16-18 years)	324 (80)	169 (82.4)	155 (77.5)	
Not in school or other	17 (4.2)	5 (2.4)	12 (6)	
Average school grade, n (%)				.88
A (90-100)	287 (70.9)	144 (70.2)	143 (71.5)	
B (80-89)	100 (24.7)	52 (25.4)	48 (24)	
C (70-79)	15 (3.7)	7 (3.4)	8 (4)	
D (60-69)	2 (0.5)	1 (0.5)	1 (0.5)	
F (<60)	1 (0.2)	1 (0.5)	0 (0)	
Poor or fair mental health, n (%)				.35
No	298 (73.6)	155 (75.6)	143 (71.5)	
Yes	107 (26.4)	50 (24.4)	57 (28.5)	
Physical health, n (%)				.19
Excellent	93 (22.9)	50 (24.4)	43 (21.5)	
Very good	177 (43.7)	85 (41.5)	92 (46)	
Good	106 (26.2)	58 (28.3)	48 (24)	
Fair	24 (5.9)	8 (3.9)	16 (8)	
Poor	5 (1.2)	4 (1.9)	1 (0.5)	
Ever prescribed opioids, n (%)				.41
Yes	106 (26.2)	50 (24.4)	56 (28)	
No	299 (73.8)	155 (75.6)	144 (72)	
Friends who use prescription opioids, n (%)				.49
Yes	68 (16.8)	37 (18)	31 (15.5)	
No	337 (83.2)	168 (81.9)	169 (84.5)	
Participated in a prescription opioid prevention program, n (%)				.34

XSL•FO
RenderX

Demographic and baseline characteristics	Overall (N=405)	P4T[a] (n=205)	JTT[b] (n=200)	P value
Yes	11 (2.7)	4 (1.9)	7 (3.5)	
No	394 (97.3)	201 (98)	193 (96.5)	

[a]P4T: POP4Teens.

[b]JTT: JustThinkTwice.

Program Effectiveness (Within-Condition Effects)

Drug Use Outcome Expectancies (Attitudes)

The P4T program produced significant improvements from baseline across the follow-up time points (1, 3, and 6 months) for all positive expectancies of POs. The JTT program produced a similar pattern of results except for *reduce boredom* (baseline-1 month: P=.11) and *weight loss* (baseline-6 months: P=.06). The P4T program produced significant improvements from baseline across all follow-up time points for most negative expectancies. The exceptions were a pattern of nonsignificant results across all time points for *get in trouble with parents* and nonsignificance of *feel sleepy* (baseline-1 month: P=.41). The JTT program produced significant improvements from baseline across all follow-up time points for a few negative expectancies (eg, feel sleepy or overdose). The exceptions were a pattern of significant differences observed at the 1 and 6 month follow-ups, and nonsignificant differences at the 3-month follow-up (eg, spend too much money, feel sick, or pass out). The same pattern of nonsignificant differences across all time points for *get in trouble with parents* held for the JTT program. Finally, there were significant differences from baseline to 1 month that were not sustained at subsequent follow-ups for *do poorly in school*. For greater detail on within-condition effects, see Multimedia Appendix 2 (means) and Multimedia Appendix 3 (mixed models).

Perceived Risks

The P4T program significantly increased *perceived physical risks* from baseline to 1 month and baseline to 3 months; however, significant increases were not maintained at the third follow-up (P=.14). The program also significantly increased *perceived other risks* from baseline to 3 months; however, this effect was not observed for baseline to 1 month (P=.06) or baseline to 6 months (P=.06). The JTT program produced significant and sustained increases in perceived physical and other risks from baseline across all time points.

Knowledge

The P4T program significantly increased knowledge about POs from baseline to all time points. The JTT program produced no significant changes in knowledge at any follow-up time point.

Skills

The P4T program produced significant and sustained changes in the desired direction across all skill sets (ie, assessments of difficulty and ability to refuse offers and requests) and time points except for *ability to refuse an offer to misuse POs* at 3 months (P=.06). The JTT program produced significant and sustained changes in the desired direction across all skill sets except for *perceived difficulty to refuse an offer to misuse POs* at 1 month (P=.14).

Intentions

Both programs produced significant and sustained decreased intentions to use POs in the next 12 months across all time points.

Assessment of Usage

Participants in the P4T group completed an average of 6.43 (SD 2.98) of 8 possible modules and a total of 80.79% (1312/1624) possible modules. We were unable to calculate a similar metric in the control sample (JTT), as we did not control that website. However, we were able to calculate the average number of log-ins on the control site (a conservative estimate as participants may have chosen to access the site without using the portal on the study website [mean 4.06, SD 3.89]).

Effectiveness of P4T Compared With JTT

Drug Use Outcome Expectancies (Attitudes)

There were no significant differences between groups for any positive expectancies at any time points. There were no significant differences between groups across the negative expectancies observed from baseline to 1 month. From baseline to 3 months, there were statistically significant increases in negative expectancies of POs favoring P4T: do poorly in school (P=.01), spend money (P=.01), feel sick (P=.03), and pass out (P=.004). From baseline to 6 months, the effect for do poorly in school was maintained (P=.01).

Perceived Risks

P4T produced a greater reduction (from baseline to 1 month) in the odds of perceiving a low *other* risk from POs compared with JTT (P=.03). There were no other significant differences between groups at any other time point for measures of risk.

Knowledge

There was a significant difference between groups in change from baseline for knowledge gains favoring P4T at the first 2 follow-up time points: baseline to 1 month (P<.001) and baseline to 3 months (P=.001). The effect somewhat diminished at the 6-month follow-up (P=.06).

Skills

There were no significant differences between groups across skills *except* for *difficulty refusing an offer to misuse* between baseline and 1 month (P=.05) favoring P4T. Differences were not maintained at later time points.

Intentions

There were no significant differences between groups in change from baseline to any of the 3 follow-up time points for *intention*

XSL•FO
RenderX

to use POs: baseline to 1 month (P=.68), baseline to 3 months (P=.99), and baseline to 6 months (P=.75).

Feedback Surveys

With one exception, evaluations of the P4T and JTT programs were generally positive, with most mean scores at 7 or higher (scale 1-10) and not significantly different. However, the P4T program was rated as significantly easier to use than the JTT program (P<.001). In addition, the P4T youth liked using laptops, phones, and/or tablets significantly more to access the program than JTT youth (P<.001). No mean scores fell below 6, except for the applicability of program content to participants' lives (Multimedia Appendices 2 [means] and 3 [mixed models]).

Four P4T program dimensions that youth liked best (in order of preference) were stories, ease of use, quizzes, and video. The stories were preferred for how they provided *real-life, relatable* examples and helped make *the opioid epidemic just seem more real.* Youth liked P4T because of how easy it was to *access* and *navigate.* They also reported enjoying the quizzes for how *the material began to make sense,* and the repetitive nature of the quiz questions helped youth *master* the information. Finally, some youth appreciated that the videos were upfront with being *a little bit corny at times* and were a *great educational tool.* The functionality of quizzes and videos surfaced among program dimensions was the least. Specifically, the timed and repetitive nature of the quizzes surfaced for critique:

> *The timed portion of the quizzes created unnecessary stress, increasing speed of the questions…lead to testing for your reaction time rather than processing information, the questions seemed very repetitive.*

Videos were least liked by some because they were found to be *dull.*

Baseline Differences: Characterizing the Sample in Terms of At-Risk Subgroups

Drug Use Outcome Expectancies (Attitudes)

Scores for youth reporting past-year medical use of POs, having friends who engage in NMUPO, and having poor mental health illuminated their at-risk status at baseline compared with youth who were not in those groups and therefore warrant attention. For negative expectancies, those with poor mental health agreed more strongly than those without poor mental health that overdose is likely when one engages in NMUPO. There was a pattern across those with a history of past year medical use of POs and having friends who engage in NMUPO that NMUPO would likely result in trouble with parents and doing poorly in school compared with those without such histories. In addition, those with a history of past year medical use of POs agreed more strongly that NMUPO would make one feel sleepy. For positive expectancies from POs, there is a clear pattern for both those with histories of poor mental health and having friends who engage in NMUPO demonstrating greater agreement with almost all such items compared with those without these histories, indicating an increased risk for these groups of youth. The only positive expectancy score for those with past year medical use of POs histories higher than their comparators was for reducing physical pain.

Perceived Risks, Knowledge, Skills, and Intentions

There were no differences between groups at risk (poor mental health, past year medical use of POs, and having friends who engage in NMUPO) and those not at risk in these categories for perceived risks with the lone exception among those with friends who engage in NMUPO who demonstrated slightly weaker agreement than those without friends who engage in NMUPO that there may be *other* risks (beyond physical), if one NMUPO. There were also no differences between youth in these 3 at-risk groups and their counterparts without such histories on knowledge about POs. Similar to the pattern of findings for positive expectancies, those with poor mental health and friends who engage in NMUPO scored at greater risk with respect to skills compared with participants without such histories. There were no differences between those with past year medical use of POs and those without such histories. Finally, all 3 subgroups reported stronger intentions to NMUPO in the next 12 months compared with youth not in those groups (Table 2).

XSL•FO
RenderX

Table 2. Primary and secondary measures of at-risk youth at baseline (mean).

Baseline value[a]	Poor mental health			Medical prescription opioid use			Friend who NMUPO[b]		
	Yes, mean (SD)	No, mean (SD)	P value	Yes, mean (SD)	No, mean (SD)	P value	Yes, mean (SD)	No, mean (SD)	P value
Intention to use prescription opioids	1.63 (0.77)	1.43 (0.69)	.02	1.62 (0.76)	1.44 (0.70)	.03	1.76 (0.92)	1.43 (0.66)	.01
Negative expectancies									
Spend too much (US $)	3.92 (1.07)	3.81 (1.11)	.41	3.73 (1.18)	3.88 (1.07)	.23	3.85 (1.20)	3.84 (1.08)	.93
Overdose	3.09 (1.13)	2.77 (1.11)	.01	2.87 (1.16)	2.86 (1.11)	.92	2.79 (1.10)	2.87 (1.13)	.60
Trouble with parents	4.58 (0.82)	4.53 (0.95)	.58	4.37 (0.99)	4.60 (0.88)	.03	4.19 (1.14)	4.61 (0.85)	.01
Pass out	3.44 (0.93)	3.30 (0.93)	.18	3.22 (0.98)	3.38 (0.92)	.15	3.24 (0.95)	3.35 (0.93)	.35
Do poorly in school	3.93 (0.99)	3.96 (1.09)	.79	3.70 (1.11)	4.04 (1.03)	.01	3.60 (1.15)	4.02 (1.03)	.01
Feel sick	3.79 (0.94)	3.80 (0.90)	.95	3.80 (0.97)	3.80 (0.88)	.98	3.75 (1.00)	3.81 (0.89)	.65
Feel sleepy	3.85 (0.89)	3.75 (0.90)	.34	3.95 (0.96)	3.72 (0.87)	.03	3.79 (0.99)	3.78 (0.88)	.89
Positive expectancies									
Reduce anxiety	3.12 (1.17)	2.73 (1.05)	.003	2.86 (1.12)	2.82 (1.08)	.77	3.22 (1.08)	2.75 (1.08)	.003
Reduce boredom	2.45 (1.14)	2.10 (1.07)	.01	2.08 (1.11)	2.23 (1.10)	.23	2.49 (1.23)	2.14 (1.07)	.03
Escape problems	2.35 (1.10)	2.01 (1.01)	.01	2.18 (1.07)	2.07 (1.03)	.35	2.44 (1.15)	2.03 (1.01)	.01
Feel good	2.95 (1.04)	2.56 (1.07)	.001	2.84 (1.14)	2.60 (1.04)	.06	3.03 (1.15)	2.59 (1.04)	.004
Reduce physical pain	3.67 (1.10)	3.48 (1.04)	.12	3.81 (0.94)	3.43 (1.08)	.001	3.78 (1.08)	3.48 (1.05)	.04
Reduce sadness	2.76 (1.04)	2.44 (1.00)	.01	2.42 (0.99)	2.56 (1.03)	.24	2.76 (1.01)	2.47 (1.02)	.03
Improve social situations	2.31 (0.96)	1.93 (0.93)	<.001	2.02 (1.02)	2.03 (0.93)	.92	2.25 (1.06)	1.98 (0.92)	.06
Lose weight	2.57 (1.00)	2.45 (0.99)	.31	2.46 (1.05)	2.49 (0.97)	.79	2.51 (1.04)	2.48 (0.98)	.80
Perceived risks									
Physical	4.28 (0.87)	4.37 (0.81)	.33	4.30 (0.95)	4.37 (0.78)	.53	4.22 (0.91)	4.38 (0.81)	.20
Other	4.35 (0.78)	4.44 (0.76)	.28	4.40 (0.82)	4.42 (0.74)	.77	4.21 (0.94)	4.46 (0.72)	.04
Skills									
How hard to refuse offer	1.81 (0.95)	1.49 (0.77)	.002	1.58 (0.79)	1.58 (0.85)	.96	1.79 (0.96)	1.54 (0.80)	.04
Able to refuse offer	1.69 (0.78)	1.38 (0.72)	<.001	1.50 (0.88)	1.45 (0.70)	.59	1.69 (0.87)	1.42 (0.71)	.02
How hard to refuse request for prescription	1.88 (1.05)	1.52 (0.82)	.002	1.58 (0.85)	1.62 (0.92)	.62	1.93 (1.06)	1.55 (0.86)	.01
Able to refuse request for prescription	1.67 (0.89)	1.46 (0.80)	.03	1.60 (0.97)	1.49 (0.77)	.28	1.72 (0.96)	1.48 (0.80)	.06
Knowledge	13.31 (2.00)	13.18 (2.18)	.58	13.32 (2.15)	13.18 (2.12)	.56	13.28 (2.30)	13.20 (2.09)	.80

[a]Likert item anchors or values: expectancies: strongly disagree=1, disagree=2, not sure=3, agree=4, strongly agree=5; perceived risks: no risk=1, not sure=3, great risk=5; skills or difficulty: very easy=1, easy=2, not sure=3, hard=4, very hard=5; skills or ability: definitely would=1, would=2, not sure=3, would not=4, definitely would not=5; intentions: definitely will not=1, will not=2, not sure=3, will=4, definitely will=5.

[b]NMUPO: nonmedical use of prescription opioids.

Discussion

Principal Findings

To our knowledge, we developed the first interactive, web-based, and mobile-friendly program focused specifically on the prevention of PO misuse among adolescents. This study is the first to evaluate the effectiveness of this novel tool designed to help meet the challenges of the evolving opioid crisis in the United States.

The study yielded 4 key findings. First, the results demonstrate that both the P4T and the comparison JTT programs performed well in targeting variables known to impact NMUPO. Both programs were effective in steadily and significantly decreasing agreement with positive expectancies (in all cases for P4T/in most cases for JTT) and increasing agreement with negative expectancies potentially associated with PO use over time. The exception was the potential to get in trouble with parents, and yet the baseline scores for this item were already only a half-point from the full agreement, highlighting a ceiling effect

XSL•FO
RenderX

limiting significant change. A recent longitudinal study of 3396 ethnically diverse high school students (9th and 10th grades) found that increases in positive substance use expectancies among those who never used each respective substance predicted increased odds of onset during late adolescence (11th and 12th grades; alcohol: odds ratio [OR]$_{ß}$ 7.73, P<.001; tobacco: OR$_{ß}$ 5.58, P<.001; marijuana: OR$_{ß}$ 2.49, P=.001) [62]. Mounting evidence has suggested that outcome expectancies associated with substance use play a role in the initiation and progression of drug use [77]. Thus, change in expectancies is an important target in general substance use prevention and likely for PO misuse prevention as well.

Similar to the ceiling effect for the negative expectancy potential to get in trouble with parents, youth in both conditions perceived great risks at baseline; thus, limiting room for statistically significant change. However, both programs encouraged youth to increase their negative expectations around PO use. A systematic review of risk and protective factors associated with nonmedical use of prescription drugs (including opioids) among youth in the United States identified perceived risks as among the strongest and most consistent contributing factors at the individual level [78]. Both programs effectively impacted youth's assessment of the difficulty and one's ability to refuse offers to misuse POs as well as requests for personally prescribed POs. Scores across these variables were already close to optimal at baseline with little room for change. However, both programs produced statistically significant changes.

The second key finding is that P4T outperformed JTT on knowledge and negative expectancy changes as well as perceived ease of use, ease of understanding, and how much youth liked using technology to access the program. Knowledge-based, stand-alone interventions do not prevent later substance use. Programs that combine knowledge-based components with social competence and social influence approaches have greater support for their efficacy [79]. In addition, the JTT program focused on opioids among other drugs. However, it is significant that the program was effective in helping youth attain and sustain mastery of this important content.

The P4T youth also performed better than the JTT in increasing negative expectancy changes. Cross-sectional and longitudinal research findings indicate that the level of negative substance use expectancies is negatively associated with substance use behavior and use onset [80]. As noted above, although expectancy change and its effect on substance use onset may be more salient for positive expectancies (relative to negative), including both proximal and distal, negative expectancy items may increase the power of negative expectancies to impact substance use as powerfully as positive expectancies. The P4T program enlisted the stories of 8 youths detailing countless negative effects, both proximal and distal, of their experiences with NMUPO. It is possible that the level of detail with respect to negative effects through long periods of time (in most instances) served to more significantly impact the P4T scores in this area. The expectation-experience discrepancy (violating expectations) literature suggests that although it may seem reasonable in prevention efforts to stress the physical, psychological, and other harms of using drugs, this approach requires caution. Some marijuana-focused evidence suggests that if adolescents choose to use marijuana and the promised harms fall short of the severity they were led to expect; expectation changes may strengthen usage intentions [72]. However, legitimate opioid use before high school graduation is independently associated with a 33% increase in the risk of future opioid misuse after high school, and this association is concentrated among individuals who have little to no history of drug use and strong disapproval of illegal drug use at baseline [4]. It is highly plausible that expectation-experience discrepancies combined with the highly reinforcing properties of POs may have functioned to alter attitudes in this study.

Finally, the youth significantly preferred the P4T program to the JTT program in terms of ease of use cannot be underestimated. Drug prevention programs are arguably not at the top of the average adolescent's list of preferred activities. Thus, lowering the barriers to a positive user experience in terms of program navigability and applicability to the lives of users is advantageous.

We drew a predominantly White, female, highly educated, middle adolescent (15-17 years), US sample, some of whom had personal histories of medical use of POs, friends who used POs nonmedically, and/or reported poor mental health. Aside from the highly educated [81,82] and White [83] characterizations of the sample, reporting past year medical use of POs, having friends who engage in NMUPO, and poor mental health are all often noted adolescent risk factors for NMUPO. The third key finding is that many youths in these latter 3 subgroups compared with youth not in the groups (comparators) scored consistently higher in terms of risk at baseline across variables of interest. Across all 3 subgroups (past year medical use of POs, having friends who engage in NMUPO, and poor mental health), intentions to use POs were significantly greater than those of comparators (P<.05). Youth with poor mental health and friends who engage in NMUPO underscored their elevated risk status by scoring almost all positive expectancies at a significantly higher rate than comparators, and perceived greater difficulty and lower ability to refuse offers and requests to use POs relative to comparators. These data support the growing evidence characterizing youth at greater risk for PO misuse. In addition, youth with past year medical use of POs and friends who engage in NMUPO tended toward disagreement with several negative expectancies relative to comparators (*trouble with parents* and *do poorly in school*). Those with a history of past year medical use of POs were in greater agreement than the comparators that POs might make a person *feel sleepy.* Thus, youth who had a history of past year medical use of POs may have more realistic expectations given their actual experience with the medications and be aware that POs do make you sleepy and that schoolwork does not have to suffer. Notably, youth who had past year medical use of Pos also scored only one positive expectancy (*reduce pain*) significantly higher relative to comparators. Consistent with the realistic expectations noted above, youth with medication experience may be aware that POs do indeed reduce pain. Although the evidence supports their at-risk status [4], the scores for youth who had past year medical use of POs function as a control on the other at-risk

XSL•FO
RenderX

subgroups, as their differences from comparators are all consistent with having experienced the real consequences of past year medical use of POs.

The fourth key finding is that neither of the programs performed well in changing negative expectancy scores for *overdose*. Youth hovered around *not sure* across time points and programs for the negative expectancy of *overdose*, indicating that the risk of overdose does not seem to be a particularly salient concern in this sample. This is alarming and highlights an area where both programs need to do a better job of making this very real risk clearer. Between 1999 and 2016, the pediatric mortality rate from prescription and illicit opioids increased by 262.8% (N=approximately 9000) [84]. Annual rates were the highest for youth between 15 years and 19 years: 6755 (85.3%) of the deaths were unintentional, whereas 381 (4.8%) were attributed to suicide. In this group, the rates of fatal heroin, PO, and synthetic opioid poisonings increased by 404.8%, 94.7%, and 2925.0%, respectively. From 2014 to 2016, there were 1508 opioid deaths among adolescents between ages 15 and 19 years; of these, 468 (31.0%) were attributed to synthetic opioids (eg, fentanyl). One positive finding for one at-risk subgroup identified by our analyses was that youth with poor mental health scored significantly higher at baseline in agreement with the expectancy of overdose. Youth with poor mental health are at an increased risk of suicide [85]; thus, increasing awareness of the risks associated with NMUPO among this subgroup is reassuring. A 2018 survey of pediatric residents (N=69) examined knowledge, attitudes, and barriers to overdose prevention in the clinical setting. In total, 82% reported frequent exposure to pediatric patients using opioids and at risk of overdose. Although 94% of the trainees felt they had the responsibility to educate patients about overdose risk, only 42% discussed overdose prevention, and a mere 10% ever prescribed naloxone [86]. These findings reinforce that programs such as P4T and JTT must improve messaging around the potential for overdose when using POs medically and nonmedically.

Limitations and Future Research

This study has several limitations. Recruitment advertisements were directed toward parents (after advertisements directed toward youth failed to yield interested families) skewing the sample in favor of youth from homes with parents attentive to substance misuse issues, that is, youth potentially less likely to have problems with substances or at least youth with concerned parents, impacting the generalizability of the sample. Participants were also not blinded; thus, knowledge of group assignment may have affected their behavior in the trial (eg, motivation to access the control group website) as well as their responses to subjective outcome measures. In addition, the knowledge questions posed to participants in both groups were derived from content in the P4T program and might have consequently inflated the differences between groups at multiple time points. Finally, although the measures used in this study have been used in previous studies, one limitation is that they have not been clinically validated.

Improving prevention focused on the risk factors for opioid misuse and OUD is one of the 4 central areas proposed by the National Institute on Drug Abuse, which is necessary for reversing the opioid crisis [87]. The US opioid crisis is now in its third wave [88] of opioid overdose-related deaths (wave 1: POs, wave 2: heroin, wave 3: synthetic opioids, eg, illicit fentanyl). As the nature of opioids driving the crisis evolves, prevention efforts will also need to evolve to respond effectively. On the basis of the results of our study, future research might include an enhanced approach to target at-risk subgroups (eg, poor mental health, having friends who engage in NMUPO, past year medical use of POs), as the pattern of findings point to different risk factors based on the different risk profiles (we plan to publish the differential effects of group assignment on at-risk subgroups). Moreover, due to the ceiling effect we found with respect to perceived physical and other risks and the floor effect for intentions to use, it may be useful to recruit youth who may already be experimenting and/or who endorse increased sensation-seeking or self-medicating profiles [89]. Recognizing the heterogeneity among NMUPO will enhance the evolution and efficient tailoring of prevention programs. Future research may also require an enhanced approach to target the misuse of synthetic opioids such as illicit fentanyl by youth, as underscored by the dramatic increase in death rates noted above. Another potential direction for future research is to attend to how parents contribute to or can protect against adolescent NMUPO. Parental NMUPO is associated with adolescent NMUPO [90], and parental use of opioids is associated with a doubling of the risk of a suicide attempt by their children (OR 1.99, 95% CI 1.71-2.33) [91]. However, parental involvement and disapproval of substance use [92] and monitoring as well as reduced conflict [90] are associated with reduced odds of lifetime NMUPO.

Public Health Implications

The results suggest that the P4T program, which is freely available and accessible on the internet, may help prevent future misuse of PO among adolescents. Digital technology may provide scalable, effective prevention of PO misuse among teens.

Conflicts of Interest

LM is affiliated with both HealthSim, LLC (the business that received the NIH STTR grant and developed the web-based intervention described in this manuscript) as well as Dartmouth College. This relationship is extensively managed by her academic institution, Dartmouth College. JB is a member of the board of directors and treasurer of MySafeRx Inc., a non-profit scientific research organization, and does not receive any form of compensation (financial or otherwise) from this organization. All other authors have no conflicts to declare.

Multimedia Appendix 1

XSL•FO
RenderX

CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 96 KB - publichealth_v7i2e18487_app1.pdf]

Multimedia Appendix 2
Primary and secondary outcome (means).
[DOCX File , 33 KB - publichealth_v7i2e18487_app2.docx]

Multimedia Appendix 3
Primary and secondary outcome (mixed models).
[DOCX File , 44 KB - publichealth_v7i2e18487_app3.docx]

References

1. McCabe SE, West BT, Veliz P, McCabe VV, Stoddard SA, Boyd CJ. Trends in Medical and Nonmedical Use of Prescription Opioids Among US Adolescents: 1976-2015. Pediatrics 2017 Apr;139(4) [FREE Full text] [doi: 10.1542/peds.2016-2387] [Medline: 28320868]
2. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth Risk Behavior Surveillance - United States, 2017. MMWR Surveill Summ 2018 Dec 15;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
3. Murphy SM, Friesner DL, Rosenman R. Opioid misuse among adolescents: new evidence from a misclassification analysis. Appl Health Econ Health Policy 2015 Apr;13(2):181-192 [FREE Full text] [doi: 10.1007/s40258-015-0151-z] [Medline: 25617182]
4. Miech R, Johnston L, O'Malley PM, Keyes KM, Heard K. Prescription Opioids in Adolescence and Future Opioid Misuse. Pediatrics 2015 Nov;136(5):e1169-e1177 [FREE Full text] [doi: 10.1542/peds.2015-1364] [Medline: 26504126]
5. Austic E, McCabe SE, Stoddard SA, Ngo QE, Boyd C. Age and Cohort Patterns of Medical and Nonmedical Use of Controlled Medication Among Adolescents. J Addict Med 2015;9(5):376-382 [FREE Full text] [doi: 10.1097/ADM.0000000000000142] [Medline: 26291544]
6. McCabe SE, West BT, Morales M, Cranford JA, Boyd CJ. Does early onset of non-medical use of prescription drugs predict subsequent prescription drug abuse and dependence? Results from a national study. Addiction 2007 Dec;102(12):1920-1930 [FREE Full text] [doi: 10.1111/j.1360-0443.2007.02015.x] [Medline: 17916222]
7. Dash GF, Wilson AC, Morasco BJ, Feldstein Ewing SW. A Model of the Intersection of Pain and Opioid Misuse in Children and Adolescents. Clin Psychol Sci 2018 Sep;6(5):629-646 [FREE Full text] [doi: 10.1177/2167702618773323] [Medline: 30333942]
8. Volkow ND. Teen prescription drug abuse a major health concern. Tenn Med 2009 Apr;102(4):28-29 [FREE Full text] [Medline: 19385366]
9. Substance Abuse and Mental Health Services Administration. Results from the 2013 national survey on drug use and health: summary of national findings. NSDUH Series H-48, HHS Publication No. (SMA). Rockville, MD: SAMHSA; 2014:14.
10. Center for Behavioral Health Statistics and Quality. 2015 National Survey on Drug Use and Health: Detailed tables. 2016. URL: https://www.samhsa.gov/data/report/results-2015-national-survey-drug-use-and-health-detailed-tables [accessed 2021-01-01]
11. Wu L, Pilowsky DJ, Patkar AA. Non-prescribed use of pain relievers among adolescents in the United States. Drug Alcohol Depend 2008 Apr 01;94(1-3):1-11 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.09.023] [Medline: 18054444]
12. Boyd CJ, Esteban McCabe S, Teter CJ. Medical and nonmedical use of prescription pain medication by youth in a Detroit-area public school district. Drug Alcohol Depend 2006 Jan 04;81(1):37-45 [FREE Full text] [doi: 10.1016/j.drugalcdep.2005.05.017] [Medline: 16040201]
13. McCabe SE, Boyd CJ, Young A. Medical and nonmedical use of prescription drugs among secondary school students. J Adolesc Health 2007 Jan;40(1):76-83 [FREE Full text] [doi: 10.1016/j.jadohealth.2006.07.016] [Medline: 17185209]
14. Sung H, Richter L, Vaughan R, Johnson PB, Thom B. Nonmedical use of prescription opioids among teenagers in the United States: trends and correlates. J Adolesc Health 2005 Jul;37(1):44-51. [doi: 10.1016/j.jadohealth.2005.02.013] [Medline: 15963906]
15. Whiteside LK, Walton MA, Bohnert ASB, Blow FC, Bonar EE, Ehrlich P, et al. Nonmedical prescription opioid and sedative use among adolescents in the emergency department. Pediatrics 2013 Nov;132(5):825-832 [FREE Full text] [doi: 10.1542/peds.2013-0721] [Medline: 24167166]
16. Muhuri PK, Gfroerer JC, Davies MC. Associations of Nonmedical Pain Reliever Use and Initiation of Heroin Use in the United States. CBHSQ Data Review.: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality; 2013. URL: https://www.samhsa.gov/data/sites/default/files/DR006/DR006/nonmedical-pain-reliever-use-2013.htm [accessed 2021-01-01]

XSL•FO
RenderX

17. Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. The changing face of heroin use in the United States: a retrospective analysis of the past 50 years. JAMA Psychiatry 2014 Jul 01;71(7):821-826. [doi: 10.1001/jamapsychiatry.2014.366] [Medline: 24871348]
18. Mars SG, Bourgois P, Karandinos G, Montero F, Ciccarone D. "Every 'never' I ever said came true": transitions from opioid pills to heroin injecting. Int J Drug Policy 2014 Mar;25(2):257-266 [FREE Full text] [doi: 10.1016/j.drugpo.2013.10.004] [Medline: 24238956]
19. Palamar JJ, Shearston JA, Dawson EW, Mateu-Gelabert P, Ompad DC. Nonmedical opioid use and heroin use in a nationally representative sample of us high school seniors. Drug Alcohol Depend 2016 Jan 01;158:132-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.11.005] [Medline: 26653341]
20. Curtin SC, Tejada-Vera B, Warmer M. Drug Overdose Deaths Among Adolescents Aged 15-19 in the United States: 1999-2015. NCHS Data Brief 2017 Aug(282):1-8 [FREE Full text] [Medline: 29155681]
21. Han B, Compton WM, Jones CM, Cai R. Nonmedical Prescription Opioid Use and Use Disorders Among Adults Aged 18 Through 64 Years in the United States, 2003-2013. JAMA 2015 Oct 13;314(14):1468-1478. [doi: 10.1001/jama.2015.11859] [Medline: 26461997]
22. Jones CM. The paradox of decreasing nonmedical opioid analgesic use and increasing abuse or dependence - An assessment of demographic and substance use trends, United States, 2003-2014. Addict Behav 2017 Feb;65:229-235. [doi: 10.1016/j.addbeh.2016.08.027] [Medline: 27561431]
23. Hadland SE, Wharam JF, Schuster MA, Zhang F, Samet JH, Larochelle MR. Trends in Receipt of Buprenorphine and Naltrexone for Opioid Use Disorder Among Adolescents and Young Adults, 2001-2014. JAMA Pediatr 2017 Aug 01;171(8):747-755 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0745] [Medline: 28628701]
24. Moore SK, Grabinski M, Bessen S, Borodovsky JT, Marsch LA. Web-Based Prescription Opioid Abuse Prevention for Adolescents: Program Development and Formative Evaluation. JMIR Form Res 2019 Jul 19;3(3):e12389 [FREE Full text] [doi: 10.2196/12389] [Medline: 31325289]
25. Marsch LA, Borodovsky JT. Technology-based Interventions for Preventing and Treating Substance Use Among Youth. Child Adolesc Psychiatr Clin N Am 2016 Dec;25(4):755-768. [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
26. Marsch LA, Guarino H, Grabinski MJ, Syckes C, Dillingham ET, Xie H, et al. Comparative Effectiveness of Web-Based vs. Educator-Delivered HIV Prevention for Adolescent Substance Users: A Randomized, Controlled Trial. J Subst Abuse Treat 2015 Dec;59:30-37 [FREE Full text] [doi: 10.1016/j.jsat.2015.07.003] [Medline: 26293644]
27. Palamar JJ, Shearston JA, Dawson EW, Mateu-Gelabert P, Ompad DC. Nonmedical opioid use and heroin use in a nationally representative sample of us high school seniors. Drug Alcohol Depend 2016 Jan 01;158:132-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.11.005] [Medline: 26653341]
28. Das JK, Salam RA, Arshad A, Finkelstein Y, Bhutta ZA. Interventions for Adolescent Substance Abuse: An Overview of Systematic Reviews. J Adolesc Health 2016 Oct;59(4S):S61-S75 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.06.021] [Medline: 27664597]
29. Binder C. Behavioral fluency: Evolution of a new paradigm. Behav Anal 1996;19(2):163-197 [FREE Full text] [Medline: 22478257]
30. Schepis TS, Krishnan-Sarin S. Sources of prescriptions for misuse by adolescents: differences in sex, ethnicity, and severity of misuse in a population-based study. J Am Acad Child Adolesc Psychiatry 2009 Aug;48(8):828-836 [FREE Full text] [doi: 10.1097/CHI.0b013e3181a8130d] [Medline: 19564803]
31. Hudgins JD, Porter JJ, Monuteaux MC, Bourgeois FT. Prescription opioid use and misuse among adolescents and young adults in the United States: A national survey study. PLoS Med 2019 Nov;16(11):e1002922 [FREE Full text] [doi: 10.1371/journal.pmed.1002922] [Medline: 31689290]
32. McCabe SE, Cranford JA, Boyd CJ, Teter CJ. Motives, diversion and routes of administration associated with nonmedical use of prescription opioids. Addict Behav 2007 Mar;32(3):562-575 [FREE Full text] [doi: 10.1016/j.addbeh.2006.05.022] [Medline: 16843611]
33. Boyd CJ, McCabe SE, Cranford JA, Young A. Prescription drug abuse and diversion among adolescents in a southeast Michigan school district. Arch Pediatr Adolesc Med 2007 Mar;161(3):276-281. [doi: 10.1001/archpedi.161.3.276] [Medline: 17339509]
34. Netemeyer R, Burton S, Delaney B, Hijjawi G. The Legal High: Factors Affecting Young Consumers' Risk Perceptions and Abuse of Prescription Drugs. Journal of Public Policy & Marketing 2015 Apr;34(1):103-118. [doi: 10.1509/jppm.14.073]
35. Daniulaityte R, Falck R, Carlson RG. "I'm not afraid of those ones just 'cause they've been prescribed": perceptions of risk among illicit users of pharmaceutical opioids. Int J Drug Policy 2012 Sep;23(5):374-384 [FREE Full text] [doi: 10.1016/j.drugpo.2012.01.012] [Medline: 22417823]
36. Lord S, Marsch LA. Emerging trends and innovations in the identification and management of drug use among adolescents and young adults. Adolesc Med State Art Rev 2011 Dec;22(3):649-69, xiv [FREE Full text] [Medline: 22423469]
37. Marsch LA, Gustafson DH. The Role of Technology in Health Care Innovation: A Commentary. J Dual Diagn 2013;9(1):101-103 [FREE Full text] [doi: 10.1080/15504263.2012.750105] [Medline: 23599690]
38. Ellickson P. Schools. In: Handbook on drug abuse prevention: a comprehensive strategy to prevent the abuse of alcohol and other drugs. Needham Heights, MA: Allyn & Bacon; 1995.

XSL•FO
RenderX

39. Carey KB, Scott-Sheldon LAJ, Elliott JC, Bolles JR, Carey MP. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
40. Pradhan AM, Park L, Shaya FT, Finkelstein J. Consumer Health Information Technology in the Prevention of Substance Abuse: Scoping Review. J Med Internet Res 2019 Jan 30;21(1):e11297 [FREE Full text] [doi: 10.2196/11297] [Medline: 30698526]
41. Anderson M, Jiang J. Teens, Social Media & Technology 2018. Pew Research Center. 2018. URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-01-01]
42. Nesvåg S, McKay JR. Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
43. Carey KB, Scott-Sheldon LAJ, Elliott JC, Bolles JR, Carey MP. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
44. Champion KE, Newton NC, Barrett EL, Teesson M. A systematic review of school-based alcohol and other drug prevention programs facilitated by computers or the internet. Drug Alcohol Rev 2013 Mar;32(2):115-123. [doi: 10.1111/j.1465-3362.2012.00517.x] [Medline: 23039085]
45. Walton MA, Resko S, Barry KL, Chermack ST, Zucker RA, Zimmerman MA, et al. A randomized controlled trial testing the efficacy of a brief cannabis universal prevention program among adolescents in primary care. Addiction 2014 May;109(5):786-797 [FREE Full text] [doi: 10.1111/add.12469] [Medline: 24372937]
46. Marsch LA, Bickel WK, Grabinski MJ. Application of interactive, computer technology to adolescent substance abuse prevention and treatment. Adolesc Med State Art Rev 2007 Aug;18(2):342-56, xii. [Medline: 18605650]
47. Schwinn TM, Schinke SP, Di Noia J. Preventing drug abuse among adolescent girls: outcome data from an internet-based intervention. Prev Sci 2010 Mar;11(1):24-32 [FREE Full text] [doi: 10.1007/s11121-009-0146-9] [Medline: 19728091]
48. Schinke SP, Fang L, Cole KC, Cohen-Cutler S. Preventing substance use among Black and Hispanic adolescent girls: results from a computer-delivered, mother-daughter intervention approach. Subst Use Misuse 2011;46(1):35-45 [FREE Full text] [doi: 10.3109/10826084.2011.521074] [Medline: 21190404]
49. Newton NC, Teesson M, Vogl LE, Andrews G. Internet-based prevention for alcohol and cannabis use: final results of the Climate Schools course. Addiction 2010 Apr;105(4):749-759. [doi: 10.1111/j.1360-0443.2009.02853.x] [Medline: 20148791]
50. Gulliver A, Farrer L, Chan JKY, Tait RJ, Bennett K, Calear AL, et al. Technology-based interventions for tobacco and other drug use in university and college students: a systematic review and meta-analysis. Addict Sci Clin Pract 2015 Feb 24;10:5 [FREE Full text] [doi: 10.1186/s13722-015-0027-4] [Medline: 25928221]
51. Marsch LA, Moore SK, Borodovsky JT, Solhkhah R, Badger GJ, Semino S, et al. A randomized controlled trial of buprenorphine taper duration among opioid-dependent adolescents and young adults. Addiction 2016 Aug;111(8):1406-1415 [FREE Full text] [doi: 10.1111/add.13363] [Medline: 26918564]
52. Dougherty KM, Johnston JM. Overlearning, fluency, and automaticity. Behav Anal 1996;19(2):289-292 [FREE Full text] [Medline: 22478265]
53. Johnson KR, Layng TV. Breaking the structuralist barrier. Literacy and numeracy with fluency. Am Psychol 1992 Nov;47(11):1475-1490. [Medline: 1482008]
54. Meier EA, Troost JP, Anthony JC. Extramedical use of prescription pain relievers by youth aged 12 to 21 years in the United States: national estimates by age and by year. Arch Pediatr Adolesc Med 2012 Sep 01;166(9):803-807 [FREE Full text] [doi: 10.1001/archpediatrics.2012.209] [Medline: 22566515]
55. Cantrell MA, Lupinacci P. Methodological issues in online data collection. J Adv Nurs 2007 Dec;60(5):544-549. [doi: 10.1111/j.1365-2648.2007.04448.x] [Medline: 17973718]
56. Ahern NR. Using the Internet to conduct research. Nurse Res 2005;13(2):55-70. [doi: 10.7748/nr2005.10.13.2.55.c5968] [Medline: 16416980]
57. Drug Enforcement Administration. DEA Launches Teen Web Page: Urges Teens to “Just Think Twice,” About Meth and Marijuana “It’s not All Right!”. Drug Enforcement Administration. 2005 Aug 30. URL: https://www.dea.gov/sites/default/files/states/newsrel/denver083005.html [accessed 2021-01-01]
58. Voepel-Lewis T, Boyd CJ, McCabe SE, Zikmund-Fisher BJ, Malviya S, Grant J, et al. Deliberative Prescription Opioid Misuse Among Adolescents andEmerging Adults: Opportunities for Targeted Interventions. J Adolesc Health 2018 Nov;63(5):594-600 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.07.007] [Medline: 30348282]
59. Cooper ML, Kuntsche E, Levitt A, Barber L, Wolf S. Motivational models of substance use: A review of theory and research on motives for using alcohol, marijuana, and tobacco. In: The Oxford Handbook of Substance Use and Substance Use Disorders. New York, NY: The Oxford University Press; 2016:A.
60. Jones BT, Corbin W, Fromme K. A review of expectancy theory and alcohol consumption. Addiction 2001 Jan;96(1):57-72. [doi: 10.1046/j.1360-0443.2001.961575.x] [Medline: 11177520]
61. Callas PW, Flynn BS, Worden JK. Potentially modifiable psychosocial factors associated with alcohol use during early adolescence. Addict Behav 2004 Nov;29(8):1503-1515. [doi: 10.1016/j.addbeh.2004.02.028] [Medline: 15451120]

XSL•FO
RenderX

62. Montes KS, Witkiewitz K, Pearson MR, Leventhal AM. Alcohol, tobacco, and marijuana expectancies as predictors of substance use initiation in adolescence: A longitudinal examination. Psychol Addict Behav 2019 Feb;33(1):26-34 [FREE Full text] [doi: 10.1037/adb0000422] [Medline: 30407027]
63. Johnston L, Miech R, O'Malley P, Bachman J, Schulenberg J, Patrick M. (2020). Monitoring the Future national survey results on drug use 1975- 2019;Overview, key findings on adolescent drug use. Ann Arbor. [doi: 10.3998/2027.42/150621]
64. Bachman J, Johnston L, O'Malley P, Schulenberg JE. The Monitoring the Future project after thirty-seven years: Design and procedures. In: (Monitoring the Future Occasional Paper No. 76). Ann Arbor, MI: Institute for Social Research; 2011:93.
65. Arria AM, Caldeira KM, Vincent KB, O'Grady KE, Wish ED. Perceived harmfulness predicts nonmedical use of prescription drugs among college students: interactions with sensation-seeking. Prev Sci 2008 Sep;9(3):191-201 [FREE Full text] [doi: 10.1007/s11121-008-0095-8] [Medline: 18633709]
66. Terry-McElrath YM, O'Malley PM, Patrick ME, Miech RA. Risk is still relevant: Time-varying associations between perceived risk and marijuana use among US 12th grade students from 1991 to 2016. Addict Behav 2017 Nov;74:13-19 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.026] [Medline: 28558335]
67. Yeomans-Maldonado G, Patrick ME. The Effect of Perceived Risk on the Combined Used of Alcohol and Marijuana: Results from Daily Surveys. Addict Behav Rep 2015 Dec 01;2:33-36 [FREE Full text] [doi: 10.1016/j.abrep.2015.05.004] [Medline: 26086039]
68. Parker MA, Anthony JC. Population-level predictions from cannabis risk perceptions to active cannabis use prevalence in the United States, 1991–2014. Addictive Behaviors 2018 Jul;82:101-104 [FREE Full text] [doi: 10.1016/j.addbeh.2018.02.030]
69. Schwinn TM, Schinke SP, Di Noia J. Preventing drug abuse among adolescent girls: outcome data from an internet-based intervention. Prev Sci 2010 Mar;11(1):24-32 [FREE Full text] [doi: 10.1007/s11121-009-0146-9] [Medline: 19728091]
70. Fearnow-Kenney M, Hansen WB, McNeal RB. Comparison of Psychosocial Influences on Substance Use in Adolescents: Implications for Prevention Programming. Journal of Child & Adolescent Substance Abuse 2002 Jun;11(4):1-24. [doi: 10.1300/j029v11n04_01]
71. Haerens L, Deforche B, Vandelanotte C, Maes L, De Bourdeaudhuij I. Acceptability, feasibility and effectiveness of a computer-tailored physical activity intervention in adolescents. Patient Educ Couns 2007 Jun;66(3):303-310 [FREE Full text] [doi: 10.1016/j.pec.2007.01.003] [Medline: 17298873]
72. Skenderian JJ, Siegel JT, Crano WD, Alvaro EE, Lac A. Expectancy change and adolescents' intentions to use marijuana. Psychol Addict Behav 2008 Dec;22(4):563-569 [FREE Full text] [doi: 10.1037/a0013020] [Medline: 19071982]
73. Armitagexclusions, losses TFICT. Shapiro L. T. editor. Clinical Trials, Issues and Approaches. New York: Marcel Dekker, Inc; 1983:E.
74. Marsch LA, Bickel WK, Badger GJ. Applying Computer Technology to Substance Abuse Prevention Science: Results of a Preliminary Examination. Journal of Child & Adolescent Substance Abuse 2007 Mar 06;16(2):69-94. [doi: 10.1300/j029v16n02_04]
75. Julious S. Sample Sizes for Clinical Trials. Chapman & Hall/CRC. Boca Raton, FL 2010. [doi: 10.1201/9781584887409]
76. Chow S, Shao J, Wang H. 2008. Sample Size Calculations in Clinical Research (Second Edition). Chapman & Hall/CRC. Boca Raton, FL 2008.
77. Jones BT, Corbin W, Fromme K. A review of expectancy theory and alcohol consumption. Addiction 2001 Jan;96(1):57-72. [doi: 10.1046/j.1360-0443.2001.961575.x] [Medline: 11177520]
78. Nargiso JE, Ballard EL, Skeer MR. A systematic review of risk and protective factors associated with nonmedical use of prescription drugs among youth in the United States: a social ecological perspective. J Stud Alcohol Drugs 2015 Jan;76(1):5-20. [Medline: 25486389]
79. Faggiano F, Minozzi S, Versino E, Buscemi D. Universal school-based prevention for illicit drug use. Cochrane Database Syst Rev 2014(12):CD003020. [doi: 10.1002/14651858.CD003020.pub3] [Medline: 25435250]
80. Doran N, Khoddam R, Sanders PE, Schweizer CA, Trim RS, Myers MG. A prospective study of the Acquired Preparedness Model: the effects of impulsivity and expectancies on smoking initiation in college students. Psychol Addict Behav 2013 Sep;27(3):714-722 [FREE Full text] [doi: 10.1037/a0028988] [Medline: 22686965]
81. Agrawal A, Gardner CO, Prescott CA, Kendler KS. The differential impact of risk factors on illicit drug involvement in females. Soc Psychiatry Psychiatr Epidemiol 2005 Jun;40(6):454-466. [doi: 10.1007/s00127-005-0907-0] [Medline: 16003595]
82. Martins SS, Kim JH, Chen L, Levin D, Keyes KM, Cerdá M, et al. Nonmedical prescription drug use among US young adults by educational attainment. Soc Psychiatry Psychiatr Epidemiol 2015 May;50(5):713-724 [FREE Full text] [doi: 10.1007/s00127-014-0980-3] [Medline: 25427665]
83. Ford JA, Rigg KK. Racial/Ethnic differences in factors that place adolescents at risk for prescription opioid misuse. Prev Sci 2015 Jul;16(5):633-641. [doi: 10.1007/s11121-014-0514-y] [Medline: 25344348]
84. Gaither JR, Shabanova V, Leventhal JM. US National Trends in Pediatric Deaths From Prescription and Illicit Opioids, 1999-2016. JAMA Netw Open 2018 Dec 07;1(8):e186558 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6558] [Medline: 30646334]

XSL•FO
RenderX

85. Gomes AP, Soares ALG, Kieling C, Rohde LA, Gonçalves H. Mental disorders and suicide risk in emerging adulthood: the 1993 Pelotas birth cohort. Rev Saude Publica 2019;53:96 [FREE Full text] [doi: 10.11606/s1518-8787.2019053001235 6] [Medline: 31644774]
86. Wilson JD, Berk J, Adger H, Feldman L. Identifying Missed Clinical Opportunities in Delivery of Overdose Prevention and Naloxone Prescription to Adolescents Using Opioids. J Adolesc Health 2018 Aug;63(2):245-248 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.05.011] [Medline: 30149925]
87. Volkow ND, Jones EB, Einstein EB, Wargo EM. Prevention and Treatment of Opioid Misuse and Addiction. JAMA Psychiatry 2019 Feb 01;76(2):208. [doi: 10.1001/jamapsychiatry.2018.3126]
88. Ciccarone D. The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. Int J Drug Policy 2019 Sep;71:183-188 [FREE Full text] [doi: 10.1016/j.drugpo.2019.01.010] [Medline: 30718120]
89. Young A, McCabe SE, Cranford JA, Ross-Durow P, Boyd CJ. Nonmedical use of prescription opioids among adolescents: subtypes based on motivation for use. J Addict Dis 2012;31(4):332-341 [FREE Full text] [doi: 10.1080/10550887.2012.735564] [Medline: 23244552]
90. Griesler PC, Hu M, Wall MM, Kandel DB. Nonmedical Prescription Opioid Use by Parents and Adolescents in the US. Pediatrics 2019 Mar;143(3). [doi: 10.1542/peds.2018-2354] [Medline: 30804077]
91. Brent DA, Hur K, Gibbons RD. Association Between Parental Medical Claims for Opioid Prescriptions and Risk of Suicide Attempt by Their Children. JAMA Psychiatry 2019 May 22. [doi: 10.1001/jamapsychiatry.2019.0940] [Medline: 31116357]
92. Monnat SM, Rigg KK. Examining Rural/Urban Differences in Prescription Opioid Misuse Among US Adolescents. J Rural Health 2016;32(2):204-218 [FREE Full text] [doi: 10.1111/jrh.12141] [Medline: 26344571]

Abbreviations

CHIT: consumer health information technology
DEA: Drug Enforcement Administration
JTT: JustThinkTwice
NMUPO: nonmedical use of prescription opioids
OR: odds ratio
OUD: opioid use disorder
PO: prescription opioid
P4T: POP4Teens

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by T Schwinn, D Leightley; comments to author 26.10.20; revised version received 21.12.20; accepted 13.01.21; published 25.02.21.

Please cite as:
Marsch LA, Moore SK, Grabinski M, Bessen SY, Borodovsky J, Scherer E
Evaluating the Effectiveness of a Web-Based Program (POP4Teens) to Prevent Prescription Opioid Misuse Among Adolescents: Randomized Controlled Trial
JMIR Public Health Surveill 2021;7(2):e18487
URL: https://publichealth.jmir.org/2021/2/e18487
doi:10.2196/18487
PMID:33629961

©Lisa A Marsch, Sarah K Moore, Michael Grabinski, Sarah Y Bessen, Jacob Borodovsky, Emily Scherer. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 25.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

COVID-19 and Black, Asian, and Minority Ethnic Communities: A Complex Relationship Without Just Cause

Peter Phiri[1,2], DipHE, BSc, PhD; Gayathri Delanerolle[3], BSc, MBBCHIR, MSc; Ayaat Al-Sudani[1], MSc; Shanaya Rathod[1], MSc, MRCPsych, MD

[1]Research & Development Department, Tom Rudd Unit, Moorgreen Hospital, Southern Health NHS Foundation Trust, Southampton, United Kingdom

[2]University of Southampton, Primary Care, Population Sciences and Medical Education, Faculty of Medicine, Aldermoor Health Centre, Southampton, United Kingdom

[3]Oxford Brain Health Clinical Trials Unit, Department of Psychiatry, Warnford Hospital, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Peter Phiri, DipHE, BSc, PhD
Research & Development Department
Tom Rudd Unit, Moorgreen Hospital
Southern Health NHS Foundation Trust
Research & Development Dept.
Clinical Trials Facility, Tom Rudd Unit,
Southampton, SO30 3JB
United Kingdom
Phone: 44 02380475112
Email: peter.phiri@southernhealth.nhs.uk

Abstract

Emerging evidence has indicated a negative and disproportionate impact of COVID-19 on Black, Asian, and minority ethnic (BAME) communities. Previous studies have already reported that biological and social risk factors increase disease susceptibility, particularly in BAME communities. Despite frontline workers in ethnic minority communities in the United Kingdom's National Health Service attempting to quell the pandemic, disproportionate numbers of BAME physicians and other health care workers have died of COVID-19. This unprecedented situation highlights ethical and moral implications, which could further augment the impact of the pandemic on their mental health. While the government attempts to mitigate the rate of virus transmission, certain key factors inadvertently augment the negative impact of the pandemic on the mental health and general well-being of BAME communities. This study examined the available literature to explore the association between, and the wider impact of, COVID-19 on BAME communities. Furthermore, this study aims to raise awareness and provide a deeper insight into current scientific discussions.

(JMIR Public Health Surveill 2021;7(2):e22581) doi:10.2196/22581

KEYWORDS

BAME; COVID-19; ethnicity; health care professionals; health care worker; impact; inequalities; minority; risk

The Disproportionate Impact of COVID-19 on Black, Asian, and Minority Ethnic Communities

The first 10 health care professionals in the United Kingdom to die of COVID-19 were from the Black, Asian, and minority ethnic (BAME) communities. This raised concerns of a potential association between ethnicity and a disproportionate impact of COVID-19 on these communities [1-3]. The Intensive Care National Audit and Research Centre published UK data sets from April 10, 2020, showing that one-third of COVID-19 patients admitted to critical care units belonged to an ethnic minority group. Among 3883 COVID-19 patients, 14% (n=486) were Asian and 12% (n=402) were Black [4]. Similarly, *The Guardian* [5] reported that among 12,593 patients, 19% (n=2393) of those who died of COVID-19 in hospital up till April 10, 2020, were from BAME communities. These data are discordant to those of BAME individuals in the general population in the United Kingdom (14%) [1,6]. In addition, *The Washington Post* [1] reported that in the United States, Black-majority counties have 3-fold the number COVID-19 cases and almost 6-fold the number of COVID-19–related deaths compared to White-majority counties. However, caution should

XSL•FO
RenderX

be exercised when generalizing across different health care systems [1]. A Public Health England (PHE) report [7] acknowledges the disproportionate impact of COVID-19 on BAME communities, reporting increased mortality, despite the absence of measures to potentially address the concerns identified therein.

Deaths Among National Health Service Health Care Professionals in BAME Communities

On March 12, 2020, chief medical officers in the United Kingdom elevated the country's risk status from moderate to high, and on March 23, 2020, the prime minister imposed a nationwide lockdown [8]. Since then, 181 National Health Service (NHS) workers have died of COVID-19 [9]. This figure, however, mainly relies on reports from hospitals within England only. Numerous reports have claimed that approximately 62%-75% of COVID-19–related deaths among health care workers occurred in BAME workers [1,2,4]. This is alarming because only one-third of NHS physicians in hospitals or community services in England belong to Asian (27.2%) or Black (6.95%) communities [10].

Disease Susceptibility and Predictability in BAME Communities

Many experts, including Duncan Young (professor, Intensive Care Medicine, University of Oxford), Dr Riyaz Patel (associate professor, Cardiology, University College London), and Naveed Sattar (professor, Metabolic Medicine, University of Glasgow), have suggested that ethnic minorities are at an increased risk of SARS-CoV-2 infection, severe disease, and poor outcomes owing to socially and biologically relevant reasons [11]. First, ethnicity could play a major role in disease transmission owing to cultural, behavioral, and societal differences including those in health-seeking behaviors [12], cohabiting lifestyle [13], and lower socioeconomic status. The disease transmission risk is further increased among NHS workers. Furthermore, individuals in ethnic minority communities are disproportionally employed in fields including those associated with public transport or delivery services, where there is a known, markedly higher risk of virus transmission. It is also common for BAME households to have several generations cohabiting within close confinement as culture and family are potentially important aspects of identity in these communities. Thus, it could be challenging for BAME communities to follow social distancing protocols [14,15]. Furthermore, complexities in other comorbidities such as diabetes, hypertension, and cardiovascular diseases are commonly associated with South Asians [11]. Another theory is based on the similarities observed between the risk of morality during the last major influenza crisis—the H1N1 epidemic—and ethnic minority communities in 2009-2010 and during the first postpandemic season of 2010-2011 in England. From this data set, 67 of 337 (19.9%) individuals were from BAME communities. Furthermore, ethnic minorities have been at a higher mortality risk than the Caucasian population during the 2009-2010 pandemic, with individuals of Pakistani descent being at the highest risk [16].

In addition, early studies on the disproportionate prevalence and severity of respiratory diseases among BAME communities suggest predictable health outcomes based on socioeconomic status [17]. Social stressors and environmental adversity appear to be linked to an elevated risk of cardiovascular disease and other comorbidities [16]. According to Carol Cooper, the head of equality, diversity and human rights at Birmingham Community Healthcare NHS Trust [18]:

> *Many of us knew that BAME people would be overrepresented - given their proportion of the population - in the mortality and morbidity figures because of the comorbidities that exist in our communities, because of the location of our communities in terms of the workforce being on the frontline [and] because of the amount of people that are caught in the poverty trap and live in households that have higher occupancy.*

Despite previous warnings and the need for public health authorities to identify at-risk populations, a literature review indicated that only 2 of 29 (7%) publications reported disaggregated data on ethnicity (case series without ethnicity-specific outcomes) [13]. The countries that initially reported the highest number of COVID-19 cases did not report data on ethnicity [13]. Researchers in the United Kingdom did not acquire or publish information on ethnicity until concerns of an association between COVID-19 and ethnicity began to emerge [3,13,19]. As is the case in many diseases, researchers and policy makers do not often consider ethnicity as a core factor until deaths among BAME communities become prominent in mortality data or media reports.

In addition to reports from the United Kingdom [11], those from the United States have indicated that chronic conditions such as diabetes, asthma, hypertension, kidney disease, and obesity are more common in Black American than in White American populations [20]. These conditions are associated with poor outcomes in COVID-19 cases. Moreover, Kirby [20] reported:

> *The risks of COVID-19 to Indigenous communities could not be clearer. More than 1 in 3 Indigenous Australian adults report having either cardiovascular disease, diabetes, or renal disease, and onset of these diseases often occurs 20 years earlier than the non-Indigenous population.*

The NHS Long Term Plan has identified and prioritized more common conditions including diabetes, hypertension, and obesity, but has overlooked other, more specific health conditions that increase disease severity in BAME communities, such as asthma and kidney and cardiovascular diseases [21,22]. Similarly, these conditions do not seem to be prioritized by US health care authorities [23].

XSL•FO
RenderX

Inequalities and Their Psychological Impact

Stress-related physiological and general psychological responses, such as recurrent experiences of discrimination, can significantly impact health by increasing the risk of heart disease, diabetes, and infections [24]. The PHE report of 2020 [22] asserts that this pandemic did not generate health inequalities but merely exposed and exacerbated the longstanding health and socioeconomic inequalities affecting BAME communities in the United Kingdom. Although this statement might hold true, perceptions of the underlying causal relationships vary greatly and are difficult to unravel [7]. Compared to the United Kingdom, hate crimes against Asian Americans have increased in the United States [25]. A study from the Healthforce Center at the University of California San Francisco [25] reported that 25% of working nurses are either Asian-born or Asian American in California alone. In the United Kingdom, over 1 in 5 allied health professionals, such as nurses, health visitors, and midwives, belong to a BAME community [26]. Similarly, Cook et al [27] reported that among 119 NHS deaths recorded in the NHS staff, 35 occurred among nurses, 27 among health care workers, and 18 among physicians. Furthermore, nurses belonging to BAME communities are more likely to report higher levels of stress and show signs of posttraumatic stress or other common mental health disorders than their non-BAME counterparts [28]. Furthermore, social media platforms reporting these stories can potentially influence global communities [29].

In an interview with *Nursing Times*, Carol Cooper added, "BME staff feel that they are being put on Covid wards and exposed to patients with Covid over and above their colleagues" [18]. The NHS Staff Survey of 2020 and data from the NHS Workforce Race Equality Standard (WRES) consistently provide evidence on staff in BAME communities, reporting instances of discrimination, harassment, and victimization from other staff members and the general public [26]. Consequently, some NHS staff from BAME communities may not feel confident in requesting necessary items such as personal protective equipment (PPE) and COVID-19 tests to ensure their safety. Feedback from BAME staff [14] also suggested that some forms of PPE may not be suitable. For example, some Muslim health care professionals wear a head covering (referred to as a "hijab"); hence, wearing face masks and visors may be difficult. In addition, some Muslim and Sikh men may have a beard, which could also affect the fit of face masks or PPE. Furthermore, key measures or restrictions to prevent the spread of infection could potentially be more hazardous to ethnic minorities than to others, such as withdrawing key services, implementing no-visitor policies, and social isolation or quarantine [14]. Without translation or language support, some patients are unable to articulate their health needs. Isolation is difficult owing to the multigenerational households in these communities. Furthermore, the closure of religious or community centers can impact psychological well-being. Compared to nonattenders, individuals from BAME communities who visit religious places of worship or community centers presented reduced suicide rates [30]. Closure of churches, community centers, and mosques can therefore result in poor mental health; this can, in turn, affect physical health, thus reducing the chances of survival among individuals of BAME communities [24,31].

In addition, Greenberg et al [32] reported that health care professionals are at an increased risk of "moral injury" and mental health difficulties. Health care workers are faced with detrimental decision making and extreme pressures both before and during the pandemic. Health care workers must be able to ensure the welfare of both themselves and others. They must maintain a balance between "desire" and "duty," while working with insufficient resources, particularly during COVID-19 [32]. Furthermore, Greenberg et al [32] reported that individuals developing moral injuries were likely to have negative thoughts about themselves or others as well as feelings of shame, guilt, or disgust. Negative thoughts can often lead to the development of common mental health problems, including depression and posttraumatic stress disorder, in turn affecting physiological health and even leading to suicidal ideation.

Health workers in Pakistan who have been under physical and psychological pressure have seen an increased incidence of mental health symptoms, such as heightened fears and anxiety, which could have long-term and detrimental effects on overall well-being [28]. Rana et al [28] suggested that intervention might involve the development and delivery of online content by mental health professionals to spread awareness of the psychological impact of pandemics. Psychological factors such as fear and prejudices associated with COVID-19 have led to notable levels of xenophobia, and this might have led to Zahidul Islam—a 36-year-old Bangladeshi man—committing suicide on March 25, 2020 [33,34]. This again may support the negative effects of closing religious and community centers or other places of importance, and social isolation. Zahidul may have also believed it was his "moral duty" not to transmit the virus, although his tests revealed that he did not harbor the infection [34].

Accordingly, researchers from National University Health System and Yong Loo Lin School of Medicine used a self-administered questionnaire to examine psychological distress, depression, and anxiety among health care workers in Singapore [35]. Tan et al [35] reported that during the peak of the pandemic, the incidence of anxiety was increased among nonmedical health care workers, probably owing to "less first-hand medical information on the outbreak and less intensive training on PPE and infection control measures." These suggestions are consistent with those of a systematic literature review [36] wherein numerous studies highlighted the importance of preparation, including training and work experience, during a crisis. Brooks et al [37] aimed to identify the social and occupational factors affecting the psychological well-being of health care workers involved in the severe acute respiratory syndrome crisis. They reported that "those who perceived their training as inadequate were more likely to experience burnout, posttraumatic stress and longer continuing perceived risk" [37]. This may be why death rates among NHS workers in BAME communities have been so high, and the WRES [26] similarly indicated that White NHS workers are more likely to have greater access to nonmandatory training

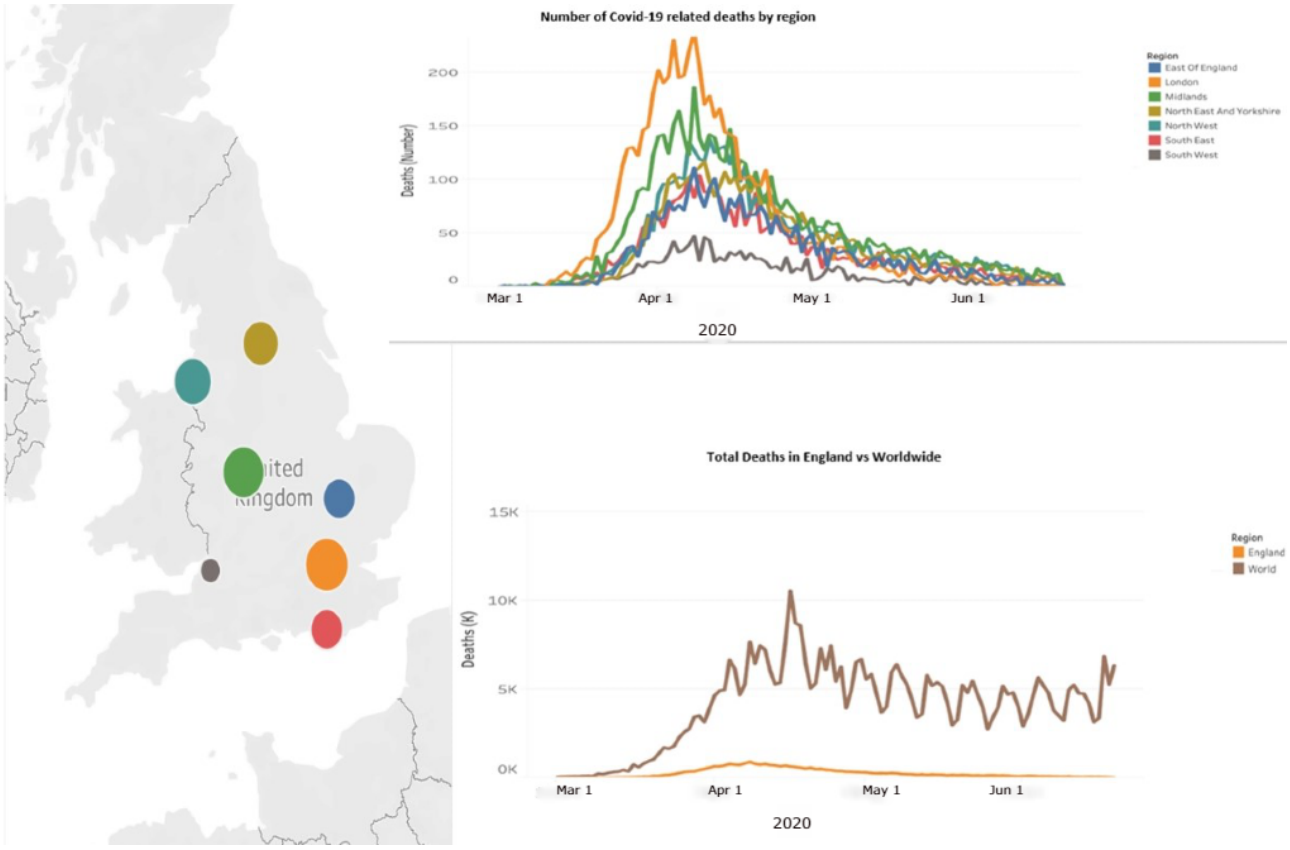
XSL•FO
RenderX

and continuous professional development than their BAME counterparts. Thus, White people may be better informed and therefore better able to cope with pandemics than their BAME counterparts.

The COVID-19 pandemic has threatened the health and lives of millions of individuals worldwide, and data from the Johns Hopkins University have reported 38,272,349 confirmed cases and 1,088,051 global deaths as of June 30, 2020 [38]. As of October 2, 2020, the United Kingdom has reported 42,369 deaths [39]. Figures 1 and 2 show the total deaths at NHS hospitals in England. As such, a virus may not discriminate among individuals; however, society apparently does, as emphasized in a recent editorial [40].

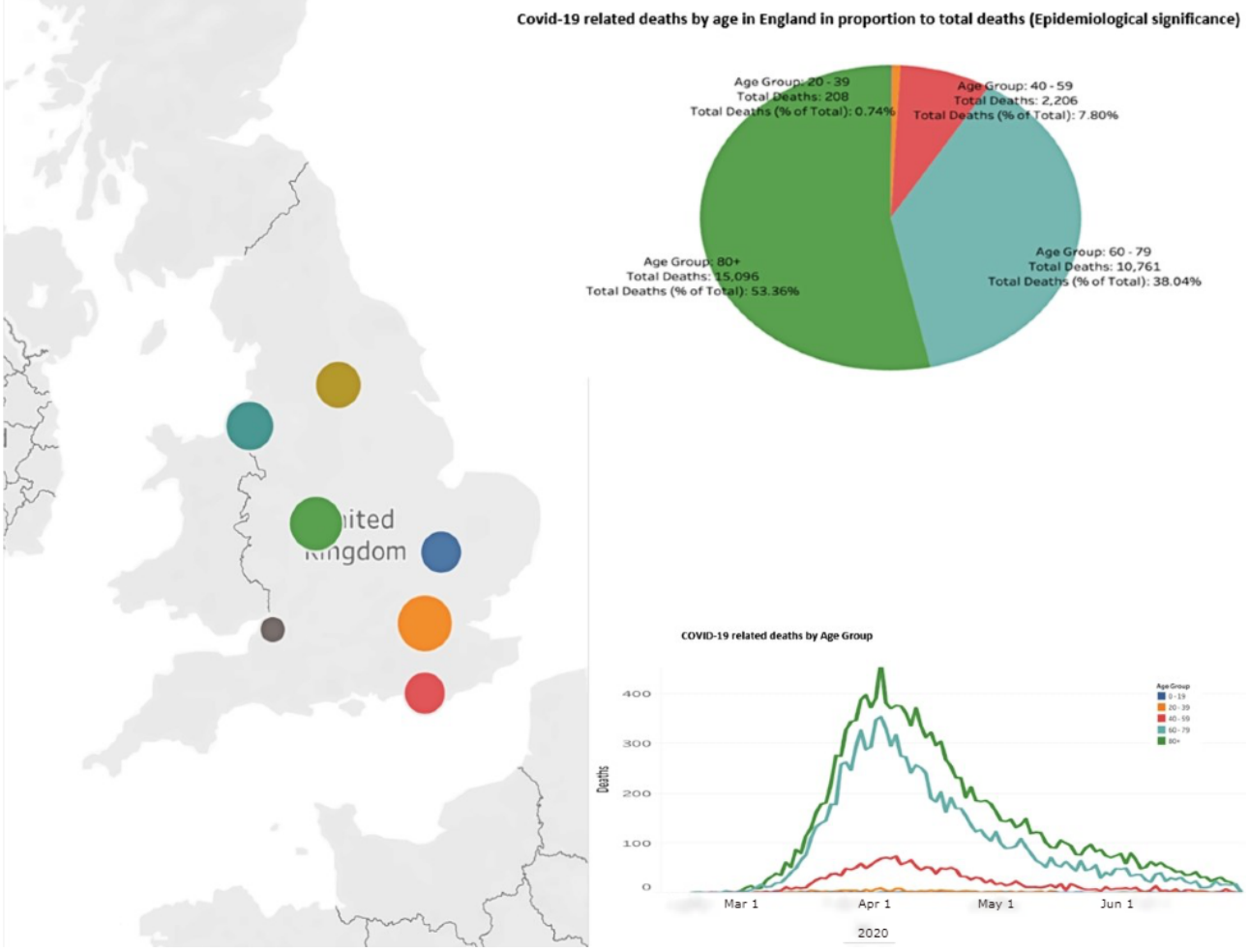
Negative effects of government measures to mitigate virus transmission albeit unintentionally increase health inequalities, including mental health inequalities [14,24,30,31]. Inequalities in wages and career development in the NHS, combined with discrimination at work and in society, restrict BAME workers to certain roles potentially rendering their specialties and services critical for combatting COVID-19 and other pandemics [40]. As previously indicated, health care workers in BAME communities may not have the confidence to voice their concerns regarding inadequate PPE, long hours of work, and low wages, which unnecessarily places them at a higher risk of succumbing to the infection [40]. Even at higher-level positions, BAME physicians and health care workers are perhaps as vulnerable as those at lower socioeconomic positions [40]. Stress, inequality, trauma, discrimination, and marginalization can negatively impact the working environment of individuals belonging to BAME communities. Furthermore, inadequacies apparent within the research sphere fail to address the inequalities among BAME groups. This was observed even during past epidemics such as the H1N1 influenza (2009), polio (2014), Ebola in West Africa (2014), Zika (2016), and Ebola in the Democratic Republic of the Congo (2019) [27], and notable effects were also observed among individuals in BAME communities in the United Kingdom [16]. Inequalities in addressing biological or physiological risk factors will inevitably lead to poor outcomes among ethnic minorities than among other populations [11,40].

Figure 1. Total COVID-19-related deaths at National Health Service Hospitals in England (data up to June 19, 2020).



XSL•FO
RenderX

Figure 2. COVID-19-related deaths by age in England (data up to June 19, 2020).



Conclusion

Emerging data strongly suggest a disproportionate impact of COVID-19 on BAME communities in the United Kingdom [1-11]. A disproportionate number of BAME physicians and health care workers in the NHS have died during the pandemic [9,10]. Nonmedical staff are at the highest risk of psychological distress, while health care professionals from BAME communities are more likely to succumb to the pandemic owing to discrimination and inequalities both at the workplace and in society [35]. These effects are further bolstered by recent government measures such as social isolation, social distancing, and lockdowns to prevent virus transmission. Stress-related physiological and psychological responses can significantly affect health by increasing the risk for heart disease, diabetes, and infection; this is a primary biological explanation for the aforementioned disproportionate impact [11]. Although mental health problems may be a secondary outcome of the pandemic, they need to be a primary focus area owing to their longer-lasting effects than those of COVID-19. The crisis thus presents an opportunity to improve mental health and bridge the aforementioned inequality gap [40]. More evidence, research, and global data on ethnicity are needed to confirm speculations of conjectured associations between COVID-19 and BAME communities, and to inform these communities about the current policies and practices. This is particularly important for regions where the pandemic is at an early stage or in those preparing for a second wave of infection. While the United Kingdom may have been late in preventing virus transmission compared to other countries, psychological intervention need not be delayed.

Authors' Contributions

PP conceived the study. PP and AAS wrote the first draft of the manuscript. PP, AAS, GD, and SR revised the manuscript critically for important intellectual content. All authors read and approved the final version of the manuscript.

Conflicts of Interest

PP received grants from NovoNordisk, Queen Mary University of London, and John Wiley & Sons during the conduct of the study.

XSL•FO
RenderX

References

1. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? BMJ 2020 Apr 20;369(4):m1548-m1394 [FREE Full text] [doi: 10.1136/bmj.m1548] [Medline: 32312785]
2. Nagpaul C. The disproportionate impact of covid-19 on ethnic minority healthcare workers. The BMJ Opinion. 2020 Apr 20. URL: https://blogs.bmj.com/bmj/2020/04/20/chaand-nagpaul-the-disproportionate-impact-of-covid-19-on-ethnic-minority-healthcare-workers/ [accessed 2020-04-28]
3. Siddique H. UK government urged to investigate coronavirus deaths of BAME doctors. The Guardian. 2020 Apr 10. URL: https://www.theguardian.com/society/2020/apr/10/uk-coronavirus-deaths-bame-doctors-bma [accessed 2020-04-28]
4. ICNARC report on COVID-19 in critical care: 10 April 2020. Intensive Care National Audit & Research Centre. URL: https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports [accessed 2020-04-29]
5. Barr C, Kommenda N, McIntyre N, Voce A. Ethnic minorities dying of Covid-19 at higher rate, analysis shows. The Guardian. 2020 Apr 22. URL: https://www.theguardian.com/world/2020/apr/22/racial-inequality-in-britain-found-a-risk-factor-for-covid-19 [accessed 2020-04-29]
6. 2011 Census. Office for National Statistics. 2011. URL: https://www.ons.gov.uk/census/2011census [accessed 2020-04-29]
7. Disparities in the risk and outcomes of COVID-19. Public Health England. 2020 Aug. URL: https://tinyurl.com/y54koq4l [accessed 2020-04-29]
8. Prime Minister's statement on coronavirus (COVID-19): 23 March 2020. Government of the United Kingdom. 2020 Mar 23. URL: https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020 [accessed 2020-02-10]
9. Allen-Kinross P. We don't know exactly how many NHS workers have died from Covid-19. Full Fact. 2020 May 22. URL: https://fullfact.org/health/we-dont-know-exactly-how-many-nhs-workers-have-died-covid-19/ [accessed 2020-06-01]
10. NHS Workforce Statistics - March 2019 (Including supplementary analysis on pay by ethnicity). NHS Workforce Statistics. 2019 Jul 27. URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics/nhs-workforce-statistics---march-2019-provisional-statistics [accessed 2020-06-01]
11. Expert comments on BME groups and risk of hospitalisation with COVID-19. Science Media Centre. 2020 Apr 09. URL: https://www.sciencemediacentre.org/expert-comments-on-bme-groups-and-risk-of-hospitalisation-with-covid-19/ [accessed 2020-04-29]
12. Rathod S, Kingdon D, Phiri P, Gobbi M. Developing culturally sensitive cognitive behaviour therapy for psychosis for ethnic minority patients by exploration and incorporation of service users' and health professionals' views and opinions. Behav Cogn Psychother 2010 Oct;38(5):511-533. [doi: 10.1017/S1352465810000378] [Medline: 20630118]
13. Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, et al. Ethnicity and COVID-19: an urgent public health research priority. Lancet 2020 May 02;395(10234):1421-1422 [FREE Full text] [doi: 10.1016/S0140-6736(20)30922-3] [Medline: 32330427]
14. The impact of COVID-19 on BME communities and health and care staff. NHS Confederation. 2020 Apr. URL: https://tinyurl.com/yd5p2r3l [accessed 2020-04-30]
15. Catney G, Sabater A. Ethnic minority disadvantage in the labour market. Joseph Rowntree Foundation. 2015 Mar 30. URL: https://www.jrf.org.uk/report/ethnic-minority-disadvantage-labour-market [accessed 2020-04-30]
16. Zhao H, Harris RJ, Ellis J, Pebody RG. Ethnicity, deprivation and mortality due to 2009 pandemic influenza A(H1N1) in England during the 2009/2010 pandemic and the first post-pandemic season. Epidemiol Infect 2015 Dec;143(16):3375-3383. [doi: 10.1017/S0950268815000576] [Medline: 25850904]
17. Myers EM. Compounding Health Risks and Increased Vulnerability to SARS-CoV-2 for Racial and Ethnic Minorities and Low Socioeconomic Status Individuals in the United States. Preprints. Preprint posted online on April 15, 2020 [FREE Full text] [doi: 10.1080/17437199.2015.1138093]
18. Ford M. Exclusive: BME nurses 'feel targeted' to work on Covid-19 wards. Nursing Times. 2020 Apr 17. URL: https://www.nursingtimes.net/news/coronavirus/exclusive-bme-nurses-feel-targeted-to-work-on-covid-19-wards-17-04-2020/ [accessed 2020-04-29]
19. Croxford R. Coronavirus cases to be tracked by ethnicity. BBC News. 2020 Apr 18. URL: https://www.bbc.co.uk/news/health-52338101 [accessed 2020-04-29]
20. Kirby T. Evidence mounts on the disproportionate effect of COVID-19 on ethnic minorities. Lancet Respir Med 2020 Jun;8(6):547-548 [FREE Full text] [doi: 10.1016/S2213-2600(20)30228-9] [Medline: 32401711]
21. NHS Long Term Plan. NHS. URL: https://www.longtermplan.nhs.uk/ [accessed 2020-10-01]
22. Beyond the data: Understanding the impact of COVID-19 on BAME groups. Public Health England. 2020 Jun. URL: https://tinyurl.com/y797mv8c [accessed 2021-01-26]
23. National Health Initiatives, Strategies and Action Plans. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/publichealthgateway/strategy/index.html [accessed 2020-10-01]
24. Anderson KF. Diagnosing Discrimination: Stress from Perceived Racism and the Mental and Physical Health Effects*. Sociol Inq 2012 Sep 07;83(1):55-81 [FREE Full text] [doi: 10.1111/j.1475-682X.2012.00433.x]

XSL•FO
RenderX

25. Natividad I. Racist harassment of Asian health care workers won't cure coronavirus. Berkeley News. 2020 Apr 09. URL: https://news.berkeley.edu/2020/04/09/racist-harassment-of-asian-health-care-workers-wont-cure-coronavirus/ [accessed 2020-05-01]
26. The WRES implementation team. NHS Workforce Race Equality Standard. NHS. 2020 Feb. URL: https://www.england.nhs.uk/wp-content/uploads/2020/01/wres-2019-data-report.pdf [accessed 2020-05-04]
27. Cook T, Kursumovic E, Lennane S. Exclusive: deaths of NHS staff from covid-19 analysed. HSJ. 2020 Apr 22. URL: https://www.hsj.co.uk/exclusive-deaths-of-nhs-staff-from-covid-19-analysed/7027471.article [accessed 2020-04-29]
28. Rana W, Mukhtar S, Mukhtar S. Mental health of medical workers in Pakistan during the pandemic COVID-19 outbreak. Asian J Psychiatr 2020 Jun;51:102080 [FREE Full text] [doi: 10.1016/j.ajp.2020.102080] [Medline: 32283512]
29. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
30. VanderWeele TJ, Li S, Tsai AC, Kawachi I. Association Between Religious Service Attendance and Lower Suicide Rates Among US Women. JAMA Psychiatry 2016 Aug 01;73(8):845-851 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.1243] [Medline: 27367927]
31. Reger MA, Stanley IH, Joiner TE. Suicide Mortality and Coronavirus Disease 2019-A Perfect Storm? JAMA Psychiatry 2020 Apr 10. [doi: 10.1001/jamapsychiatry.2020.1060] [Medline: 32275300]
32. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]
33. Youth Suicide in Gaibandha Due to Doubt of Corona [Bangla]. Somoy News. 2020 Mar. URL: https://m.somoynews.tv/pages/details/204558 [accessed 2020-04-30]
34. Mamun MA, Griffiths MD. First COVID-19 suicide case in Bangladesh due to fear of COVID-19 and xenophobia: Possible suicide prevention strategies. Asian J Psychiatr 2020 Jun;51:102073 [FREE Full text] [doi: 10.1016/j.ajp.2020.102073] [Medline: 32278889]
35. Tan BY, Chew NW, Lee GK, Jing M, Goh Y, Yeo LL, et al. Psychological Impact of the COVID-19 Pandemic on Health Care Workers in Singapore. Ann Intern Med 2020 Aug 18;173(4):317-320 [FREE Full text] [doi: 10.7326/M20-1083] [Medline: 32251513]
36. Iversen AC, Fear NT, Ehlers A, Hacker Hughes J, Hull L, Earnshaw M, et al. Risk factors for post-traumatic stress disorder among UK Armed Forces personnel. Psychol Med 2008 Apr;38(4):511-522 [FREE Full text] [doi: 10.1017/S0033291708002778] [Medline: 18226287]
37. Brooks SK, Dunn R, Amlôt R, Rubin GJ, Greenberg N. A Systematic, Thematic Review of Social and Occupational Factors Associated With Psychological Outcomes in Healthcare Employees During an Infectious Disease Outbreak. J Occup Environ Med 2018 Mar;60(3):248-257. [doi: 10.1097/JOM.0000000000001235] [Medline: 29252922]
38. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Coronavirus Resource Center: Johns Hopkins University. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-06-30]
39. United Kingdom. Worldometer. URL: https://www.worldometers.info/coronavirus/country/uk/ [accessed 2020-10-02]
40. Kapilashrami A, Bhui K. Mental health and COVID-19: is the virus racist? Br J Psychiatry 2020 Aug;217(2):405-407 [FREE Full text] [doi: 10.1192/bjp.2020.93] [Medline: 32368995]

Abbreviations

BAME: Black, Asian, and minority ethnic
NHS: National Health Service
PHE: Public Health England
PPE: personal protective equipment
WRES: NHS Workforce Race Equality Standard

Edited by G Eysenbach; submitted 17.07.20; peer-reviewed by J Kaewkungwal, E Davies; comments to author 01.10.20; revised version received 15.10.20; accepted 17.01.21; published 01.02.21.

Please cite as:
Phiri P, Delanerolle G, Al-Sudani A, Rathod S
COVID-19 and Black, Asian, and Minority Ethnic Communities: A Complex Relationship Without Just Cause
JMIR Public Health Surveill 2021;7(2):e22581
URL: https://publichealth.jmir.org/2021/2/e22581
doi:10.2196/22581
PMID:33481752

XSL•FO
RenderX

©Peter Phiri, Gayathri Delanerolle, Ayaat Al-Sudani, Shanaya Rathod. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 01.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Characterization of Patients With COVID-19 in Primary Care in Catalonia: Retrospective Observational Study

Miguel Angel Mayer[1], MD, MPH, MSci, PhD; Josep Vidal-Alaball[2,3], MD, MPH, PhD; Anna Puigdellívol-Sánchez[4,5,6], MD, PhD; Francesc X Marín Gomez[2,3], MD, PhD; Angela Leis[1], PsyM; Jacobo Mendioroz Peña[2,7], MD, MPH

[1]Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Faculty of Health and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

[2]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[3]Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

[4]Facultat de Medicina, Universitat de Vic-Universitat Central de Catalunya, Vic, Spain

[5]CAP Anton de Borja, Consorci Sanitari de Terrassa, Terrassa, Spain

[6]Unitat d'Anatomia Humana Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain

[7]COVID-19 Response Unit, Department of Health, Generalitat de Catalunya, Barcelona, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, MPH, PhD
Health Promotion in Rural Areas Research Group
Gerència Territorial de la Catalunya Central
Institut Català de la Salut
Carrer Pica d'Estats, 13-15
Sant Fruitós de Bages, 08272
Spain
Phone: 34 936930040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: The country of Spain has one of the highest incidences of COVID-19, with more than 1,000,000 cases as of the end of October 2020. Patients with a history of chronic conditions, obesity, and cancer are at greater risk from COVID-19; moreover, concerns surrounding the use of angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin type II receptor blockers (ARBs) and its relationship to COVID-19 susceptibility have increased since the beginning of the pandemic.

Objective: The objectives of this study were to compare the characteristics of patients diagnosed with COVID-19 to those of patients without COVID-19 in primary care; to determine the risk factors associated with the outcome of mortality; and to determine the potential influence of certain medications, such as ACEIs and ARBs, on the mortality of patients with COVID-19.

Methods: An observational retrospective study of patients diagnosed with COVID-19 in the Catalan Central Region of Spain between March 1 and August 17, 2020, was conducted. The data were obtained from the Primary Care Services Information Technologies System of the Catalan Institute of Health in Barcelona, Spain.

Results: The study population included 348,596 patients (aged >15 years) registered in the Primary Care Services Information Technologies System of the Catalan Central Region. The mean age of the patients was 49.53 years (SD 19.42), and 31.17% of the patients were aged ≥60 years. 175,484/348,596 patients (50.34%) were women. A total of 23,844/348,596 patients (6.84%) in the population studied were diagnosed with COVID-19 during the study period, and the most common clinical conditions of these patients were hypertension (5267 patients, 22.1%) and obesity (5181 patients, 21.7%). Overall, 2680/348,596 patients in the study population (0.77%) died during the study period. The number of deaths among patients without COVID-19 was 1825/324,752 (0.56%; mean age 80.6 years, SD 13.3), while among patients diagnosed with COVID-19, the number of deaths was 855/23,844 (3.58%; mean age 83.0 years, SD 10.80) with an OR of 6.58 (95% CI 6.06-7.15).

Conclusions: We observed that women were more likely to contract COVID-19 than men. In addition, our study did not show that hypertension, obesity, or being treated with ACEIs or ARBs was linked to an increase in mortality in patients with COVID-19. Age is the main factor associated with mortality in patients infected with SARS-CoV-2.

XSL•FO
RenderX

(JMIR Public Health Surveill 2021;7(2):e25452) doi:10.2196/25452

KEYWORDS

COVID-19; risk factors; primary health care; angiotensin-converting enzyme inhibitors; angiotensin II type 2 receptor blockers; risk; characteristic; retrospective; observational; Spain; mortality

Introduction

A highly pathogenic coronavirus, SARS-CoV-2, was first described in Wuhan in late December 2019 and has since spread worldwide [1,2]. As a result, the second meeting of the International Health Regulations Emergency Committee was convened by the Director-General of the World Health Organization (WHO) regarding the outbreak of this coronavirus to declare it a public health emergency of international concern [3], and it was declared a pandemic on March 11, 2020 [4]. Spain, particularly some geographical areas such as the autonomous communities of Catalonia and Madrid, has been one of the countries most affected by COVID-19, with more than one million cases by the end of October 2020 [5].

Since the beginning of the pandemic, it has been clearly shown that COVID-19 disproportionately affects patients with a history of chronic conditions such as cardiovascular disease, chronic obstructive pulmonary disease (COPD), hypertension, and diabetes mellitus [6,7]. Obesity and cancer have also been linked to worse outcomes and prognosis for COVID-19 [8]. The relationship between smoking and the severity of COVID-19 infection is also controversial. At the beginning of the pandemic, studies pointed to a protective relationship between smoking and COVID-19 and suggested that nicotine could have a protective effect against COVID-19 due to its minor anti-inflammatory properties. This effect could even be more marked considering that it could be masked by smoking-related toxicity and cessation of smoking when patients are severely ill with COVID-19 [9-11]. The information linking smoking and protection against COVID-19 prompted the WHO to issue a warning that the severity of COVID-19 disease was higher among smokers [12]. Other studies have supported this statement, indicating that tobacco smoking increases the lung gene expression of angiotensin-converting enzyme II (ACE2) and therefore increases the severity of COVID-19 [13-16].

Although angiotensin-converting enzyme (ACE) and ACE2 are distinct enzymes with different mechanisms of action [17], ACE2 has been linked to susceptibility to COVID-19 as well as its severity, as the viral protein involved in cell entry (spike S protein) binds to ACE2 [17-19]. In rat hearts, angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin-receptor blockers (ARBs) have been shown to increase the expression of ACE2 [20], and it has been suggested that these treatments can predispose patients to more serious SARS-CoV-2 infection and worse outcomes and mortality [17]. Other studies have postulated that ACEI has a protective role in COVID-19 and could even be used as a treatment to reduce lung complications resulting from the disease [21].

Although the burden of managing the COVID-19 pandemic initially fell on hospitals, this situation has gradually changed; primary care services are handling more cases, which is requiring substantial changes in the way primary care services are delivered to populations to manage the COVID-19 pandemic [22]. Patients are usually encouraged to request advice and speak first to a primary care physician by telephone or a video call to avoid in-person appointments and going to primary care settings. Countries recognized to have strong primary care systems, such as Spain, Belgium, and the United Kingdom, have experienced high COVID-19 mortality rates [23,24]. A recent study analyzing the prognostic factors of patients with COVID-19 in primary care was also conducted in Catalonia, including 322 consecutive patients with COVID-19. Being older, male sex, and autoimmune disease were the main predictors of admission to hospital or death [25]. Analysis of the characteristics of the population followed up in primary care and comparison between the characteristics of patients with and without COVID-19 can provide useful information for decision-making in pandemic situations.

The objectives of this study were (1) to compare the characteristics of adult patients diagnosed with COVID-19 compared to patients without COVID-19 in primary care; (2) to determine the risk factors for these patients associated with fatality as the outcome; and (3) to analyze the influence of taking medications such as ACEIs and ARBs on the mortality of patients with COVID-19.

Methods

Population

An observational retrospective study was conducted on patients diagnosed with COVID-19 living in Spain, specifically in the Catalan Central Region, from March 1 to August 17, 2020. The population in the region over the age of 15 years was included in the study, with 348,596 patients. Patients older than 15 years are cared for by family and community specialists in primary care in Spain. The population of the area, including patients younger than 16 years, was 415,000 at the time of the study. Among the patients included in the study, 6.8% (23,844/348,596) had been diagnosed with COVID-19.

The data were extracted from the computerized medical records of the Primary Care Services Information Technologies (IT) System of the Catalan Institute of Health in Barcelona, Spain. The Primary Care Services IT System contains primary care electronic health records (EHRs) for over 6 million people in Catalonia, covering more than 80% of the Catalan population [26]. This system applies an anonymization process to maintain complete confidentiality and privacy of these data, following the European General Data Protection Regulation 2016/679 of April 27 and the Spanish Organic Law 3/2018 on Data Protection and Guarantee of Digital Rights. Variables such as demographic characteristics (ie, age and sex) and diverse comorbidities, such as hypertension, diabetes, obesity, dyslipidemia, COPD, being a current smoker, heart failure,

XSL•FO
RenderX

cerebrovascular disease, and ischemic heart disease, were studied. Patients being treated with ACEIs or ARBs were also considered. The different diagnoses of risk factors and COVID-19 were based on the International Statistical Classification of Diseases and Related Health Problems 10th Edition (ICD-10); the codes related to these diseases are registered in the EHR database. Specifically, the codes used were based on the WHO and the Spanish Ministry of Health codification recommendations: U07.1 COVID-19; B97.29: Other coronavirus as the cause of diseases classified elsewhere and B34.2: Coronavirus infection, unspecified. The complete list of codes used in the study is included in Multimedia Appendix 1. The death events of the patients were also obtained from this database, which records the exact moment a patient dies.

The study protocol was approved by the University Institute for Primary Care Research Jordi Gol Health Care Ethics Committee (Code 20/066-PCV).

Statistical Analysis

Categorical variables are described using frequencies and percentages. Continuous variables are described with means and standard deviation. Proportions of categorical variables were compared using the Fisher exact test, and the sample *t* test or Wilcoxon rank-sum test were used in the case of continuous variables. Multiple logistic regression analysis for predicting binary outcomes from continuous and categorical variables was applied. Multivariate logistic regression models were applied for the comparison between groups (patients with and without COVID-19, and deceased and living patients with COVID-19), in which risk factors were adjusted for age and sex. The level of significance used was $P<.05$. The statistical analysis was conducted using R version 3.6.3 (R Project) and Jamovi version 1.2.24.0.

Results

The study population included 348,596 patients over 15 years of age registered in the Primary Care Services IT System of the Central Catalan Region in Spain. The mean age of the patients was 49.53 years (SD 19.4), and 108,762/348,596 (31.2%) of the patients were aged ≥60 years. Overall, 175,484/348,596 patients (50.3%) were female. The most common comorbidities and clinical conditions were hypertension (75,699/348,596, 21.7%), dyslipidemia (71,424/348,596, 20.5%), and obesity (69,501/348,596, 19.9%). Regarding patients with hypertension, 22,771/75,699 (30.1%) of them were being treated with ACEIs and 9487/75,699 (12.5%) were being treated with ARBs. A comparison of the demographics and risk factors of patients without COVID-19 and patients diagnosed with COVID-19 is shown in Table 1.

The characteristics of patients with and without COVID-19 are shown in Table 1. A total of 23,844 patients in the study population of 348,596 (6.8%) were diagnosed with COVID-19 during the study period, with a mean age of 49.9 years (SD 20.0) and no significant differences compared to patients without COVID-19 ($\chi^2_1=18.1$, $P=.10$) adjusted for sex. However, COVID-19 diagnosis was more frequent in patients between 31 and 40 years of age, between 41 and 50 years of age, between 81 and 90 years of age, and aged ≥90 years, with significant differences between the groups with and without COVID-19 ($\chi^2_7=1521.5$, $P<.001$). Of the 23,844 patients diagnosed with COVID-19, 13,763 (57.7%) were women. The most frequent clinical conditions among these 23,844 patients were hypertension (5267 patients, 22.1%), obesity (5181 patients, 21.7%) and dyslipidemia (4749 patients, 19.9%). Among the 5267 patients with COVID-19 and hypertension, 1701 (32.3%) were in treatment with ACEIs, and 729 (13.8%) were in treatment with ARBs.

Patients who had been diagnosed with COVID-19 were more likely to have diverse comorbidities. Diabetes ($P=.01$), obesity ($P<.001$), COPD ($P=.001$), cancer ($P<.001$), smoking ($P<.001$), heart failure ($P<.001$), and cerebrovascular disease ($P<.001$) were significantly more frequent among patients with COVID-19 compared to patients without COVID-19. Moreover, hypertension ($P=.77$) and ischemic heart disease ($P=.02$) were more frequent among patients with COVID-19; however, the differences were significant only in the case of ischemic heart disease. Finally, dyslipidemia ($P<.001$) was more frequent among patients without COVID-19, with a significant difference.

XSL•FO
RenderX

Table 1. Demographics and comorbidities of the study population, including patients with and without COVID-19 (N=348,596). Multivariate analysis of the risk factors was performed. Age was adjusted for sex and sex was adjusted for age. The remaining risk factors were adjusted for age and sex.

Demographics and risk factors	Patients with COVID-19 (n=23,844)	Patients without COVID-19 (n=324,752)	Adjusted odds ratio (95% CI)	*P* value
Demographics				
Age (years), mean (SD)	49.93 (19.4)	49.53 (20.0)	1.0 (0.99-1.0)	.10
Age group (years), n (%)			N/A[a]	<.001
16-30	4239 (19.2)	61,382 (20.4)		
31-40	4182 (18.9)	49,610 (16.5)		
41-50	4381 (19.8)	56,457 (18.7)		
51-60	3685 (16.6)	48,309 (16.0)		
61-70	1841 (8.3)	39,163 (13.0)		
71-80	1335 (6.0)	25,999 (8.6)		
81-90	1763 (8.0)	16,884 (5.6)		
>90	739 (3.3)	3620 (1.2)		
Sex			0.73 (0.71-0.75)	<.001
Female	13,763 (57.7)	161,721 (49.8)		
Male	10,081 (42.3)	163,031 (51.2)		
Risk factors				
Hypertension	5267 (22.1)	70,432 (21.7)	1.01 (0.97-1.04)	.77
Diabetes	2101 (8.8)	27,379 (8.4)	1.06 (1.01-1.12)	.01
Dyslipidemia	4749 (19.9)	66,675 (20.5)	0.94 (0.90-0.97)	<.001
Obesity	5181 (21.7)	64,320 (19.8)	1.10 (1.06-1.14)	<.001
Chronic obstructive pulmonary disease	685 (2.9)	8193 (2.5)	1.23 (1.14-1.34)	.001
Cancer	1765 (7.4)	21,785 (6.7)	1.09 (1.04-1.15)	<.001
Current smoker	3578 (15.0)	46,659 (14.3)	1.09 (1.05-1.13)	<.001
Heart failure	515 (2.2)	4301 (1.3)	1.63 (1.48-1.79)	<.001
Cerebrovascular disease	176 (0.7)	1658 (0.5)	1.48 (1.26-1.72)	<.001
Ischemic heart disease	627 (2.6)	8122 (2.5)	1.11 (1.02-1.20)	.02

[a]N/A: not applicable.

A total of 2680/348,596 patients (0.77%) in the overall study population died during the study period. The number of deaths among patients without COVID-19 was 1825/324,752 (0.56%); meanwhile, the number of deaths among patients diagnosed with COVID-19 was 855 (3.72%), with an odds ratio (OR) of 6.58 (χ^2_1=2658.4, *P*<.001, 95% CI 6.06-7.15). However, the number of deaths among patients aged ≥60 years without COVID-19 was 1681 (1.67%), and the number of deaths among patients with COVID-19 was 828 (14.02%), with an OR of 8.36 (χ^2_1=3168.5, *P*<.001, 95% CI 7.66-9.12). Moreover, the number of deaths among patients aged <60 years without COVID-19 was 144 (0.07%), and that among patients diagnosed with COVID-19 was 17 (0.16%), with an OR of 2.44 (χ^2_1=18.1, *P*<.001, 95% CI 1.62-3.69).

The frequency of risk factors and the statistical analysis of the association of the different risk factors with mortality adjusted for age and sex among patients diagnosed with COVID-19 are shown in Table 2. The mean age of patients with COVID-19 who died was 83 years (SD 10.80), with significant differences compared to the mean age (48.7 years, SD 19.22) of patients who survived ($t_{1066.4}$=–87.84, *P*<.001). Men with COVID-19 died more frequently than women compared to the group of patients who survived. The mean age of death in patients with COVID-19 (83.0 years, SD 10.80) was higher than the mean age of death in patients without COVID-19 (80.6 years, SD 13.3), with a significant difference ($t_{2015.5}$=–4.9936, *P*<.001).

XSL•FO
RenderX

Table 2. Characteristics of deceased and living patients diagnosed with COVID-19.

Demographics and risk factors	Deceased patients with COVID-19 (n=855)	Living patients with COVID-19 (n=22,989)	Adjusted odds ratio (95% CI)	*P* value[a]
Demographics				
Age (years), mean (SD)	83 (10.8)	48.7 (19.2)	1.12 (1.11-1.13)	<.001
Sex			2.22 (1.90-2.60)	<.001
Male	468 (54.7)	13,295 (57.8)		
Female	387 (45.3)	9694 (42.2)		
Risk factors				
Hypertension	587 (68.7)	4680 (20.4)	1.16 (0.98-1.37)	.09
Diabetes	278 (32.5)	1823 (7.9)	1.69 (1.43-1.99)	<.001
Dyslipidemia	383 (44.7)	4367 (19.0)	1.19 (1.03-1.39)	.03
Obesity	250 (29.2)	4931 (21.5)	1.08 (0.91-1.27)	.38
Chronic obstructive pulmonary disease	105 (12.3)	580 (2.5)	1.18 (0.92-1.51)	.19
Cancer	210 (24.6)	1555 (6.8)	1.19 (0.99-1.43)	.06
Current smoker	37 (4.3)	3541 (15.4)	0.98 (0.67-1.39)	.93
Heart failure	127 (14.9)	388 (1.7)	1.59 (1.26-1.99)	<.001
Cerebrovascular disease	21 (2.5)	155 (0.7)	0.84 (0.50-1.33)	.47
Ischemic heart disease	93 (10.9)	534 (2.3)	1.20 (0.93-1.54)	.15
Treatment with ACEIs[b]	152 (17.8)	1549 (6.7)	0.90 (0.74-1.09)	.32
Treatment with ARBs[c]	75 (8.8)	654 (2.8)	1.00 (0.76-1.30)	.98

[a]In the multivariate analysis, sex was adjusted for age and age was adjusted for sex. The remaining risk factors were adjusted for age and sex.

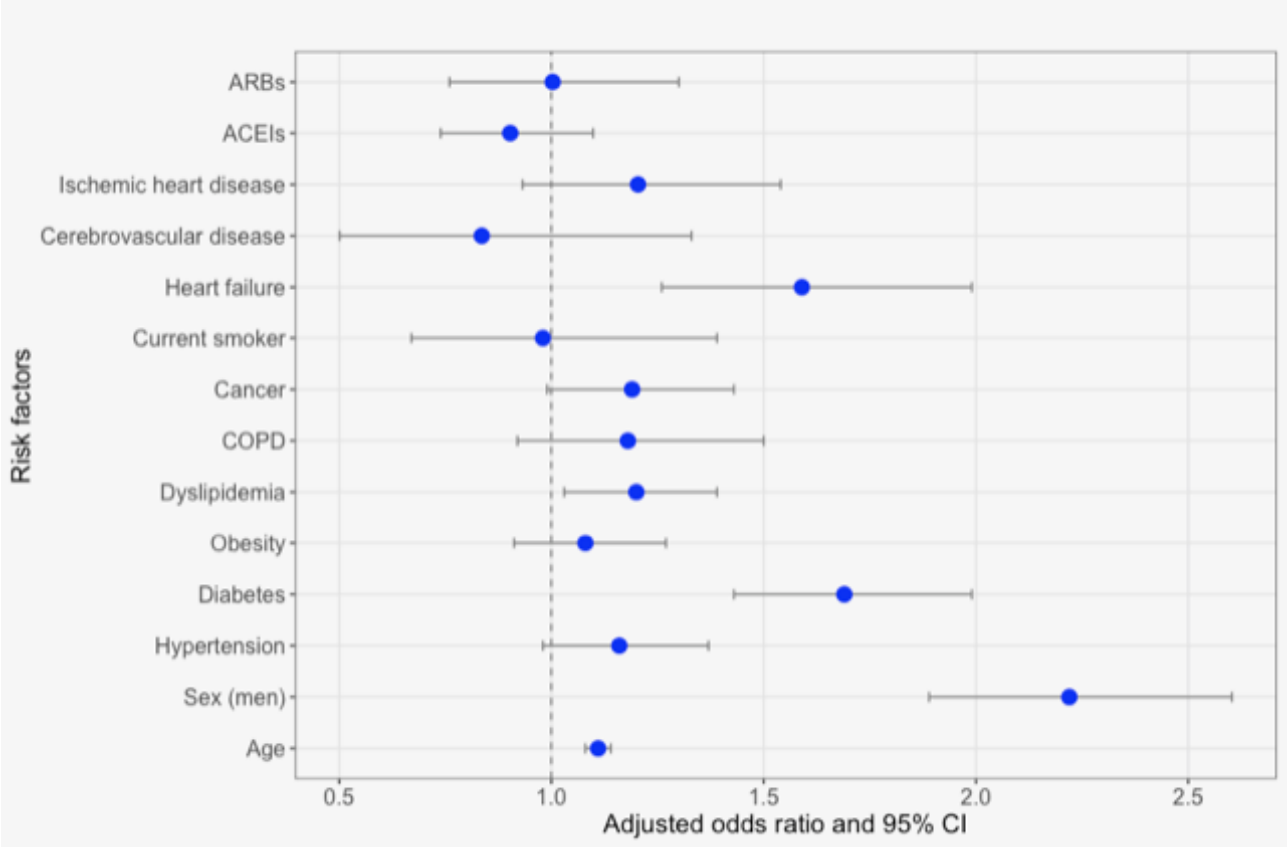
[b]ACEIs: angiotensin-converting-enzyme inhibitors.

[c]ARBs: angiotensin II receptor blockers.

The logistic regression analysis showed that several conditions, such as diabetes, dyslipidemia, and heart failure, were associated with increased death risk with significant differences. Age and (male) sex were also associated with an increased death risk, with significant differences. Meanwhile, hypertension, obesity, COPD, cancer, being a current smoker, cerebrovascular disease, and ischemic heart disease did not show an increased risk of mortality among patients diagnosed with COVID-19. Figure 1 shows the ORs and 95% confidence intervals for the association between these risk factors and mortality described in Table 2. The use of ACEIs and ARBs for the treatment of hypertension, did not show an increased risk of mortality among patients diagnosed with COVID-19.

XSL•FO
RenderX

Figure 1. Risk factors associated with mortality in patients with COVID-19. The associations of each risk factor were estimated using logistic regression models adjusted for age and sex; sex was adjusted for age, and age was adjusted for sex. ACEIs: angiotensin-converting enzyme inhibitors; ARBs: angiotensin II receptor blockers; COPD: chronic obstructive pulmonary disease.



Discussion

Principal Findings

We report the clinical characteristics and mortality rates of 23,844 patients diagnosed with COVID-19 in primary care, who constituted 6.84% of the population studied in the Central Catalan Region aged ≥16 years. This is one of only a few larger studies of COVID-19 patients conducted in primary care settings [25,27,28] and the only one to compare the characteristics and risk factors of patients diagnosed with COVID-19 with those of other patients without COVID-19 in the same health care area. Although there is no significant difference in age between patients with and without COVID-19, age is the main factor associated with mortality in patients with COVID-19, along with male sex, diabetes, dyslipidemia and heart failure. On the other hand, although obesity is more frequent among patients with COVID-19, it did not show an increase in the risk of mortality in this study. Moreover, ACEI or ARB treatment did not show an increase in the risk of mortality.

In our study, we report lower rates of comorbidity among patients diagnosed with COVID-19 compared to another study carried out in Catalonia in primary care, with rates of 33.9% for hypertension, 14.3% for diabetes, 5.9% for COPD, and 11.5% for cancer. In contrast, they reported lower rates of dyslipidemia (13.7%) and obesity (14.3%) [25]. In this study, they reported a mean age of 56.7 years, more than 7 years older than the mean age in our study; this aspect could explain the differences in the frequency of the mentioned comorbidities.

Numerous studies have shown that male sex and older age are associated with higher COVID-19–related mortality, a conclusion which our study strongly supports [14,29,30]. Although studies based on hospital data have found that the frequency of SARS-CoV-2 infection was higher in men [2,31-35], we found that women more frequently contracted COVID-19. Regarding mortality and age, we report that the death rate among patients aged ≥60 years with COVID-19 was 8.36 times higher than that among patients aged ≥60 years without COVID-19, with age being a strong predictor of mortality among people diagnosed with COVID-19, in particular those older than a certain age (≥60 years).

In our study, diabetes was more common in patients with COVID-19 and was associated with an increased risk of death. This finding is in accordance with a meta-analysis that analyzed the association between comorbid diabetes and disease severity or death [35,36]. As shown in our study, diabetes is one of the most prevalent comorbidities in patients with COVID-19, and there is a significant association with a greater severity of the illness or death in primary care; this association is in line with in-hospital mortality for patients with COVID-19, for which up to a 3-fold increased risk has been found [37]. Likewise, in our study, heart failure was more frequent among patients with COVID-19 and was associated with an increased risk of mortality. Other studies have reported similar results [30]. Although we found that patients with obesity were more common among those diagnosed with COVID-19, contrary to other studies, we did not find an increased risk of mortality in these patients [38]. In addition, surprisingly, hypertension, one

XSL•FO
RenderX

of the most important factors reported as a mortality risk factor [31], and ischemic heart disease were neither more frequent nor associated with a higher mortality risk in our study.

As for smoking, we found that smokers were more frequently infected with COVID-19, although with no increased risk of mortality. Regarding risk factors such as cancer and cerebrovascular disease, patients diagnosed with COVID-19 were not at a higher risk of mortality. In addition, our results are in keeping with those showing that ACEIs and ARBs are not significantly associated with an increased risk of death among patients with COVID-19 [2,35,39] when adjusting for age and sex. Although the frequency of the different risk factors analyzed, except in the case of dyslipidemia, was higher among patients with COVID-19, this aspect should be considered as important to take into account in the management of these patients; we have seen that mortality is linked to some of these factors, such as diabetes, dyslipidemia, and heart failure. In addition, the health characteristics of the population followed up in primary care settings can be useful to have a better knowledge of

Limitations

The study has several limitations. The diagnostics of COVID-19 registered in the Primary Care Services IT System during the early months of the pandemic were based on both clinical and polymerase chain reaction (PCR) testing. This is a pragmatic approach due to the fact that PCR testing was not fully available in primary care during the first few weeks of the pandemic. After the first months of pandemic the number of PCR tests performed increased, and for this reason the number of COVID-19 diagnoses may be underestimated during the period of our study. Although we included the majority of relevant risk factors and comorbidities associated with COVID-19 in the study, additional conditions should be considered, which may have an impact on the analysis and its interpretation. Although the coverage of patients followed up by primary care physicians is approximately 80% of the population in Catalonia, including this health care area studied, it is unlikely but possible that the remaining 20% of patients have other demographic or health characteristics that could affect the results. In addition, as can occur in other diseases, the registration and mortality of cases can be underestimated or can be affected by factors such as gender; this may modify some of the results presented.

Conclusions

At the beginning of the COVID-19 outbreak, attention was focused on the characteristics of patients diagnosed with COVID-19 who were admitted to hospital. However, after the period of our study, the attention shifted to community and primary care services, coinciding with the work overload in primary care settings and the possibility of more extensive SARS-CoV-2 testing. Our study is focused on patients in primary care with COVID-19, unlike most previous studies on COVID-19, which are based on hospital data. As we observed in our study, hypertension, one of the risk factors associated with COVID-19, was neither more frequent nor associated with higher mortality in patients with COVID-19 in primary care. We observed that women were more affected by COVID-19, unlike the majority of studies, which reported that men more frequently contracted the disease. In addition, our study did not find an associated risk between obesity and COVID-19, another risk factor associated with increased COVID-19–related mortality. Furthermore, treatment with ACEIs or ARBs was not associated with a higher mortality rate among patients infected with SARS-CoV-2. Age was the most important factor associated with mortality in patients with COVID-19.

Further studies focused on community and primary care are needed to provide new insights into SARS-CoV-2 infection and how to address outbreaks and improve health care strategies in pandemic situations.

Acknowledgments

We are grateful to the staff at the Technical and Support Area of *Gerència Territorial de la Catalunya Central* for their support during the data collection phase.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full list of diagnosis codes used in the study.
[DOCX File , 67 KB - publichealth_v7i2e25452_app1.docx]

References

1. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Feb 3;579(7798):265-269. [doi: 10.1038/s41586-020-2008-3]
2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]
3. Jee Y. WHO International Health Regulations Emergency Committee for the COVID-19 outbreak. Epidemiol Health 2020;42:e2020013 [FREE Full text] [doi: 10.4178/epih.e2020013] [Medline: 32192278]

XSL•FO
RenderX

4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-01-31]
5. Spain: WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/region/euro/country/es [accessed 2020-10-27]
6. Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA 2020 Apr 28;323(16):1612-1614 [FREE Full text] [doi: 10.1001/jama.2020.4326] [Medline: 32191259]
7. Lam PH, Milam AJ, Tou E, Dhillon NK, Toscano S, Abaalkhail N. Characteristics of mechanically ventilated patients with COVID-19 and persons under investigation negative for COVID-19 at an academic medical center: A retrospective cross-sectional pilot study. J Clin Anesth 2020 Dec;67:110029 [FREE Full text] [doi: 10.1016/j.jclinane.2020.110029] [Medline: 32858344]
8. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
9. Tsigaris P, Teixeira da Silva JA. Smoking prevalence and COVID-19 in Europe. Nicotine Tob Res 2020 Aug 24;22(9):1646-1649 [FREE Full text] [doi: 10.1093/ntr/ntaa121] [Medline: 32609839]
10. Farsalinos K, Barbouni A, Poulas K, Polosa R, Caponnetto P, Niaura R. Current smoking, former smoking, and adverse outcome among hospitalized COVID-19 patients: a systematic review and meta-analysis. Ther Adv Chronic Dis 2020 Jun 25;11:2040622320935765 [FREE Full text] [doi: 10.1177/2040622320935765] [Medline: 32637059]
11. Grundy EJ, Suddek T, Filippidis F, Majeed A, Coronini-Cronberg S. Smoking, SARS-CoV-2 and COVID-19: a review of reviews considering implications for public health policy and practice. Tob Induc Dis 2020 Jul 3;18(July):58 [FREE Full text] [doi: 10.18332/tid/124788] [Medline: 32641924]
12. Resources for tobacco use control as part of COVID-19 response. World Health Organization. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/technical-guidance/resources-for-tobacco-use-control-as-part-of-covid-19-response [accessed 2020-09-07]
13. Sanchez-Ramirez DC, Mackey D. Underlying respiratory diseases, specifically COPD, and smoking are associated with severe COVID-19 outcomes: A systematic review and meta-analysis. Respir Med 2020 Sep;171:106096 [FREE Full text] [doi: 10.1016/j.rmed.2020.106096] [Medline: 32763754]
14. Guo F. Active smoking is associated with severity of coronavirus disease 2019 (COVID-19): An update of a meta-analysis. Tob Induc Dis 2020;18:37 [FREE Full text] [doi: 10.18332/tid/121915] [Medline: 32382258]
15. Cai G, Bossé Y, Xiao F, Kheradmand F, Amos CI. Tobacco smoking increases the lung gene expression of ACE2, the receptor of SARS-CoV-2. Am J Respir Crit Care Med 2020 Jun 15;201(12):1557-1559. [doi: 10.1164/rccm.202003-0693le]
16. Leung JM, Sin DD. Smoking, ACE-2 and COVID-19: ongoing controversies. Eur Respir J 2020 Jul 19;56(1):2001759 [FREE Full text] [doi: 10.1183/13993003.01759-2020] [Medline: 32430431]
17. Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin–Angiotensin–Aldosterone System Blockers and the Risk of Covid-19. N Engl J Med 2020 Jun 18;382(25):2431-2440. [doi: 10.1056/nejmoa2006923]
18. de Abajo FJ, Rodríguez-Martín S, Lerma V, Mejía-Abril G, Aguilar M, García-Luque A, et al. Use of renin–angiotensin–aldosterone system inhibitors and risk of COVID-19 requiring admission to hospital: a case-population study. Lancet 2020 May;395(10238):1705-1714. [doi: 10.1016/S0140-6736(20)31030-8]
19. Zhang J, Dong X, Cao Y, Yuan Y, Yang Y, Yan Y, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020 Jul;75(7):1730-1741. [doi: 10.1111/all.14238] [Medline: 32077115]
20. Ferrario CM, Jessup J, Chappell MC, Averill DB, Brosnihan KB, Tallant EA, et al. Effect of Angiotensin-Converting Enzyme Inhibition and Angiotensin II Receptor Blockers on Cardiac Angiotensin-Converting Enzyme 2. Circulation 2005 May 24;111(20):2605-2610. [doi: 10.1161/circulationaha.104.510461]
21. Gurwitz D. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug Dev Res 2020 Aug 04;81(5):537-540 [FREE Full text] [doi: 10.1002/ddr.21656] [Medline: 32129518]
22. Krist AH, DeVoe JE, Cheng A, Ehrlich T, Jones SM. Redesigning primary care to address the COVID-19 pandemic in the midst of the pandemic. Ann Fam Med 2020 Jul 13;18(4):349-354 [FREE Full text] [doi: 10.1370/afm.2557] [Medline: 32661037]
23. Goodyear-Smith F, Kinder K, Mannie C, Strydom S, Bazemore A, Phillips RL. Relationship between the perceived strength of countries' primary care system and COVID-19 mortality: an international survey study. BJGP Open 2020 Sep 08;4(4):bjgpopen20X101129. [doi: 10.3399/bjgpopen20x101129]
24. Majeed A, Maile EJ, Bindman AB. The primary care response to COVID-19 in England's National Health Service. J R Soc Med 2020 Jun;113(6):208-210 [FREE Full text] [doi: 10.1177/0141076820931452] [Medline: 32521196]
25. Sisó-Almirall A, Kostov B, Mas-Heredia M, Vilanova-Rotllan S, Sequeira-Aymar E, Sans-Corrales M, et al. Prognostic factors in Spanish COVID-19 patients: A case series from Barcelona. PLoS One 2020 Aug 21;15(8):e0237960 [FREE Full text] [doi: 10.1371/journal.pone.0237960] [Medline: 32822413]

XSL•FO
RenderX

26. Coma E, Medina M, Méndez L, Hermosilla E, Iglesias M, Olmos C, et al. Effectiveness of electronic point-of-care reminders versus monthly feedback to improve adherence to 10 clinical recommendations in primary care: a cluster randomized clinical trial. BMC Med Inform Decis Mak 2019 Nov 29;19(1):245 [FREE Full text] [doi: 10.1186/s12911-019-0976-8] [Medline: 31783854]
27. Leal F, Mendes-Correa M, Buss L. A primary care approach to the COVID-19 pandemic: clinical features and natural history of 2,073 suspected cases in the Corona Sao Caetano programme, Sao Paulo, Brazil. medRxiv. Preprint posted online on June 23, 2020 2020. [doi: 10.1101/2020.06.23.20138081]
28. de Lusignan S, Dorward J, Correa A, Jones N, Akinyemi O, Amirthalingam G, et al. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. Lancet Infect Dis 2020 Sep;20(9):1034-1042 [FREE Full text] [doi: 10.1016/S1473-3099(20)30371-6] [Medline: 32422204]
29. Mallapaty S. The coronavirus is most deadly if you are older and male - new data reveal the risks. Nature 2020 Sep 28;585(7823):16-17. [doi: 10.1038/d41586-020-02483-2] [Medline: 32860026]
30. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. PLoS Med 2020 Sep 10;17(9):e1003321 [FREE Full text] [doi: 10.1371/journal.pmed.1003321] [Medline: 32911500]
31. 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. Lancet 2020 Mar;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3]
32. Docherty A, Harrison E, Green C, Pius R, Norman L, Holden KA, ISARIC4C Investigators, et al. Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. medRxiv. Preprint posted online on April 28, 2020 2020 [FREE Full text] [doi: 10.1101/2020.04.23.20076042]
33. 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):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
34. Li L, Huang T, Wang Y, Wang Z, Liang Y, Huang T, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol 2020 Jun 23;92(6):577-583 [FREE Full text] [doi: 10.1002/jmv.25757] [Medline: 32162702]
35. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahão MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun 2020 Oct 06;11(1):5009 [FREE Full text] [doi: 10.1038/s41467-020-18849-z] [Medline: 33024121]
36. Guo L, Shi Z, Zhang Y, Wang C, Do Vale Moreira NC, Zuo H, et al. Comorbid diabetes and the risk of disease severity or death among 8807 COVID-19 patients in China: A meta-analysis. Diabetes Res Clin Pract 2020 Aug;166:108346 [FREE Full text] [doi: 10.1016/j.diabres.2020.108346] [Medline: 32710998]
37. Mantovani A, Byrne CD, Zheng M, Targher G. Diabetes as a risk factor for greater COVID-19 severity and in-hospital death: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2020 Jul 24;30(8):1236-1248 [FREE Full text] [doi: 10.1016/j.numecd.2020.05.014] [Medline: 32571616]
38. Cornejo-Pareja IM, Gómez-Pérez AM, Fernández-García JC, Barahona San Millan R, Aguilera Luque A, de Hollanda A, et al. Coronavirus disease 2019 (COVID-19) and obesity. Impact of obesity and its main comorbidities in the evolution of the disease. Eur Eat Disord Rev 2020 Nov 25;28(6):799-815. [doi: 10.1002/erv.2770] [Medline: 32974994]
39. Morales D, Conover M, You S, Pratt N, Kostka K, Duarte-Salles T, et al. Renin-angiotensin system blockers and susceptibility to COVID-19: a multinational open science cohort study. medRxiv. Preprint posted online on June 12, 2020 2020 [FREE Full text] [doi: 10.1101/2020.06.11.20125849] [Medline: 32587982]

Abbreviations

ACE: angiotensin-converting enzyme
ACE2: angiotensin-converting enzyme II
ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin type II receptor blocker
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Edition
IT: information technology
OR: odds ratio
PCR: polymerase chain reaction
WHO: World Health Organization

XSL•FO
RenderX

Edited by T Sanchez; submitted 02.11.20; peer-reviewed by F Lopez Segui, P Nguyen; comments to author 27.11.20; revised version received 02.01.21; accepted 04.01.21; published 08.02.21.

Please cite as:
Mayer MA, Vidal-Alaball J, Puigdellívol-Sánchez A, Marín Gomez FX, Leis A, Mendioroz Peña J
Clinical Characterization of Patients With COVID-19 in Primary Care in Catalonia: Retrospective Observational Study
JMIR Public Health Surveill 2021;7(2):e25452
URL: http://publichealth.jmir.org/2021/2/e25452/
doi:10.2196/25452
PMID:33496668

©Miguel Angel Mayer, Josep Vidal-Alaball, Anna Puigdellívol-Sánchez, Francesc X Marín Gomez, Angela Leis, Jacobo Mendioroz Peña. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 08.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Collaborating in the Time of COVID-19: The Scope and Scale of Innovative Responses to a Global Pandemic

Theresa Bernardo[1*], DVM, MSc; Kurtis Edward Sobkowich[1*], HBSc; Russell Othmer Forrest[1], HBSc, MSc; Luke Silva Stewart[1*], HBSc; Marcelo D'Agostino[2*], MSc; Enrique Perez Gutierrez[3*], DVM, MSc, MPVM, PhD; Daniel Gillis[4*], BSc, MSc, PhD

[1]Department of Population Medicine, University of Guelph, Guelph, ON, Canada

[2]Information Systems for Health, Evidence and Intelligence for Action in Health, Pan American Health Organization and World Health Organization Regional Office for the Americas, Washington, DC, United States

[3]Health Information and Risk Assessment Unit, Pan American Health Organization and World Health Organization Regional Office for the Americas, Washington, DC, United States

[4]School of Computer Science, University of Guelph, Guelph, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Theresa Bernardo, DVM, MSc
Department of Population Medicine
University of Guelph
50 Stone Rd E
Guelph, ON, N1G 2W1
Canada
Phone: 1 519 824 4120 ext 54184
Email: theresabernardo@gmail.com

Abstract

The emergence of COVID-19 spurred the formation of myriad teams to tackle every conceivable aspect of the virus and thwart its spread. Enabled by global digital connectedness, collaboration has become a constant theme throughout the pandemic, resulting in the expedition of the scientific process (including vaccine development), rapid consolidation of global outbreak data and statistics, and experimentation with novel partnerships. To document the evolution of these collaborative efforts, the authors collected illustrative examples as the pandemic unfolded, supplemented with publications from the JMIR COVID-19 Special Issue. Over 60 projects rooted in collaboration are categorized into five main themes: knowledge dissemination, data propagation, crowdsourcing, artificial intelligence, and hardware design and development. They highlight the numerous ways that citizens, industry professionals, researchers, and academics have come together worldwide to consolidate information and produce products to combat the COVID-19 pandemic. Initially, researchers and citizen scientists scrambled to access quality data within an overwhelming quantity of information. As global curated data sets emerged, derivative works such as visualizations or models were developed that depended on consistent data and would fail when there were unanticipated changes. Crowdsourcing was used to collect and analyze data, aid in contact tracing, and produce personal protective equipment by sharing open designs for 3D printing. An international consortium of entrepreneurs and researchers created a ventilator based on an open-source design. A coalition of nongovernmental organizations and governmental organizations, led by the White House Office of Science and Technology Policy, created a shared open resource of over 200,000 research publications about COVID-19 and subsequently offered cash prizes for the best solutions to 17 key questions involving artificial intelligence. A thread of collaboration weaved throughout the pandemic response, which will shape future efforts. Novel partnerships will cross boundaries to create better processes, products, and solutions to consequential societal challenges.

(JMIR Public Health Surveill 2021;7(2):e25935) doi:10.2196/25935

KEYWORDS

crowdsourcing; artificial intelligence; collaboration; personal protective equipment; big data; AI; COVID-19; innovation; information sharing; communication; teamwork; knowledge; dissemination

XSL•FO
RenderX

Introduction

COVID-19, caused by the transmission of SARS-CoV-2, was declared a pandemic in March 2020. In the early days of the COVID-19 pandemic, Karl Gude, former director of Infographics at Newsweek magazine, wanted to use his talents to help the public protect themselves against the onslaught of the new virus. He collaborated with a nurse and an epidemiologist, using Centers for Disease Control and Prevention resources [1] as a base to create the infographic "Break the Chain of Infection," which was widely circulated on social media in March 2020 [2]. In response to demand, translations were crowdsourced into Arabic, Chinese, French, German, Italian, Japanese, Korean, Malayalam, Portuguese, and Spanish [3]. The United Nations subsequently called upon the global power of creatives to help stop the spread of COVID-19 [4,5].

Scientists have also taken advantage of unprecedented capabilities for global information sharing to accelerate their research. China sequenced the genome of the virus and shared it with the world on January 12, 2020, igniting investigation into the characteristics of the virus and enabling development of diagnostic tests. The gene sequence data was submitted for posting on Virological [6], a hub for prepublication data [7].

Similarly, in the race to combat the pandemic, researchers have posted their preliminary findings as preprints for global online scrutiny in advance of peer review and journal publication. The two most popular preprint servers, bioRxiv and medRxiv, posted nearly 3000 COVID-19 studies by May 7, 2020 [8]. Nature launched an open-source platform, The Outbreak Science Rapid PREreview, which allows scientists with ORCID IDs to submit their reviews as they read the preprints [9]. A new journal from MIT Press, Rapid Reviews: COVID-19, will use artificial intelligence (AI) to categorize preprints by discipline, novelty, and importance prior to review by humans [10]. Innovation to expedite review processes also includes peer-reviewed journals. A preprint analysis of 14 medical research journals found that average turnaround times had fallen from 117 to 60 days [8].

The JMIR journals, which are fully online, have set a precedent by partnering with the World Health Organization (WHO) on an e-collection *Theme Issue 2020: Novel Coronavirus (COVID-19) Outbreak Rapid Reports*, giving priority to the review and publication of COVID-19 articles on an ongoing basis [11]. The JMIR COVID-19 Special Issue citation library can be easily accessed via its download button. As of July 28, 2020, it included 147 papers that were screened by two of the authors to identify relevant references to supplement examples of collaboration collected by the authors throughout the pandemic, as well as through personal experience gained through the development of a COVID-19 dashboard for Canada [12].

The objective of this paper is to provide an illustrative, rather than exhaustive, overview of the range and types of collaboration stimulated by COVID-19 on a variety of fronts (research, public health information, personal protective equipment [PPE] shortages, etc) and involving individuals, academia, governments, and the private sector (Table 1). Finding the right partners for collaboration may require reaching outside existing networks to add complementary skills and perspectives [13]. Case in point, several of the authors of this paper have never met each other in person. Although forming successful new relationships entails risk and cost, it can also unlock creative potential to confront complex challenges [13].

XSL•FO
RenderX

Table 1. Overview of the scope and scale of collaboration to combat COVID-19.

Section	Scale of collaboration	Outcomes
Knowledge dissemination	Global	• Multilingual infographics • >40 million shares of cocreated articles • COVID-19 eLearning
Data		
The collection, and dissemination of COVID-19 data	Global	• Dashboards and data visualization
Sharing data among countries in the Americas	International (Western Hemisphere)	• Data sharing agreements/policies
Crowdsourcing		
The power of the crowd	Global	• Hackathons • Software development • Hive mentality problem solving
Crowdsourcing and telemedicine	Global	• Communicate, work, interact, share information, and generate and exchange knowledge
Crowdsourcing and artificial intelligence	Global	• >200,000 articles in shared data set available for data and text mining
Artificial intelligence		
Predicting the spread of a global pandemic	Global	• Predictive models assessing international transmission
Predicting those most vulnerable	Global	• Predictive models assessing those most at risk
Hardware		
Personal protective equipment	National[a] (Canada)	• 10 million face shields • 90,000 medical gowns
Ventilators	International[a] (Canada, US, Italy)	• 10,000 medical ventilators
Test development and deployment	Global	• Noninvasive saliva-based antigen test for COVID-19

[a]This refers to the examples used in the text; the World Health Organization distributed these items worldwide.

COVID-19 Public Health Information and Education

The quantity of information on COVID-19 available to the public on virtually every communication channel can be truly overwhelming, while quality varies widely from peer-reviewed research to purposeful misinformation [14]. Even renowned publications such as the Economist or the Atlantic, whose content is usually behind a paywall, have provided free access to their COVID-19 coverage [15]. On February 15, 2020, the WHO recognized that the COVID-19 outbreak was accompanied by an infodemic (an overabundance of information, only some of which is accurate) and created a framework to manage it based on an online, crowdsourced technical consultation [16].

Amid the tsunami of information, a remarkable and highly successful example of collaboration to inform the public was initiated by Tomas Pueyo, a French and Spanish writer, engineer, and businessman based in San Francisco. Familiar with exponential growth rates, he published a data-driven article on Medium titled “Coronavirus: Why You Must Act Now” on March 10, 2020, which was viewed by 40 million people and translated into 30 languages before the month ended [17].

Collaboration can take many forms. It may be the stated intent from the onset or can be stimulated by the actions of an individual. Tomas later said, “I’m nobody. I’m just a guy in the right place and right time with enough storytelling ability and analytics ability to put it together” [17]. He sounded the alarm in a manner that resonated with many. Others took up the call. By March 14, 2020, the Khan Academy, a nonprofit that aims to provide free world-class education [18], had produced a video with a detailed explanation of concepts from Pueyo’s analysis [19]. Pueyo’s [20] second Medium post on March 19 was described as a team effort involving over a dozen normal citizens working around the clock to find all relevant research and structure it into a cohesive whole.

Among the numerous online courses about COVID-19, FutureLearn had 233,575 participants enrolled as of August 2020 [21]. There are also highly specialized courses to strategically encourage expert collaboration, such as Johns

XSL•FO
RenderX

Hopkins's course on epidemiology for data scientists [22] or Stanford's course on COVID-19 and AI [23].

Data

The Collection and Dissemination of COVID-19 Data

Managing and responding to a pandemic requires getting the right information to the right people at the right time and in the right format. In times of crisis, however, quality and disaggregated data are not always readily available from standard sources such as public health or other governmental agencies. Data-sharing policies and lack of transparency might also interfere with the ability to conduct timely analyses [24,25]. In the case of SARS-CoV-2, the inability to rapidly collect robust public health data and share information on the spread of the pandemic could have constrained the advance of scientific knowledge. In particular, it could have affected the immediacy of care, the availability of PPE for frontline practitioners, and the success of public health mitigation strategies.

In response to the lack of public information in the early stages of the COVID-19 pandemic, citizens and academics took to scraping and compiling data from the internet. In early January 2020, Avi Schiffman, a then 17-year-old high school student from Washington State, United States, published a website to track the disease as it moved around the world. His site was viewed by over 2 million unique visitors by March 3 [26], growing to 3 million in another 8 days, demonstrating the public's appetite for such information [27].

Interestingly, when data began appearing on government websites, they were not presented in a consistent manner, and as the pandemic continued, there were changes in the types of data reported and the manner in which they were provided. As local, regional, and federal public health agencies came to grips with the reality of the pandemic, their data became more readily available, interoperable, and consistent (although not without discrepancies and challenges [28]). Official Canadian data published on the Government of Canada websites, for example, were not downloadable until March 30, 2020, and the list of variables available changed over time. Depending on the source of information (ie, federal or provincial) and the time of the query, the data could have been presented in aggregate form with new or cumulative cases; with or without case status; and possibly with age, gender, race, or socioeconomic status [29]. These changes meant that anyone attempting to automate analyses and visualizations would be frequently forced to update their web scrapers and code.

When waiting for improvements to provincial and federal repositories to better understand how SARS-CoV-2 was spread, researchers and journalists resorted to using the Wayback Machine to access archived snapshots of pandemic data, including journalists at Maclean's magazine who used archived data to generate a pandemic curve of daily new cases [30]. Researchers at the University of Toronto, University of Guelph, and The Hospital for Sick Children (SickKids) developed a Canadian repository of data by painstakingly scraping media sources for details about COVID-19 cases [29]. In the United States, researchers at Johns Hopkins have been at the forefront, compiling and managing a global data set of pandemic metrics [31]. Updated daily, it has been used by numerous organizations for modeling, visualization, and knowledge mobilization [31]. Other online data sources include the Centers for Disease Control and Prevention, Worldometer, and Our World in Data (to name a few).

To make these data meaningful to the public, researchers and citizen scientists published online dashboards to visualize various aspects of the pandemic, including (but not limited to) spread, daily and cumulative case and mortality counts, hospitalization rates, testing, doubling times, and positive case rates [12,32,33]. Some of the earliest dashboards were designed by the researchers behind the Johns Hopkins data set [31] and the team of University of Toronto, University of Guelph, and SickKids researchers behind the Canadian COVID-19 data set [32]. These dashboards are subsequently addressed in the "Crowdsourcing" section. Developers, however, had to design or redesign their dashboards based on how the data that supported them were obtained (ie, manually collected and managed; automatically pulled using web scrapers; or directly accessed through downloadable files, database queries, or application programming interface statements). The COVID-19 Dashboard in Canada [12] required several updates to account for changes in the formats and types of data available as the pandemic progressed.

Although there is a danger that errors may be propagated, the open nature also serves as a safeguard, as others can discover errors with fresh eyes. For example, a reanalysis of code and data found that presymptomatic infections are spread over a longer time period before symptom onset than previously thought. Thus, tracing contacts from 2 or 3 days before symptom onset may not be sufficient to find all secondary cases. Detection of this error was only possible because the original code and data were available and accessible, and the authors noted that it was highly likely that this error was propagated in derivative works [34].

Data sharing is a key input for COVID-19 models; subsequent sharing of modeling code is equally important. Although it has been said that all models are wrong but some are useful [35], their influence on public policy cannot be ignored [36]. Transparency around model assumptions, parameter estimates, and sources of data and code can allow others to check, build upon, and improve on them as additional information about the virus and its spread becomes available. Although the structure of the models does not appear to have changed since August 2020, Our World in Data provides weekly updates of four major models based on current data [37], which is made available through their code on GitHub [38-41]. Models are produced by scientists, who may or may not have experience in public health policy. Covid Act Now is a distributed team of volunteers including technologists, epidemiologists, public health experts, and public policy leaders working to provide disease intelligence, data analysis, and modeling on COVID-19 for the United States [42].

As the pandemic progressed, larger organizations and citizen scientists created symptom tracking websites such as COVIDNearYou.org and Flatten.ca, the former developed by

XSL•FO
RenderX

volunteers from Amazon, Apple, and Google [24], and the latter by Shrey Jain at the University of Toronto [43]. Smartphone and web-based data collection tools have extraordinary potential to accumulate large amounts of data in short periods of time. However, authors in the JMIR COVID-19 Special Issue have expressed a need for a global policy regarding citizen science data collection methods to avoid issues of privacy intrusion and potential harm [44].

Sharing Data Among Countries in the Americas

Designed to prevent and cope with major international public health threats, 196 countries are signatories to the International Health Regulations. Its stated purpose is "to prevent, protect against, control and provide a response to public health threats through improved surveillance, reporting and international cooperation, and to do so in ways which avoid unnecessary interference with international traffic and trade" [45]. Under Article 44, which deals with collaboration and assistance, the countries of the Americas have shared data for research purposes and to prepare for the potential importation of cases.

The Pan American Health Organization (PAHO), which serves as the Regional Office for the Americas of the WHO and as the health organization of the Inter-American System, facilitates a communication network of national focal points, providing technical cooperation and logistical support for the detection, assessment, and response to events. Between January and November 2020, the countries of the Americas exchanged 643 communications related to the COVID-19 pandemic, particularly to advise other Member States of confirmed cases and contacts during a flight (41%) and confirmed cases and contacts during a cruise or on a ship or vessel (13%), to report travel restrictions or quarantine protocols at points of entry (4%), to request information or verification of an event (9%), to report cases from mass gatherings (2%), for repatriation or evacuations of citizens from one country to another (1%), and for contacts or confirmed cases with or without a travel history (30%).

In addition, since the beginning of the pandemic, the countries of the Americas have reported the daily counts of COVID-19 cases and deaths from official government public sources, which PAHO and the WHO collates, analyzes, and publishes on its regional dashboard [46]. PAHO supports the public health authorities to monitor trends in COVID-19 cases and deaths, identify clusters especially among vulnerable populations, and guide the implementation and adjustment of targeted public health control measures. The dashboard provides geographic distribution of cases and deaths, effective reproductive numbers, and epidemiologic graphs and trends. In-depth analyses on age, gender, testing patterns, and severity is also reported. As of November 6, 2020, details were available for 83% (17,525,352/21,168,405) of total reported cases and 56% (365,540/650,705) of total reported deaths through a shared electronic line list. PAHO also supports countries providing information on the quality of COVID-19 surveillance by monitoring performance indicators such as timelines, completeness, and representativeness of surveillance data. This type of collaboration between Member States and a multilateral organization supports a meaningful interpretation of the surveillance data on a regional basis.

Crowdsourcing

The Power of the Crowd

Despite the many parallels between the COVID-19 pandemic and the Spanish flu of 1918, there stands a major distinction: the human population today communicates with unprecedented speed and effectiveness regardless of geographical distance and in the absence of face-to-face contact. This global connectedness has fostered an extraordinary response to the pandemic that vastly outpaces its predecessor just 100 years ago.

Anyone with a connection to the internet could collect information, analyze data, and develop software and hardware. As new needs arose, eager volunteers would band together to identify potential solutions and courses of action. In April 2020, in the course of a week, volunteers from tech giants such as Google, Amazon, and Apple, in collaboration with researchers at Harvard University, Boston Children's Hospital, and the University of Toronto, had developed "COVID Near You," an app allowing users to share their symptoms and view the number of potential COVID-19 cases in their community [47]. Early efforts to map the disease by the University of Toronto relied solely on voluntary online self-assessment data [48].

Another call to action was for helping make sense of the accumulating data and drawing meaningful conclusions. Some of the authors of this paper were among the first volunteers to take the raw data being offered by the Government of Canada and translate it into an informational dashboard of preliminary epidemiological analyses interpretable by all [12]. As mentioned in the "Data" section, teams from the University of Toronto in conjunction with the University of Guelph also collaborated in producing a dashboard of critical epidemiological measures [49]. As of January 2021, 28 volunteer-driven data visualization resources have been produced in Canada alone to translate raw data into plots, figures, and tables that can be interpreted by the masses [50]. The collaborative spirit in Canada is demonstrated by the more than 6800 registered volunteers and over 200 volunteer-driven projects (as of January 2021) included in the "COVID-19 Resources" project aimed at matching volunteers nationwide to projects [50].

A larger-scale crowdsourcing effort was undertaken by Google, Apple, and local health authorities in developing contact tracing and exposure notification apps for smartphones [51]. By keeping a record of the contacts a person has had, these notifications can occur retroactively should an individual test or screen positive for the virus. Smartphone contact tracing allows for a secure and constant record of contacts to be kept based on proximity, thus eliminating the need for manual recording. Recognizing the universal need for apps such as these, Google and Apple, rivals in the tech industry, formed a partnership to expedite the software development with the intention that it be adapted and modified by individual public health authorities. Crowdsourcing projects such as these are passive in nature, as they require minimal user input following the initial setup but could play a role in curbing the pandemic with sufficient uptake.

More traditional forms of crowdsourcing, in which a direct invitation to participate is made to the public by an organization,

XSL•FO
RenderX

have also been prevalent over the course of the pandemic. Hackathons are competitions in which individuals with experience in software development, design, or other computer-based skills collaborate in a sprint to propose a solution to a challenge. The COVID Global Hackathon, occurring over a 5-day long weekend in March, attracted over 18,000 volunteers from 175 countries, proposing solutions to over 1500 distinct challenges ranging from developing health-monitoring apps for smartwatches to software that connects gamers to fight social isolation [52]. The Montreal General Hospital in collaboration with McGill University challenged volunteers to design a new ventilator, and shortly thereafter, numerous teams produced designs, some at a tenth of the price of traditional machines [53].

The humanitarian efforts were not only limited to designs and theories but also featured the physical production of PPE and machines. Cooper3D, makers of antibacterial 3D printing filaments, recognized the need for alternative forms of respirators, as the global stock of N95s dwindled. Their team designed a 3D model for a new respirator, able to be printed on most at-home and industrial printers, and made the design free to the public [54]. Enthusiasts of 3D printing also jumped on the opportunity to help designing and publishing over 4000 printable files with the tag "COVID-19" to the popular community-driven platform "Thingiverse" in the first 6 months of the pandemic. These files include face masks, door pulls, and air purifiers ready to be printed by anyone with a 3D printer at home [55]. For those with a desire to help but no access to a 3D printer, "Get Us PPE" served as a switchboard, connecting people offering masks (in small or large quantities) or materials to make masks to underresourced communities. The Get Us PPE organization has facilitated the donation of over 5.5 million units of PPE as of January 2021 [56].

As the COVID-19 pandemic persisted, health officials have recognized the health risks of social isolation, and communities have had to adapt to stay connected. Online social communities have boomed as countries such as the United States have seen an increase in social media use by roughly 50% of the population [57]. These online communities not only keep individuals connected but serve as a highway for ideas and information. Reddit users, in just over 3 months, have created the fastest growing community (also known as a "subreddit") "r/coronavirus," with over 2 million members voluntarily offering information, support, and high-quality data visualizations [58]. Health officials have warned against misinformation on social media platforms, and through crowdsourcing, this Reddit community has amassed a team of moderators consisting of PhD students, virology experts, and doctors to monitor the tens of thousands of posts a day [59]. Communities of COVID-19 survivors such as "The Survivor Corps" have also emerged, allowing for discussion of experiences to aid health care workers while serving as a support system for the trauma some have endured [60].

Crowdsourcing and Telemedicine

With border closures, quarantines, and nonpharmaceutical interventions, billions of people are socially isolated. Individuals, governments, and health institutions have turned to information and communication technologies to communicate, work, interact, share information, and generate and exchange knowledge [61]. Telemedicine in particular has proved key to sustaining the continuity of health care services, especially for those with chronic noncommunicable diseases and mental health issues [62].

A crowdsourcing exercise was conducted to obtain goods, services, and ideas for telemedicine from a network of experts across the Americas. The main public good obtained was a "Tool for assessing the maturity level of health institutions to implement telemedicine services" [63], which, although it was designed for the Americas region, was quickly disseminated throughout all continents. It was created by PAHO and the Inter-American Development Bank in collaboration with institutions and experts in telemedicine and the use of information technology in public health from the Region of the Americas and Spain: Italian Hospital of Buenos Aires (PAHO and WHO Collaborating Center for Information Systems and Digital Health), Open University of Catalonia (PAHO and WHO Collaborating Center for eHealth), Center for Health Informatics, University of Illinois (PAHO and WHO Collaborating Center on Information Systems for Health), Salud.uy, Agency for eGovernment and the Information Society of Uruguay, Telemedicine Network from Brazil, the Central American Health Informatics Network, and 10 experts from the PAHO Information Systems for Health network.

The most important result from the crowdsourcing exercise was the transformation of the original idea of a "maturity assessment" into an "accreditation service." Although this was not originally envisioned, it emanated from debate with experts and national health authorities, and proved to be a powerful aid for governments and institutions that wanted to provide telemedicine services in an effective, legal, and safe manner.

Crowdsourcing and Artificial Intelligence

Crowdsourcing is a valuable cost-effective tool and, with advancements in the field of AI, citizen scientists have new abilities to help in the global COVID-19 relief efforts. It is no wonder Eric Schmidt (former Google CEO and Executive Chairman of Alphabet Inc) predicted the next US $100 billion dollar company would be one that "uses the crowd to learn," training an AI platform on crowdsourced data to the point where the AI could outperform the crowd [64].

The purpose of the Fast Healthcare Interoperability Resources (FHIR) data format standard is to make the exchange of electronic health records and health care data easily accessible to both health care providers and individuals on a wide variety of platforms, including computers, tablets, and cell phones [65]. In the face of a global pandemic, the FHIR standard helped facilitate the international exchange of health care data relating to SARS-CoV-2, creating shared open data resources that could be mined by individuals and organizations. For example, the COVID-19 Open Research Dataset Challenge (CORD-19)–on-FHIR data set for COVID-19 research (provided by the Allen Institute for AI to support the ongoing research into COVID-19) was made available for open collaboration through GitHub [66].

XSL•FO
RenderX

The initial data set consisted of over 13,202 journal articles all relating to research on SARS-CoV-2 [67]. Coordinated by The White House Office of Science and Technology Policy, the Allen Institute for AI collaborated with the Chan Zuckerberg Initiative, Georgetown University's Center for Security and Emerging Technology, Microsoft, and the National Library of Medicine of the National Institutes of Health to expand the data set to over 200,000 scientific articles, all of which were made freely available to the scientific community.

On March 16, 2020, the CORD-19 was launched, challenging experts in the field of AI to answer 17 key questions or "Tasks" that were developed in coordination with the WHO and the National Academies of Sciences, Engineering, and Medicine's Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats [68]. Using text and data mining, AI experts used the data set to answer key questions such as "What do we know about COVID-19 risk factors?" or "What do we know about diagnostics and surveillance?" A cash prize of US $1000 was awarded for each task to whomever submitted a solution that best met the evaluation criteria set out for each of the 17 questions [67].

Artificial Intelligence

Predicting the Spread of a Global Pandemic

With the domestic and international movement of billions of humans and animals annually, the COVID-19 pandemic demonstrated how a local public health event can rapidly grow to become a global emergency. Traditional public health surveillance measures have been historically slow, often relying on reporting from health care facilities being funneled through various channels before information is communicated to the public. Furthermore, the infrastructure required to support these surveillance systems (eg, diagnostic laboratories) has (historically) limited public health surveillance to only the wealthiest of nations [69]. In the current digital era, advanced early warning public health surveillance systems powered by AI offer an alternative. Combining predictive modeling with the capacity to analyze large volumes of data in a variety of formats [70], AI is making public health surveillance faster, cheaper, and smarter.

The Canadian company BlueDot operates at the intersection of AI, public health surveillance, and epidemiology, using machine learning algorithms to collect, synthesize, and assess various data sources, including news reports written in over 60 languages, airline data, and animal disease networks, to detect outbreaks and predict the transmission patterns of disease [71]. Epidemiologists then review and verify outputs from the algorithms, ensuring they are interpretable from an epidemiological perspective. The results are disseminated to a global network of partners consisting of governmental agencies, private industry, and nongovernmental organizations, enabling them to anticipate, prepare, and manage emerging disease threats [71].

At the dawn of the global COVID-19 pandemic, BlueDot was one of the first global entities to detect the emergence of a novel coronavirus (later to be named SARS-CoV-2) in China's Hubei Province and reported its detection through ProMED-mail on December 30, 2019 [72]. BlueDot proceeded to model the international spread of COVID-19 using various different sources of data from collaborators (eg, flight and passenger travel manifests) to identify the 20 most popular international destinations for flights departing from Wuhan Tianhe International Airport [72]. Using the Infectious Disease Vulnerability Index, BlueDot assessed the capacity for each of those 20 locations to manage an outbreak should the novel coronavirus expand beyond China's borders and published the first scientific paper on the international spread of COVID-19 [73].

Predicting Those Most Vulnerable

In addition to powering disease surveillance, AI is enabling new methods to detect those most vulnerable to both emerging and existing health threats. The ClosedLoop.ai platform (powered by Amazon Web Services) uses the C-19 Index, an AI predictive model to detect those most susceptible to COVID-19 [74]. The team behind the C-19 Index developed three different predictive models, each built with different degrees of accuracy and ease of implementation. The objective for each model was the same: predict the likelihood an individual will end up in the hospital within the next 3 months due to COVID-19 [74]. The "Survey" model was the simplest of the three, built using logistic regression and made accessible to the public [75]. The "Open Source" model used gradient boosted trees and had improved accuracy compared to the "Survey" model at the cost of increased user complexity. The source code for this model was uploaded to GitHub, so the health care community could adapt the model to their changing needs [75]. Finally, the "Full" model was the most accurate and most complex of the three models. It too used gradient boosted trees and was built in the ClosedLoop.ai platform and was made available free of charge [75].

Hardware

The Private Sector Retools to Meet Demand for PPE

The pandemic has also seen the corporate sector retool and form new partnerships to meet the demands of the crisis. Inksmith was a company that used 3D printing to teach children about topics in science, technology, engineering, and mathematics [76]. Their company responded to a call from the Kitchener-Waterloo Academy of Medicine in Ontario to produce face shields for health care workers. Certification from Health Canada took just 4 days from the time the company pivoted to making face shields [77]. The company produced 1 million face shields by early May 2020 [78], en route to fulfilling an order placed by the Canadian Government for 10 million face shields by the end of August 2020 [79]. Inksmith has partnered with KWArtzLab Makerspace to coordinate a community 3D printing initiative. These additional shields with 3D printed components are donated to teachers, homeless shelters, and other organizations that need PPE [76].

Mustang Survival also retooled itself to address the needs of the pandemic. This marine apparel and goods company manufactured reusable medical gowns in response to the growing need due to COVID-19 [80]. The British Colombian

XSL•FO
RenderX

company collaborated with StedFast, a Quebecois textile company, to produce level 3 medical gowns. This partnership was facilitated by Innovation, Science and Economic Development Canada, a branch of the Canadian federal government. Outdoor apparel company Arc'teryx has also joined the effort to produce medical gowns [81]. These two companies expected to produce a total of 90,000 medical gowns.

Getting a Medical Ventilator Manufactured in Canada

Early in the COVID-19 pandemic, there were concerns that the demand for mechanical ventilators in hospitals would exceed the number available [82]. It would also be difficult to source ventilators, as demand would likely exceed the abilities of existing supply chains. Ventilators for Canadians (V4C) was launched by Canadian businessman Jim Estill of Danby Appliances Inc to address these concerns by organizing a supply chain within Canada [83]. Three other Canadian entrepreneurs joined Estill early in the formation of the V4C consortium: Rick Jamieson of ABS Friction and FTI Professional Grade Inc, Paul L'Heureux of Crystal Fountains, and Scott Shawyer of JMP Solutions [84].

JMP Solutions partnered with Medical Ventilator Milano and Nobel Laureate Arthur McDonald to design a ventilator that could be manufactured in Canada. Medical Ventilator Milano is an international team of scientists and engineers out of Canada, the United States, and Italy [85]. Medical Ventilator Milano itself originated from the Global Argon Dark Matter Collaboration, an international organization looking for an invisible component of the universe known as "dark matter." A lot of their usual research involves working with gas handling and control systems; thus, Global Argon Dark Matter Collaboration decided to repurpose their expertise to develop additional ventilators to help address the need caused by the pandemic as Medical Ventilator Milano [86]. As part of Medical Ventilator Milano, Dr McDonald led a team of three Canadian labs to develop a ventilator, including Canadian Nuclear Laboratories, TRIUMF, and SNOLAB. Ultimately, JMP Solutions and Medical Ventilator Milano partnered with Vexos Inc to produce the Medical Ventilator Milano Ventilator [87]. The Medical Ventilator Milano Ventilator was approved by the US Food and Drug Administration within just 6 weeks of development.

Meanwhile, FTI Professional Grade Inc partnered with Baylis Medical to develop and manufacture 10,000 ventilators for the Canadian government [88]. These Baylis V4C-560 ventilators are based on the design of the Medtronic PB560 ventilator, which Medtronic made open-source in late March 2020 [89]. Anyone can download, use, and sell ventilators of this design until the WHO's Public Health Emergency of International Concern ends (or until October 1, 2024) [90]. The first batch of V4C-560 ventilators were approved by Health Canada and were scheduled to arrive in August of 2020 [91].

Novel Partnerships to Accelerate Test Development and Deployment

The shortage of testing capacity and supplies has been a bottleneck for the global containment of COVID-19. Researchers from industry and academia have been working fervently to develop accurate yet fast, easy, cheap, and scalable tests. On July 28, 2020, the XPRIZE Foundation in conjunction with the nonprofit OpenCovidScreen and a coalition of partners added further incentive by announcing a 6-month competition with US $5 million in prizes to accelerate development of economically viable mass screening tests to enable a safer return to work and school [92,93]. The winning teams' testing protocols will be documented "in a free, multimedia playbook that will be disseminated globally" [92]. Furthermore, the US $50 million COVID Apollo Project led by life science investors is poised to take these innovations to market [92].

Noninvasive tests (for example, saliva tests that can achieve similar accuracy to testing using nasopharyngeal swabs in people who are asymptomatic) may increase peoples' willingness to get tested and reduce frontline workers' risk of viral exposure [94,95]. SalivaDirect is an example that arose from unique circumstances with researchers at Yale University funded by the National Basketball Association and National Basketball Players Association. The test was developed to be agnostic to the equipment on which it is run, having been validated with reagents and instruments from multiple vendors, and its protocol is available as open-source to encourage widespread adoption and production [96,97]. Laboratories around the world can easily obtain the equipment required to carry out the testing without having to rely on proprietary resources. The SalivaDirect protocol is published at protocols.io, a crowdsourced resource where researchers share knowledge and assist each other [98]. The use of a common protocol will also be used to compare data from COVID-19 vaccine candidates within the global laboratory network established by the nonprofit Coalition for Epidemic Preparedness Innovations [99].

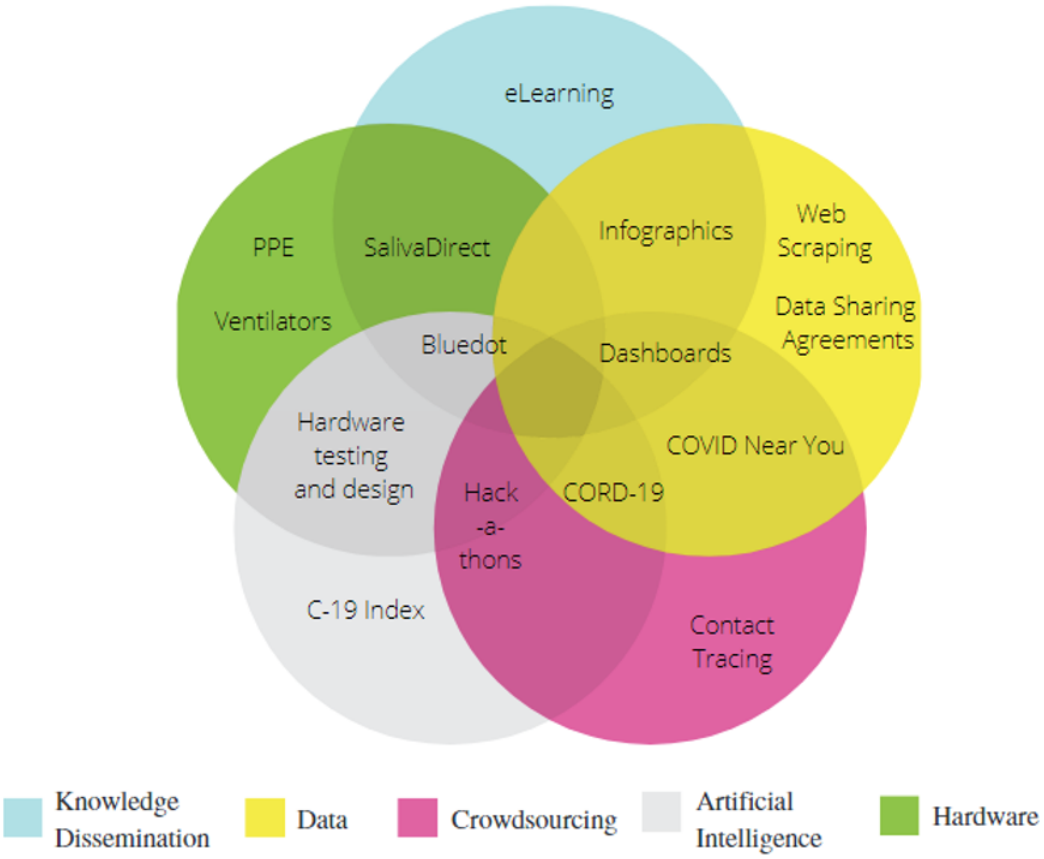
Conclusions

The thread of collaboration, cocreation, and networking that weaved throughout the pandemic response needs to be recognized as more than a series of random events and recorded to inform future efforts, pandemic or nonpandemic related. We have learned to adapt on the fly, incorporating nimbleness and agility into scientific endeavors. New ways of working together, exemplified by the surge in videoconferencing, have been thrust upon us but will have ramifications that last beyond the current pandemic. Our connectedness has allowed us to dynamically share and magnify information about the pandemic, regardless of its reliability. We also recognize that hundreds of millions of people are disconnected and that equity considerations should be a crosscutting strategy for digital transformation in health.

The pandemic has spurred the creation of large shared open repositories of health data and research, which have spawned further generations of derivative works. This dependency creates a vulnerability should there be changes or errors in the original source. Innovative processes such as crowdsourcing and competitions have been deployed to populate the repositories and to mine them for answers and solutions, using advanced analytical techniques such as AI. Examples of the intersections among the five major themes can be found in Figure 1.

XSL•FO
RenderX

Figure 1. Venn diagram of the five major themes of collaboration with examples. CORD-19: COVID-19 Open Research Dataset Challenge; PPE: personal protective equipment.



Novel partnerships, with flexible combinations of citizens, entrepreneurs, small businesses, corporations, academia, and governmental and nongovernmental organizations have crossed national boundaries and disciplinary frontiers to create new processes for working together. They have pivoted and even worked with former competitors to speed development and delivery of equipment and products in short supply. In some cases, they have taken an open-source approach with their solutions, making them available to others to reuse and modify, thus amplifying the end benefits. Born of necessity, it is hoped that this multifaceted progress will be applied not only to the pandemic but also other challenges of global proportion, such as climate change.

Conflicts of Interest

None declared.

References

1. Section 10: chain of infection. Centers for Disease Control and Prevention. 2012. URL: https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section10.html [accessed 2020-08-18]
2. Gude K, Navarro C. Coronavirus infographic. Gudethinking. 2020. URL: https://www.gudethinking.com/coronavirus-infographic [accessed 2020-08-18]
3. Spohn S. Why infographics and other visual content are saving lives from COVID-19. Contently. 2020. URL: https://contently.com/2020/06/30/visual-content-infographics-covid-19/ [accessed 2020-08-18]
4. United Nations global call out to creatives - help stop the spread of COVID-19. Talenthouse. 2020. URL: https://www.talenthouse.com/i/united-nations-global-call-out-to-creatives-help-stop-the-spread-of-covid-19 [accessed 2020-08-18]
5. Collections. United Nations COVID-19 Response Creative Content Hub. 2020. URL: https://unitednations.talenthouse.com/collections [accessed 2020-08-27]
6. Latest SARS-CoV-2 coronavirus topics. Viriological. 2020. URL: https://virological.org/c/novel-2019-coronavirus/33 [accessed 2020-08-18]

XSL•FO
RenderX

7. Schnirring L. China releases genetic data on new coronavirus, now deadly. Center for Infectious Disease Research and Policy. 2020 Jan 11. URL: https://www.cidrap.umn.edu/news-perspective/2020/01/china-releases-genetic-data-new-coronavirus-now-deadly [accessed 2020-08-18]
8. Kwon D. How swamped preprint servers are blocking bad coronavirus research. Nature 2020 May;581(7807):130-131. [doi: 10.1038/d41586-020-01394-6] [Medline: 32382120]
9. Johansson MA, Saderi D. Open peer-review platform for COVID-19 preprints. Nature 2020 Mar;579(7797):29. [doi: 10.1038/d41586-020-00613-4] [Medline: 32127711]
10. Heidt A. New journal to publish reviews of COVID-19 preprints. The Scientist. 2020. URL: https://www.the-scientist.com/news-opinion/new-journal-to-publish-reviews-of-covid-19-preprints-67675 [accessed 2020-08-18]
11. E-collection 'JMIR Theme Issue 2020: COVID-19 Special Issue'. Journal of Medical Internet Research. URL: https://www.jmir.org/themes/923 [accessed 2020-08-18]
12. Sobkowich KE, Bernardo TM, Gillis D. Monitoring COVID-19 in Canada. URL: https://kurtissobkowich.shinyapps.io/MonitoringCOVID-19inCanada/ [accessed 2020-08-18]
13. Dahlander L, Wallin M. Why now is the time for “Open Innovation”. Harvard Business Review. 2020. URL: https://hbr.org/2020/06/why-now-is-the-time-for-open-innovation [accessed 2020-08-27]
14. Cuan-Baltazar J, 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]
15. Lees N. Scientific research on the coronavirus is being released in a torrent. The Economist. 2020. URL: https://www.economist.com/science-and-technology/2020/05/07/scientific-research-on-the-coronavirus-is-being-released-in-a-torrent [accessed 2020-10-18]
16. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
17. Duan M. Tomás Pueyo, MBA ’10 on his viral post, “Coronavirus: Why You Must Act Now”. Stanford Graduate School of Business. 2020. URL: https://www.gsb.stanford.edu/experience/news-history/tomas-pueyo-mba-10-his-viral-post-coronavirus-why-you-must-act-now [accessed 2020-08-18]
18. Khan Academy. URL: https://www.khanacademy.org/ [accessed 2020-08-18]
19. Khan Academy. Estimating actual COVID 19 cases (novel corona virus infections) in an area based on deaths. YouTube. 2020. URL: https://www.youtube.com/watch?v=mCa0JXEwDEk&vl=en [accessed 2020-08-18]
20. Pueyo T. Coronavirus: the hammer and the dance. Medium. 2020. URL: https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56 [accessed 2020-08-18]
21. London School of Hygiene and Tropical Medicine. COVID-19: tackling the novel coronavirus. FutureLearn. URL: https://www.futurelearn.com/courses/covid19-novel-coronavirus [accessed 2020-08-18]
22. D'Souza G, Gurley E, Lessler J. Fighting COVID-19 with epidemiology: a Johns Hopkins teach-out. Coursera. 2020. URL: https://www.coursera.org/learn/covid19-epidemiology [accessed 2020-08-18]
23. COVID-19 and AI: a virtual conference. Stanford University Human-Centered Artificial Intelligence. 2020. URL: https://hai.stanford.edu/video-archive-covid-19-and-ai-virtual-conference [accessed 2020-08-18]
24. Farr C. How volunteers from tech companies like Amazon, Apple and Google built a coronavirus-tracking site in six days. CNBC. 2020. URL: https://www.cnbc.com/2020/03/23/coronavirus-tracking-site-built-by-amazon-apple-google-volunteers.html [accessed 2020-10-14]
25. McLennan S, Celi LA, Buyx A. COVID-19: putting the general data protection regulation to the test. JMIR Public Health Surveill 2020 May 29;6(2):e19279 [FREE Full text] [doi: 10.2196/19279] [Medline: 32449686]
26. Schlosser K. High school student near Seattle builds website to serve as a leading place for coronavirus information. GeekWire. 2020. URL: https://www.geekwire.com/2020/high-school-student-near-seattle-builds-website-serve-leading-place-coronavirus-information/ [accessed 2020-10-14]
27. Meyer J. Teen creates real-time COVID-19 tracking website that millions of people are using. Barrie 360. 2020. URL: https://barrie360.com/teen-creates-real-time-covid-19-tracking-website-that-millions-of-people-are-using/ [accessed 2020-10-14]
28. Assessment of the CDC’s new COVID-19 data reporting. The COVID Tracking Project. 2020 May. URL: https://covidtracking.com/documents/CDC_Report_CTP.pdf [accessed 2020-10-14]
29. Gillis D, Sobkowich K, Bernardo T. Better data, more lives saved. Canadian Science Policy Centre. 2020. URL: https://sciencepolicy.ca/news/better-data-more-lives-saved [accessed 2020-10-14]
30. Treble P. This coronavirus chart shows what Canada is up against in trying to 'flatten the curve'. Macleans.ca. 2020. URL: https://www.macleans.ca/society/health/this-coronavirus-chart-shows-canada-is-not-flattening-the-curve-at-least-not-yet/ [accessed 2020-10-19]
31. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]

XSL•FO
RenderX

32. Berry I, Soucy JR, Tuite A, Fisman D, COVID-19 Canada Open Data Working Group. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. CMAJ 2020 Apr 14;192(15):E420 [FREE Full text] [doi: 10.1503/cmaj.75262] [Medline: 32392510]
33. Gillis D. Visualizing COVID-19. URL: https://danielgillis.shinyapps.io/COVID/ [accessed 2020-10-14]
34. Ashcroft P, Huisman JS, Lehtinen S, Bouman JA, Althaus CL, Regoes RR, et al. COVID-19 infectivity profile correction. Swiss Med Wkly 2020 Jul 27;150:w20336 [FREE Full text] [doi: 10.4414/smw.2020.20336] [Medline: 32757177]
35. Holmdahl I, Buckee C. Wrong but useful - what Covid-19 epidemiologic models can and cannot tell us. N Engl J Med 2020 Jul 23;383(4):303-305. [doi: 10.1056/NEJMp2016822] [Medline: 32412711]
36. Enserink M, Kupferschmidt K. Mathematics of life and death: How disease models shape national shutdowns and other pandemic policies. Science. 2020. URL: https://www.sciencemag.org/news/2020/03/mathematics-life-and-death-how-disease-models-shape-national-shutdowns-and-other [accessed 2020-10-22]
37. Giattino C. How epidemiological models of COVID-19 help us estimate the true number of infections. Our World in Data. 2020. URL: https://ourworldindata.org/covid-models [accessed 2020-10-22]
38. London School of Hygiene and Tropical Medicine. centre for the mathematical modelling of infectious diseases. GitHub. 2020. URL: https://github.com/cmmid [accessed 2020-10-22]
39. Gu Y. youyanggu / covid19_projections. GitHub. 2020. URL: https://github.com/youyanggu/covid19_projections [accessed 2020-10-22]
40. Institute for Health Metrics and Evaluation. Open source public health tools and research projects created by teams at the IHME. GitHub. 2020. URL: https://github.com/ihmeuw [accessed 2020-10-22]
41. MRC Centre for Global Infectious Disease Analysis. mrc-ide / covid-sim. GitHub. 2020. URL: https://github.com/mrc-ide/covid-sim [accessed 2020-10-22]
42. Covid Act Now. 2020. URL: https://covidactnow.org [accessed 2020-10-22]
43. Friedman G. How bad is COVID-19 in your area? University students launch heat map to track virus across Canada. National Post. 2020. URL: https://nationalpost.com/news/world/how-bad-is-covid-19-in-your-area-university-students-launch-heat-map-to-track-virus-across-canada [accessed 2020-10-21]
44. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]
45. World Health Organization. International Health Regulations (2005), 2nd edition. Geneva: World Health Organization; 2008.
46. Coronavirus Disease (COVID-19) pandemic. Pan American Health Organization. 2020. URL: http://www.paho.org/en/topics/coronavirus-infections/coronavirus-disease-covid-19-pandemic [accessed 2020-11-17]
47. Boston Children's Hospital, Harvard Medical School, University of Toronto. COVID Near You. URL: https://covidnearyou.org/ca/en-CA/ [accessed 2020-08-18]
48. Flatten. 2020. URL: https://flatten.ca/?lang=en [accessed 2020-08-18]
49. Soucy JPR, Berry I. COVID-19 in Canada. 2020. URL: https://art-bd.shinyapps.io/covid19canada/ [accessed 2020-10-22]
50. Resources for researchers and clinicians responding to the COVID-19 epidemic in Canada. COVID-19 Resources Canada. 2020. URL: https://covid19resources.ca/ [accessed 2020-08-18]
51. Exposure notifications: using technology to help public health authorities fight COVID-19. Google. 2020. URL: https://www.google.com/intl/en_us/covid19/exposurenotifications/ [accessed 2020-08-18]
52. COVID-19 Global Hackathon. 2020. URL: https://covidglobalhackathon.com/ [accessed 2020-08-18]
53. Nair R. UBC students among final 10 teams in global competition to create low-cost ventilator. CBC. 2020. URL: https://www.cbc.ca/news/canada/british-columbia/low-cost-ventilator-ubc-1.5543379 [accessed 2020-08-18]
54. #HackThePandemic. Copper 3D. URL: https://copper3d.com/hackthepandemic/ [accessed 2020-08-18]
55. Thingiverse. URL: https://www.thingiverse.com/search?q=covid-19&dwh=255e6e02cf098ba&type=things&sort=relevant [accessed 2020-08-18]
56. Get Us PPE. 2020. URL: https://getusppe.org/ [accessed 2020-08-18]
57. US Adults Who Are Using Social Media More Since the Coronavirus Pandemic, March-May 2020 (% of respondents). eMarketer. 2020. URL: https://www.emarketer.com/chart/236493/us-adults-who-using-social-media-more-since-coronavirus-pandemic-march-may-2020-of-respondents [accessed 2020-08-18]
58. Whalen A. Coronavirus Reddit page becomes fastest growing subreddit with almost 400,000 new users in a week. Newsweek. 2020. URL: https://www.newsweek.com/coronavirus-reddit-live-updates-news-subreddit-covid-1490536 [accessed 2020-08-18]
59. Betuel E. Meet the mods behind the fastest-growing subreddit: r/Coronavirus. Inverse. 2020. URL: https://www.inverse.com/mind-body/the-mods-behind-the-fastest-growing-covid-19-subreddit [accessed 2020-08-18]
60. Survivor Corps. URL: https://www.survivorcorps.com [accessed 2020-08-18]
61. The potential of frequently used information technologies during the pandemic. Pan American Health Organization. 2020. URL: https://iris.paho.org/bitstream/handle/10665.2/52021/Factsheet-TICs_eng.pdf?sequence=16&isAllowed=y [accessed 2020-11-15]

XSL•FO
RenderX

62. Digital health: a strategy to maintain health care for people living with noncommunicable diseases during COVID-19. Pan American Health Organization. 2020. URL: https://iris.paho.org/bitstream/handle/10665.2/52543/PAHOEIHISCOVID-19200015_eng.pdf?sequence=1&isAllowed=y [accessed 2020-11-15]
63. COVID-19 and telemedicine. Pan American Health Organization. 2020. URL: https://www.paho.org/ish/index.php/en/telemedicine [accessed 2020-11-15]
64. Kharpal A. Alphabet's Eric Schmidt predicts what will spark the next $100 billion firm. CNBC. 2016. URL: https://www.cnbc.com/2016/06/15/alphabets-eric-schmidt-predicts-what-spark-the-next-100-billion-firm.html [accessed 2020-10-19]
65. Fast Healthcare Interoperability Resources. Wikipedia. 2020. URL: https://en.wikipedia.org/w/index.php?title=Fast_Healthcare_Interoperability_Resources&oldid=979793084 [accessed 2020-10-19]
66. FHIRCat. fhircat / CORD-19-on-FHIR. GitHub. 2020. URL: https://github.com/fhircat/CORD-19-on-FHIR [accessed 2020-10-19]
67. Jiang G. COVID-19 Open Research Dataset Challenge (CORD-19). Kaggle. 2020. URL: https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge [accessed 2020-10-19]
68. Office of Science and Technology Policy, The White House of the United States. Call to action to the tech community on new machine readable COVID-19 dataset. Montréal InVivo. 2020. URL: https://web.archive.org/web/20210120130121/https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/ [accessed 2021-02-03]
69. Edelstein M, Lee LM, Herten-Crabb A, Heymann DL, Harper DR. Strengthening global public health surveillance through data and benefit sharing. Emerging Infect Dis 2018 Jul;24(7):1324-1330. [doi: 10.3201/eid2407.151830]
70. Wong ZS, Zhou J, Zhang Q. Artificial intelligence for infectious disease big data analytics. Infect Dis Health 2019 Feb;24(1):44-48. [doi: 10.1016/j.idh.2018.10.002] [Medline: 30541697]
71. BlueDot. URL: https://bluedot.global/ [accessed 2020-10-06]
72. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. J Travel Med 2020 Mar 13;27(2) [FREE Full text] [doi: 10.1093/jtm/taaa008] [Medline: 31943059]
73. Minevich M. 15 companies that are saving the world from Covid-19. Forbes. 2020. URL: https://www.forbes.com/sites/markminevich/2020/06/25/15-companies-that-are-saving-the-world-from-covid-19/ [accessed 2021-01-10]
74. C-19 Index. ClosedLoop.ai. URL: https://closedloop.ai/c19index/ [accessed 2020-08-19]
75. DeCaprio D, Gartner J, McCall CJ, Burgess T, Kothari S, Sayed S. Building a COVID-19 Vulnerability Index. medRxiv. Preprint posted online March 21, 2020. [doi: 10.1101/2020.03.16.20036723]
76. Rogers tv. Waterloo Region: InkSmith - KWArtzlab face shield collaboration - April 23, 2020. YouTube. 2020 Apr 23. URL: https://www.youtube.com/watch?time_continue=71&v=R_72KsW5uCI&feature=emb_logo [accessed 2020-08-19]
77. Day-by-day breakdown: InkSmith's ultimate COVID-19 pivot. Waterloo EDC Blog. 2020. URL: https://blog.waterlooedc.ca/breakdown-inksmith-ultimate-covid19-pivot [accessed 2020-08-19]
78. The Canadian Shield reaches key milestone with one millionth PPE face shield manufactured. The Canadian Shield. 2020. URL: https://www.canadianshieldppe.ca/blogs/news/the-canadian-shield-one-millionth-shield [accessed 2020-08-19]
79. The Canadian Shield fulfills federal contract for 10 million medical face shields. The Canadian Shield. URL: https://www.canadianshieldppe.ca/blogs/news/the-canadian-shield-fulfills-federal-contract-for-10-million-medical-face-shields [accessed 2020-10-22]
80. Innovation, Science and Economic Development Canada. Mustang Survival and the Government of Canada fight COVID-19. YouTube. 2020. URL: https://www.youtube.com/watch?v=FDVPUmlp8ac [accessed 2020-08-19]
81. From Alpha SV to PPE. The Bird Blog - Arc'teryx. 2020. URL: https://blog.arcteryx.com/medical-gowns/ [accessed 2020-10-19]
82. Chung E, Ghebreslassie M. Hospitals scramble to secure more ventilators amid coronavirus outbreak. CBC. 2020. URL: https://www.cbc.ca/news/health/ventilator-shortage-1.5503541 [accessed 2020-08-18]
83. Ventilators for Canadians. URL: https://ventilatorsforcanadians.com/ [accessed 2020-08-18]
84. Ventilators for Canadians partnering with Baylis Medical to Manufacture Ventilators in Support of COVID-19 Response. Cision Canada. URL: https://www.newswire.ca/news-releases/ventilators-for-canadians-partnering-with-baylis-medical-to-manufacture-ventilators-in-support-of-covid-19-response-886430200.html [accessed 2021-02-03]
85. JMP Solutions plays a key role in Canadian ventilator manufacturing team to aid in the ongoing fight against COVID-19. JMP Solutions. 2020. URL: https://www.jmpsolutions.com/jmp-solutions-plays-a-key-role-in-canadian-ventilator-manufacturing-team-to-aid-in-the-ongoing-fight-against-covid-19-jmp-solutions/ [accessed 2020-08-18]
86. Who we are. Milano Ventilatore Meccanico. 2020. URL: https://mvm.care/who-we-are-en/ [accessed 2020-08-19]
87. MVM Ventilator product overview. Vexos. 2020. URL: https://www.vexos.com/mvm-ventilator [accessed 2020-08-19]
88. Baylis Medical Company Inc. Baylis Medical Partners with Ventilators for Canadians to manufacture ventilators in support of COVID-19 pandemic response. Cision. 2020. URL: https://www.newswire.ca/news-releases/

XSL•FO
RenderX

baylis-medical-partners-with-ventilators-for-canadians-to-manufacture-ventilators-in-support-of-covid-19-pandemic-response-827412147.html [accessed 2020-08-19]
89. Ventilators for Canadians. Ventilators for Canadians receive Health Canada approval for COVID-19 ventilators. Cision. 2020. URL: https://www.newswire.ca/news-releases/ventilators-for-canadians-receive-health-canada-approval-for-covid-19-ventilators-879168864.html [accessed 2020-08-19]
90. Mantey D. Medtronic open sources ventilators. Industrial Equipment News. 2020. URL: https://www.ien.com/product-development/video/21125741/medtronic-open-sources-ventilators [accessed 2020-08-18]
91. Innovation, Science and Economic Development Canada. The Ventilators for Canadians consortium and the Government of Canada fight COVID-19. YouTube. 2020. URL: https://www.youtube.com/watch?v=hFjZg3gPyEg [accessed 2020-08-19]
92. Giali J. To safely reopen, we need testing that's frequent, fast, cheap, and easy-to-use. XPRIZE. 2020. URL: https://www.xprize.org/articles/xp-rapid-covid-testing [accessed 2020-10-20]
93. MacKay MJ, Hooker AC, Afshinnekoo E, Salit M, Kelly J, Feldstein JV, et al. The COVID-19 XPRIZE and the need for scalable, fast, and widespread testing. Nat Biotechnol 2020 Sep;38(9):1021-1024 [FREE Full text] [doi: 10.1038/s41587-020-0655-4] [Medline: 32820257]
94. Craft L. Why Japanese researchers say there's "no reason" to keep doing painful COVID nasal swab tests. CBS News. 2020. URL: https://www.cbsnews.com/news/coronavirus-pcr-testing-no-reason-japan-reasearch-says-covid-saliva-tests-as-reliable/ [accessed 2020-10-20]
95. Yokota I, Shane PY, Okada K, Unoki Y, Yang Y, Inao T, et al. Mass screening of asymptomatic persons for SARS-CoV-2 using saliva. Clin Infect Dis 2020 Sep 25 [FREE Full text] [doi: 10.1093/cid/ciaa1388] [Medline: 32976596]
96. Greenwood M. Quick and affordable saliva-based COVID-19 test developed by Yale scientists receives FDA Emergency Use authorization. YaleNews. 2020. URL: https://news.yale.edu/2020/08/15/yales-rapid-covid-19-saliva-test-receives-fda-emergency-use-authorization [accessed 2020-10-20]
97. Vogels CBF, Watkins AE, Harden CA, Brackney D, Shafer J, Wang J, Yale IMPACT Research Team, et al. SalivaDirect: a simplified and flexible platform to enhance SARS-CoV-2 testing capacity. medRxiv. Preprint posted online September 28, 2020. [doi: 10.1101/2020.08.03.20167791]
98. Vogels C, Brackney DE, Kalinich CC, Ott IM, Grubaugh N, Wyllie A. SalivaDirect: RNA extraction-free SARS-CoV-2 diagnostics V.1. protocols.io. 2020 Sep 01. URL: https://www.protocols.io/view/salivadirect-rna-extraction-free-sars-cov-2-diagno-bkjgkujw [accessed 2020-10-20]
99. Kelland K. Exclusive: new global lab network will compare COVID-19 vaccines head-to-head. Reuters. 2020. URL: https://www.reuters.com/article/health-coronavirus-vaccines-comparison/exclusive-new-global-lab-network-will-compare-covid-19-vaccines-head-to-head-idUKL8N2GS1CU [accessed 2020-10-20]

Abbreviations

AI: artificial intelligence
CORD-19: COVID-19 Open Research Dataset Challenge
FHIR: Fast Healthcare Interoperability Resources
PAHO: Pan American Health Organization
PPE: personal protective equipment
V4C: Ventilators for Canadians
WHO: World Health Organization

Edited by T Sanchez; submitted 21.11.20; peer-reviewed by A Mavragani, N Finn; comments to author 06.01.21; revised version received 13.01.21; accepted 21.01.21; published 09.02.21.

Please cite as:
Bernardo T, Sobkowich KE, Forrest RO, Stewart LS, D'Agostino M, Perez Gutierrez E, Gillis D
Collaborating in the Time of COVID-19: The Scope and Scale of Innovative Responses to a Global Pandemic
JMIR Public Health Surveill 2021;7(2):e25935
URL: http://publichealth.jmir.org/2021/2/e25935/
doi:10.2196/25935
PMID:33503001

©Theresa Bernardo, Kurtis Edward Sobkowich, Russell Othmer Forrest, Luke Silva Stewart, Marcelo D'Agostino, Enrique Perez Gutierrez, Daniel Gillis. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 09.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach

Tiago de Melo[1*], PhD; Carlos M S Figueiredo[1*], PhD

Intelligent Systems Laboratory, Superior School of Technology, Amazonas State University, Manaus, Brazil

[*]all authors contributed equally

Corresponding Author:
Tiago de Melo, PhD
Intelligent Systems Laboratory
Superior School of Technology
Amazonas State University
Av Darcy Vargas, 1200
Manaus
Brazil
Phone: 55 9233487601
Email: tmelo@uea.edu.br

Abstract

Background: The COVID-19 pandemic is severely affecting people worldwide. Currently, an important approach to understand this phenomenon and its impact on the lives of people consists of monitoring social networks and news on the internet.

Objective: The purpose of this study is to present a methodology to capture the main subjects and themes under discussion in news media and social media and to apply this methodology to analyze the impact of the COVID-19 pandemic in Brazil.

Methods: This work proposes a methodology based on topic modeling, namely entity recognition, and sentiment analysis of texts to compare Twitter posts and news, followed by visualization of the evolution and impact of the COVID-19 pandemic. We focused our analysis on Brazil, an important epicenter of the pandemic; therefore, we faced the challenge of addressing Brazilian Portuguese texts.

Results: In this work, we collected and analyzed 18,413 articles from news media and 1,597,934 tweets posted by 1,299,084 users in Brazil. The results show that the proposed methodology improved the topic sentiment analysis over time, enabling better monitoring of internet media. Additionally, with this tool, we extracted some interesting insights about the evolution of the COVID-19 pandemic in Brazil. For instance, we found that Twitter presented similar topic coverage to news media; the main entities were similar, but they differed in theme distribution and entity diversity. Moreover, some aspects represented negative sentiment toward political themes in both media, and a high incidence of mentions of a specific drug denoted high political polarization during the pandemic.

Conclusions: This study identified the main themes under discussion in both news and social media and how their sentiments evolved over time. It is possible to understand the major concerns of the public during the pandemic, and all the obtained information is thus useful for decision-making by authorities.

(JMIR Public Health Surveill 2021;7(2):e24585) doi:10.2196/24585

KEYWORDS

COVID-19; Twitter; infodemiology; news; sentiment analysis; social media; Brazil; monitoring; topic modeling; entity recognition; text analysis

Introduction

Background

In December 2019, the outbreak of COVID-19 in China was reported [1]. Due to the rapid spread of SARS-CoV-2 worldwide, the World Health Organization declared a state of emergency. Recent research has confirmed that during the current pandemic, the number of infected people can double every 7 days, and each patient can spread the virus to 2.2 other people on average [2]. In Latin America, there were 937,974 cases of COVID-19 and 49,139 confirmed deaths up to May 31, 2020 [3]. In this region, Brazil is the country that is most

XSL•FO
RenderX

affected by the disease. According to the abovementioned report [3], there were 465,166 cases and 27,878 deaths in Brazil.

In past pandemic outbreaks, information exchange was relatively slow. However, with the popularization of the internet, 3.7 billion people worldwide (approximately 49.7% of the world's population) commonly use web-based information [4], and these people rely on two main sources of web-based data, namely news media web sites and social media. Using these media, people keep themselves informed about treatments, prevention, and cases, and they also participate in debates about the impact of the COVID-19 outbreak on their lives.

News media web sites are used to report crisis situations worldwide. The articles on these sites are written by journalists and subject matter experts; therefore, people trust these sources of data. However, these channels failed to keep pace with the spread of the outbreak of COVID-19 [5], and many news media channels incorrectly stated that either the pandemic would not affect countries other than China or the virus was less dangerous than influenza [6]. This coverage of the pandemic had repercussions after the spread of the disease became severe and global [7].

On the other hand, social media is a well-known channel for news and information in the timely media environment, with one in three people worldwide engaging in social media and two-thirds of people using the internet [8]. This is particularly true for health issues, with one-third of people reporting that social media are an important source of information [9]. However, recent studies have indicated that social media has also become an environment for misinformation on COVID-19 [10,11].

Currently, almost 70% of Brazilians use the internet, 90% of them access the web on a daily basis, and Brazil is the country in the western hemisphere whose residents spend most time on social media per day [12]. Thus, we envisioned that Brazil is a strategic country to study the impact of the COVID-19 pandemic through web-based media. We proposed to perform this task by applying an improved topic model and sentiment analysis methodology to news and social media compared to related work. The results of this study can help researchers understand what information about the pandemic is relevant and how people are reacting to it. Thus, this information can be useful for researchers and authorities to identify important aspects of the pandemic that can guide better action and communication policies toward the population.

Prior Work

Traditional news media focus substantial interest on health issues, especially when a new disease emerges. A number of researchers have exploited the importance of understanding the depiction of health issues in the news media. For instance, Washer [13] investigated how severe acute respiratory syndrome (SARS) was depicted in newspapers in the United Kingdom. Dias et al [14] presented a study that analyzed the representations of mental health and its treatment and the impact of the 2008 economic crisis. Ribeiro et al [15] investigated how the Zika outbreak was reported in two major newspapers in Brazil. Liu et al [16] investigated the patterns of media-directed health communications as well as the role of the media in the COVID-19 crisis in China. Gozzi et al [17] investigated the media coverage and collective internet response to the COVID-19 pandemic in four countries: Italy, the United Kingdom, the United States, and Canada.

These related studies focused on how traditional news media react to health events and the characterization of their reports. Our work differs by focusing on the analysis of social media and comparing it with traditional news media, as we are interested in showing the impact of the COVID-19 pandemic on people's lives.

The research community is also interested in correlating pandemic events with information shared by people on social networks, especially Twitter. Several examples show how useful information can be extracted from social media to help understand pandemic behavior but also to enable organizations to act to improve people's quality of life. For instance, Chew and Eysenbach [18] presented the first study using Twitter data to evaluate the H1N1 influenza pandemic in 2009, showing that this social media platform disseminated news from credible sources but also shared users' opinions and experiences. Comito et al [19] presented a study to evaluate the effectiveness of Twitter-based influenza as surveillance information. Ahmed et al [20] investigated the content shared by Twitter users during the Zika virus outbreak in 2016, and they showed that people's fears were intensified due to false news. More recently, Lwin et al [21] examined worldwide trends of several types of emotions and the narratives underlying those emotions during the COVID-19 pandemic. Abd-Alrazaq et al [22] presented a topic study of tweets in English, and Huang et al [23] analyzed the characteristics of suspected or laboratory-confirmed patients with COVID-19 who asked for help on social media; they found that it is possible to identify common patient characteristics in advance to accelerate emergency responses.

Although several previous studies have separately assessed news coverage and social media in pandemic events, only a few of them have compared news coverage with social media (in contrast to other disasters [24-28]). Particularly, Kim et al [29] investigated topic coverage and sentiment dynamics of two different media sources, Twitter and news publications, on the health issue of Ebola virus. The results reported in their paper indicate that Twitter and news media present two distinct points of view. In other work, Mondragon et al [30] presented a study on how Ebola virus was transformed from purely scientific knowledge to public thinking through media communication.

Our work follows a similar approach to that of [29] in that we use topic analysis and sentiment polarity on each data set. However, we have extended and improved the proposed methodology by generating the topic model from all data sets, aggregating them in meaningful themes, and analyzing sentiments from documents classified according to themes, which resulted in a better and more meaningful sentiment timeline.

Goals

In this study, we describe a methodological approach to analyze the content of two main sources of web-based data to better

XSL•FO
RenderX

understand the focus of each channel in disseminating information on COVID-19. Recent work in the literature (eg, [18,19]) has presented methodologies focused on social media and news comparisons based on topic models [31] and sentiment analysis [32]. We have contributed to the literature by extending these methodologies in addition to introducing specific analysis to understand the COVID-19 pandemic in Brazil. To the best of our knowledge, this is the first study to compare news and social media data in Portuguese.

The three main research questions that we are addressing in this study are:

- RQ1: Does social media cover similar categories and types of topics to traditional news media about the COVID-19 pandemic?
- RQ2: Do news web sites and social media mention the same types of entities?
- RQ3: Are there differences in the sentiments of Twitter posts and news articles? Does the degree of sentiments change over time?

To answer these questions, we collected and analyzed data from the main news media web site from Brazil, namely Universo Online (UOL), and Twitter. Twitter is a very popular social media platform worldwide, and UOL is a very popular portal for news in Brazil. We proposed the generation of topic models for each data collection, their grouping in themes for sentiment analysis, the observation of theme-sentiment evolution on a time scale, and the extraction of named entities. One challenging aspect of this research is the adaptation of the proposed methods to the Brazilian Portuguese language; therefore, we adopted some tools and developed specific trained models. By comparing all the features extracted from news and social media data sets, we present some perceptions on how the COVID-19 pandemic is affecting Brazil.

Methods

Data Collection

We collected news articles and tweets related to COVID-19 in the Portuguese language from January to May 2020. To collect the tweets, we used the TwitterScraper Python library [33] with the option *--lang* to retrieve tweets only in Portuguese. The metadata of a tweet contains a location entry; however, we noted that very few users fill in this field, and many of those users fill in nonstandard labels. Although Brazil is not the only country in which Portuguese is spoken, it represents 75% of the world's speakers, and upon manually checking the tweets that contained the user's location, we observed that only 4% were from people who spoke Portuguese and were not in Brazil. Thus, we consider that these data statistically represent this country. We also filtered the tweets with the following set of most frequently appearing keywords obtained from Google Trends for COVID-19–related topics: *azitromicina* (azithromycin), *cloroquina* (chloroquine), *comorbidade* (comorbidity), *corona*, *coronavirus*, *covid*, *covid19*, *covid-19*, di*stanciamento social* (social distancing), *ivermectina* (ivermectin), *lockdown*, *hidroxicloroquina* (hydroxychloroquine), *pandemia* (pandemic), *quarentena* (quarantine), and *tamiflu*. This search for keywords was executed at the beginning of March 2020. The final Twitter collection did not contain any retweets, and it contained 1,597,934 tweets posted by 1,299,084 users.

Regarding news collection, we gathered all the articles published in the COVID-19 section from the UOL portal. We chose UOL because this portal is responsible for publishing the *Folha de São Paulo*, which is the leading Brazilian daily newspaper by circulation [34]. In this collection, we gathered all web pages related to COVID-19; therefore, we did not need to use a set of keywords. The final news collection contained 18,413 articles.

Characterization of the Collected Data

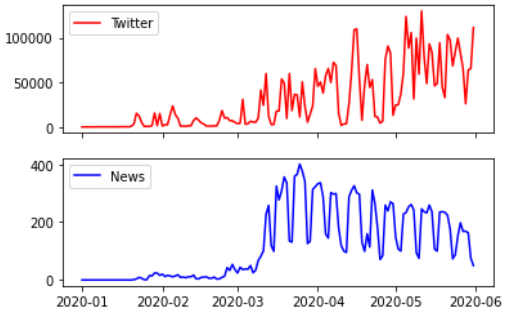
To better understand the collected data, we evaluated the statistics of the number of tokens published in each data source over time, where a token is an individual occurrence of a linguistic unit in speech or writing. The monthly distributions of the total number and percentage of tokens from both data sets are described in Table 1. One major difference between tweets and news articles is that the news presented a sharp increase in the number of tokens that decreased in May, whereas a persistent increase occurred on Twitter during the entire period of time. These findings indicate that Twitter users remained increasingly interested in the COVID-19 pandemic, while the news media began to lose interest in the month of May. This can also be observed in Figure 1, which shows the distributions of the collected data by day over five months. Additionally, this figure shows that the number of posts sharply increased at the end of March, when the first death from COVID-19 was announced in Brazil.

Table 1. Monthly statistics of tokens in news articles and tweets.

Tokens	Month, n (%)				
	January	February	March	April	May
News articles					
Unique tokens (n=134,845)	4149 (3.07)	10,093 (7.48)	38,300 (28.40)	43,327 (32.13)	38,976 (28.90)
Total tokens (n= 2,616,002)	14,550 (0.55)	62,008 (2.37)	792,175 (30.28)	953,441 (36.44)	793,828 (30.34)
Tweets					
Unique tokens (n=407,406)	16,100 (3.95)	29,511 (7.24)	84,713 (20.79)	122,169 (29.98)	154,913 (38.02)
Total tokens (n=14,155,346)	106,619 (0.75)	284,012 (2.00)	2,191,226 (15.48)	4,420,658 (31.23)	7,152,831 (50.53)

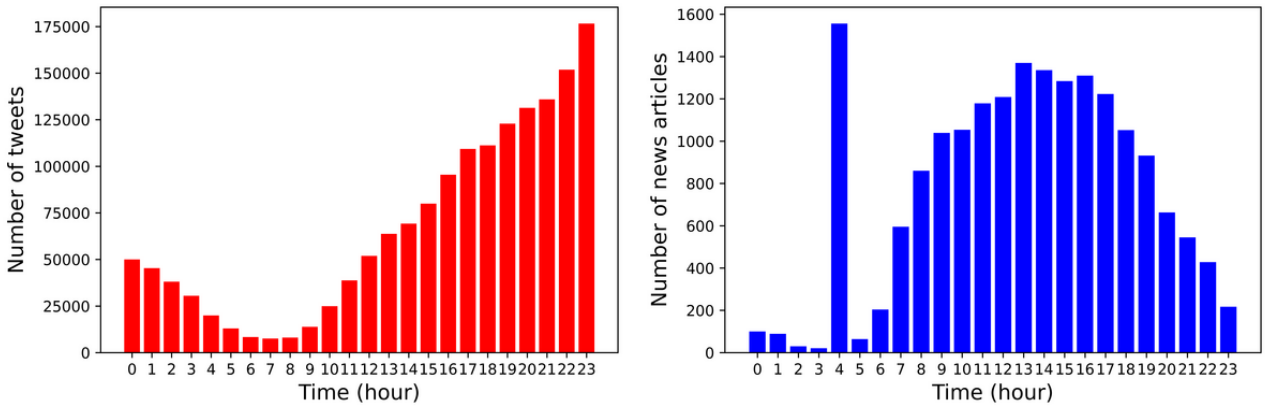
XSL•FO
RenderX

Figure 1. Distribution of tweets and news reports by day.



We also noted that the variation in the density of news and tweets over time (Figure 2) shows the 24-hour temporal distributions of news and tweets. The highest rates of tweeting occurred at night, while the highest rates of news posting occurred between noon and 6 PM. Interestingly, we verified that a peak in the news data source appeared at 4 AM on different days. This is probably due to an automated action to publish news for that morning.

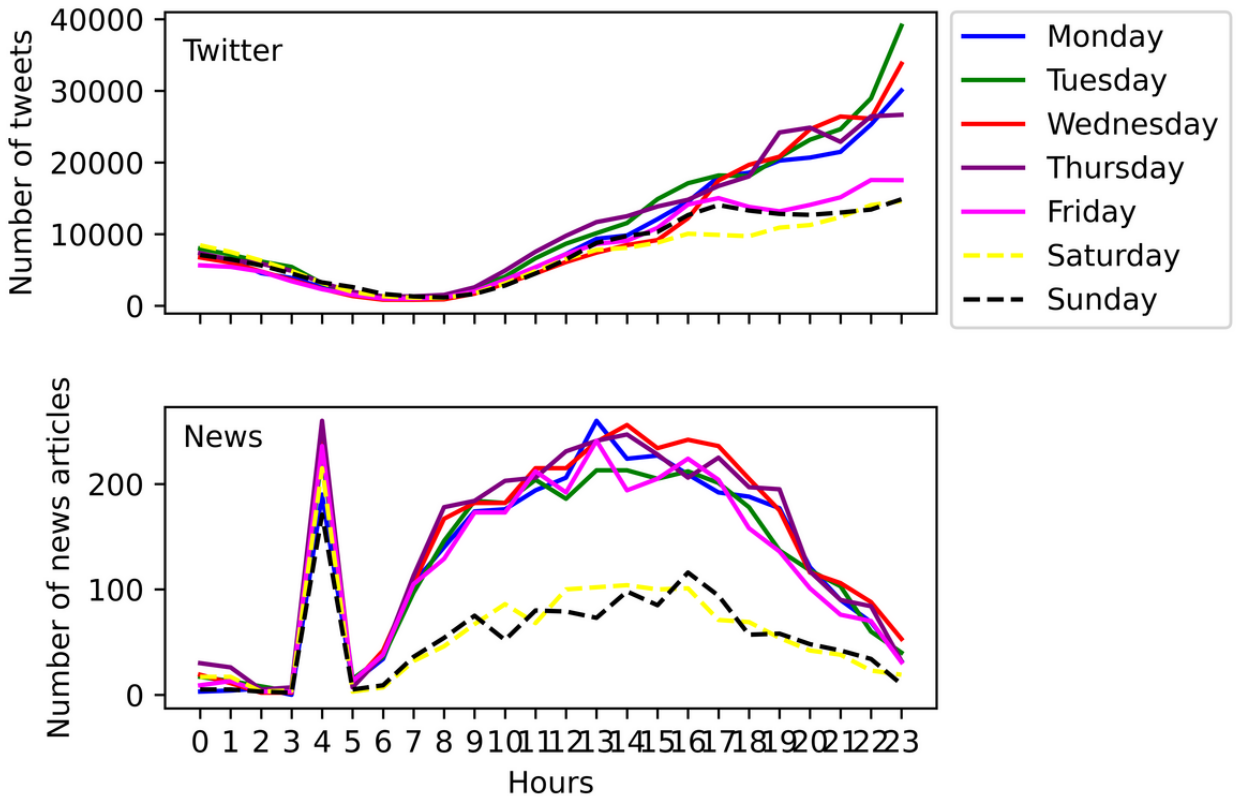
Figure 2. The 24-hour temporal distributions of posted tweets and news articles.



Although the distribution trend is relatively consistent on each day of the week, the activity was significantly different between work periods and holidays. As shown in Figure 3, during work periods, the number of tweets and news publications was significantly higher. The daily activity during holidays was quite different from that during work periods. One of the main differences observed between data sources is that Twitter users posted on the weekend until early afternoon at the same pace as during work periods, while news articles were posted at a much slower pace during holiday periods than during work periods.

XSL•FO
RenderX

Figure 3. The 24-hour temporal distribution of posts on Twitter and Universo Online.



Data Preprocessing

The collected data contained a large amount of noise that needed to be filtered out before further analysis. First, we tokenized the text, and then we adopted the following steps to normalize the texts:

1. Lowercase: All tokens were converted to lowercase. By doing this, identical tokens were merged and the dimensionality of the text was reduced.
2. URL removal: People post URLs with text to provide supporting information about the text. These URL links became noisy data during the analysis. All URL links in the texts were replaced by a space.
3. Username: Some Twitter usernames in texts start with the symbol @ and are used to tag other users. In our investigation, we were focusing on COVID-19 and not on any targeted person; therefore, we replaced all usernames with white spaces. This step was applied only to tweets.
4. Punctuation: We removed all the punctuation symbols from the collected data because they did not contribute to our evaluation.
5. Stop words: Stop words refer to the most common words used in text. We eliminated the Portuguese stop words that contributed less to our evaluation. We used a list of Portuguese stop words provided by the Natural Language Toolkit framework.
6. White spaces: We removed all the extra white spaces between tokens or at the end of lines or paragraphs of the text.

In addition to the above steps, we used lemmatization and stemming in the preprocessing of the text. However, the results were not satisfactory because there are few tools with these functions in the Portuguese language, and these tools present results with low accuracy.

Topic Modeling and Topic Similarity

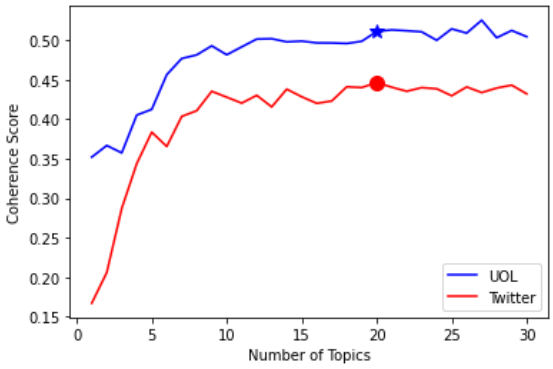
Topic models are particularly useful because they enable the inference of structure from a large data collection without the need for extensive manual interventions [35]. In the sentiment analysis domain, one of the best-known techniques to discover topics is latent Dirichlet allocation (LDA) [36]. LDA is a statistical topic model with the purpose of automatically identifying groups of related terms that approximate to real-word topics. In our research, we used LDA to uncover the main discussion topics and their trends over time.

LDA requires the user to specify the number of topics, where this parameter provides control over the granularity of the discovered topics. A larger number of topics will produce more detailed topics (finer-grained), while a smaller number of topics will produce more general topics (coarser-grained). Therefore, there is no single value of the number of topics that is appropriate in all domains and types of problems. To discover the most appropriate number of topics, we performed several different LDA experiments, varying the number of topics from 1 to 30 for both data sources. As illustrated in Figure 4, the coherence score increases steadily and quickly at the beginning, but it becomes stable at the score of 10 for both data sources. With the goal to capture broad topic trends in both data collections while keeping them distinct from each other, we set

XSL•FO
RenderX

the number of topics to 20. Our final model generated 20 topics using the MALLET (Machine Learning for Language Toolkit) implementation of LDA [37] with the default parameters, and the coherence scores were 0.511 and 0.446 for news media and Twitter, respectively. The lowest values of the score obtained for Twitter are in line with previous results reported by [31,38,39], because LDA may not necessarily perform well when handling short texts. Despite this shortcoming, we still extracted a set of representative topics about COVID-19 from Twitter.

Figure 4. Coherence scores for the latent Dirichlet allocation. UOL: Universo Online.



After topic discovery, we manually categorized the topics in themes based on the first 10 words, as these terms are ranked by their probability of appearance. The topics were categorized in the following themes: Confirmed Cases, Economic Influences, Entertainment, Medical Supplies, Medical Treatment and Research, Political, Prevention and Control, and Stories. Table 2 presents a short description of each theme. These themes are commonly used in the literature [16,22,40] to aggregate topics discussing similar subjects.

Table 2. Descriptions of the considered themes in this work.

Theme	Description
Confirmed Cases	Mentions of confirmed cases of COVID-19, such as updated numbers of cases and mortalities
Economic Influences	The influence of COVID-19 on the economy and society, such as the large number of unemployed people
Entertainment	Cultural events, sports, or food, such as the interruption of soccer championships
Medical Supplies	The medical supply situation in Brazil, such as the lack of respirators and use of masks
Medical Treatment and Research	Mentions of medical treatment and research combating COVID-19, such as the use of hydroxychloroquine
Political	Mentions of politicians and public officials and their responsibility
Prevention and Control	Different aspects of prevention and control procedures, such as social isolation and lockdowns in cities
Stories	Stories from people in Brazil who became ill or about the impact of COVID-19 on people's lives

Once we obtained the topics and themes, we assessed their similarity to understand if Twitter and traditional news media cover similar categories and types of topics related to COVID-19. To achieve this, we adopted the popular cosine similarity, which is the angle between the representation of two topics, as a measure to report the similarity among topics:

where t_a and t_b are the vector representations of topics a and b, respectively. The range of similarity is between 0 and 1, as all vector values are positive, with 1 implying total similarity.

Calculation of Topic-Based Sentiment Scores

For the sentiment analysis, we identified the polarity of the opinion or emotion expressed in the texts. One challenge we faced was the lack of robust language resources to support sentiment analysis for the Portuguese language [41]. This is due to the lack of advanced tools for Portuguese. Some studies [32,42,43] showed that methods that translate data set texts to English for sentiment analysis perform better than versions in Portuguese. In fact, this approach has become very common in research involving multilingual texts, as machine translation systems are presenting a good level of maturity [44]. Thus, we adopted this approach in this work and translated all original texts to English using the googletrans tool, a free Google Translate application programming interface [45], which is also evaluated in [44] and used in [46].

After the translation process. we used the VADER (Valence Aware Dictionary and Sentiment Reasoner) tool [47] to calculate the degree of positivity or negativity of the texts. VADER is a lexicon and rule-based sentiment analysis tool that focuses on sentiments expressed in social media. It can capture sentiment intensity based on grammatical and syntactical conventions. It performs well on social media platforms such as Twitter. Evaluations in [47] also show competitive performance on conventional text data sets. Sentiments are assigned a value on a scale of –1.0 to 1.0, the most negative and most positive polarities, where 0.0 represents neutral.

XSL•FO
RenderX

Table 3 presents some examples. It can be seen that the translation tool captures expressions and social typos well (eg, *vc* = *você* = you) while preserving entity names.

Table 3. Examples of text translation and sentiment analysis.

Portuguese text	Translated text	VADER[a] sentiment
pq é que pessoas doentes insistem em espirrar para cima dos outros? nc ouviram falar de distanciamento social quando estão doentes?	why are sick people insisting on splashing on others? haven't you heard of social distance when you're sick?	–0.8
Pode o que vc quiser, seu corpo, suas regras. Fim.	Do what you want, your body, your rules. End.	0.0
ministro Paulo Guedes diz que vamos recuperar e em "V"!!! Isso significa, rapidamente! Eu confio nele!	Minister Paulo Guedes says that we will recover and in "V" !!! That means, quickly! I trust him!	0.8

[a]VADER: Valence Aware Dictionary and Sentiment Reasoner

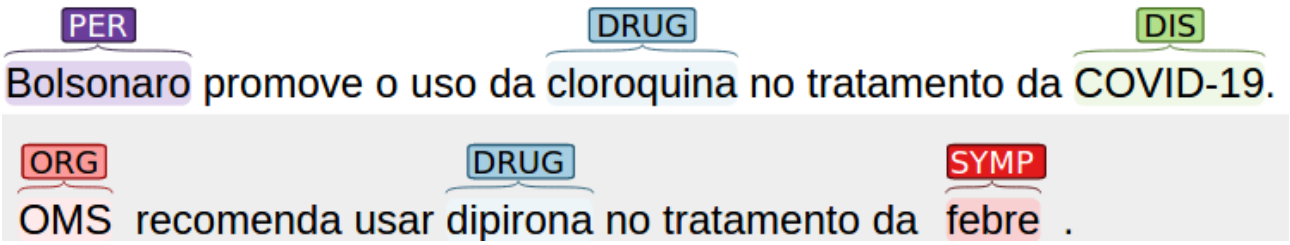
Named Entity Recognition and Network

Named entity recognition (NER) is particularly useful for identifying which terms in a text are mentions of entities in the real world and classifying them according to a set of categories. Although NER is not a new research field, it is not an easy task. The reasons for this are multifold. First, there is much work targeting English text, but studies focused on Portuguese text are still scarce [48]. Therefore, further work is still needed for several languages, such as Portuguese, due to their complex structures and relatively scarce language processing tools and data sets. Secondly, the category of a named entity is highly dependent on textual semantics and its surrounding context. The extraction of named entities from Twitter is an even more challenging task because tweets are short and are therefore more difficult to interpret compared to longer texts. In addition, short texts have many linguistic variations, and they tend to be less grammatically correct than news articles. Moreover, there are many definitions of named entity and evaluation criteria, which introduces evaluation complications [49]. Finally, we could not find a NER system that was designed to recognize the entities that we were interested in for this study.

Considering that the current state-of-the-art NER systems are based on neural architectures, we decided to use the spaCy2 library, which is based on the hierarchical attention network proposed in [50] and enables the creation of news models. The pretrained model for named entity recognition in Portuguese provided by spaCy recognizes the following entities: location, organization, person, and miscellaneous. As our goal is to identify entities in tweets and news articles related to COVID-19, we created our own model using the spaCy library. Our model is able to recognize the following entities: Person (PER), Organization (ORG), Disease (DIS), Symptoms (SYMP), and Drugs (DRUG). We chose these categories of entities because they are essential during a pandemic crisis.

We trained a new blank spaCy Portuguese language model; the initial model had no trained entities. An important issue in generating NER models is the effort involved in obtaining training data. To address this issue, we adopted a semisupervised approach to create training data that is better explained as follows. After training data generation, we then shuffled and looped over the training data. For each instance, the model was updated by calling the update function, which steps through all the words of each sentence. At each word, the update function makes a prediction. It then consults the golden standards to determine whether the prediction is right. If it is wrong, the update function adjusts its weights so that the correct action will score higher next time. Our model was built using 100 iterations with a dropout rate of 0.2. Once trained, our NER model was saved, and it can be used to recognize named entities in previously unseen tweets and news. Figure 5 illustrates an output example of our NER model, where *dipirona* (dipyrone) is a type of drug that was not used in training.

Figure 5. Sample outcome of the trained NER model.



Regarding the training data set, our strategy consisted of using a list of keywords for each entity regarding COVID-19 as a set of seeds. The algorithm in Textbox 1 describes this semi-supervised strategy. The algorithm takes two inputs. The first input is a set of pairs $\langle e, k\rangle \in P$, where e is an entity and k is a keyword. $\langle DRUG,tamiflu\rangle$ and $\langle DIS,COVID\text{-}19\rangle$ are examples of pairs. The second input is a set S of unlabeled sentences. We used the sentences from our data collection. Our algorithm returns as output a training data composed of pairs $\langle s,L\rangle$, where s is a sentence and L is a list of pairs $\langle e, k\rangle$. Each of these pairs represents that sentence s contains one or more keywords $k \in S$ associated with some entity $e \in E$.

XSL•FO
RenderX

The algorithm iterates through the set of sentences $s \in S$ (lines 4-12), attempting to match any of the sentence terms with some $k \in P$ (line 7). If there is an occurrence of k with any term of the sentence s, then the pair $\langle e, k \rangle$ is added to the list L (Line 8). After all the pairs belonging to P have been processed, a training pair $\langle s,L \rangle$ is added to T in line 11. If the keyword k does not match any term in sentence s, this sentence s is simply discarded. Note that in this case, the set L remains empty. After all the sentences $s \in S$ are processed, the algorithm outputs the training set T in line 13.

The training set generated by Algorithm 1 involves only a small degree of supervision, such as a set of keywords for each target entity, to start the learning process. To represent each type of text, we generated distinct training sets for news media and Twitter.

Textbox 1. Semisupervised learning strategy.

1 **let** E be the set of entities;

2 **let** K be a set of keywords about COVID-19;

Input: A set P = {⟨e, k | e ∈ E and k ∈ K}

Input: A set S of unlabeled sentences

Output: A set of training pairs $\{\langle s,L \rangle \mid s \in S$ and L is a list of $p \in P\}$

3 $T \leftarrow \varnothing$;

4 **foreach** $s \in S$ **do**

5 $L \leftarrow \varnothing$;

6 **foreach** $e, k \in P$ **do**

7 **if** $k \in \{s\}$ **then**

8 $L \leftarrow L \cup \{\langle e, k \rangle\}$;

9 **end**

10 **end**

11 $T \leftarrow T \cup \{\langle s,L \rangle\}$;

12 **end**

13 **return** T

Results

Overall Topic Distribution

Topics were analyzed for UOL and Twitter data sets according to the methods described in the previous section. Afterward, we organized the topics in themes as described in Table II. Topics and themes for UOL and Twitter are shown in Table 4 and Table 5, respectively. The original words in Portuguese are shown in brackets. The topic terms appear in decreasing order of density distribution. These tables show that the topic terms capture different meanings from both UOL and Twitter posts, and a diversity of themes was represented.

XSL•FO
RenderX

Table 4. Topics and themes for Universo Online.

ID	Topic	Theme
1	people (*pessoas*), mask (*máscara*), city (*cidade*), food (*alimentos*), products (*produtos*), employees (*funcionários*), local (*local*), alcohol (*álcool*), residents (*moradores*), image (*imagem*)	Prevention and Control
2	people (*gente*), do (*fazer*), stay (*ficar*), people (*pessoas*), time (*tempo*), account (*conta*), moment (*momento*), situation (*situação*), work (*trabalho*), folks (*pessoal*)	Stories
3	president (*presidente*), bolsonaro (*bolsonaro*), minister (*ministro*), state (*disse*), stated (*afirmou*), health (*saúde*), isolation (*isolamento*), interview (*entrevista*), party (*partido*), social (*social*)	Political
4	cases (*casos*), number (*número*), deaths (*mortes*), confirmed (*confirmados*), data (*dados*), total (*total*), bigger (*maior*), disease (*doença*), people (*pessoas*), deaths (*óbitos*)	Confirmed Cases
5	tests (*testes*), patients (*pacientes*), study (*estudo*), treatment (*tratamento*), research (*pesquisa*), vaccine (*vacina*), virus (*vírus*), researchers (*pesquisadores*), pain (*dor*), disease (*doença*), results (*resultados)*	Medical Treatment and Research
6	people (*pessoas*), virus (*vírus*), risk (*risco*), health (*saúde*), disease (*doença*), can (*podem*), diseases (*doenças*), avoid (*evitar*), population (*população*), seniors (*idosos*)	Prevention and Control
7	president (*presidente*), government (*governo*), states (*estados*), state (*estado*), minister (*ministro*), crisis (*crise*), pandemic (*pandemia*), congress (*congresso*), project (*projeto*), senate (*senado*)	Political
8	measures (*medidas*), isolation (*isolamento*), social (*social*), people (*pessoas*), state (*estado*), city (*cidade*), activities (*atividades*), capital (*capital*), cities (*cidades*), measure (*medida*)	Prevention and Control
9	can (*podem*), data (*dados*), information (*informação*), classes (*aulas*), access (*acesso*), possible (*possível*), do (*fazer*), work (*trabalho*), pandemic (*pandemia*), form (*forma*)	Stories
10	disease (*doença*), symptoms (*sintomas*), hospital positive (*positivo*), result (*resultado*), state (*disse*), death (*morte*), exams (*exames*), doctor (*médico*), covid (*covid*)	Medical Treatment and Research
11	state (*disse*), announced (*anunciou*), week (*semana*), pandemic (*pandemia*), march (*março*), communication (*comunicado*), events (*eventos*), april (*abril*), since (*partir*), june (*junho*)	Political
12	soccer (*futebol*), pandemic (*pandemia*), championship (*campeonato*), clubs (*clubes*), players (*jogadores*), season (*temporada*), athletes (*atletas*), games (*jogos*), return (*retorno*), english (*inglês*)	Entertainment
13	countries (*países*), state (*disse*), authorities (*autoridades*), world (*mundo*), people (*pessoas*), measures (*medidas*), worldwide (*mundial*), organization (*organização*), pandemic (*pandemia*), confinement (*confinamento*)	Political
14	economy (*economia*), companies (*empresas*), crisis (*crise*), fall (*queda*), market (*mercado*), sector (*setor*), pandemic (*pandemia*), bigger (*maior*), production (*produção*), impact (*impacto*)	Economic Influences
15	pandemic (*pandemia*), world (*mundo*), people (*pessoas*), moment (*momento*), big (*grande*), crisis (*crise*), population (*população*), form (*forma*), social (*social*), society (*sociedade*)	Stories
16	health (*saúde*), professionals (*profissionais*), patients (*pacientes*), hospitals (*hospitais*), beds (*leitos*), state (*estado*), doctors (*médicos*), attendance (*atendimento*), hospital (hospital), equipment (*equipamentos*)	Medical Supplies
17	decision (*decisão*), police (*polícia*), general (*geral*), public (*público*), ministry (*ministro*), request (*pedido*), state (*estado*), safety (*segurança*), public (*público*), measures (*medidas*)	Political
18	workers (*trabalhadores*), work (*trabalho*), government (*governo*), companies (*empresas*), payment (*pagamento*), income (*renda*), value (*valor*), caixa measure (*medida*), money (*dinheiro*)	Economic Influences
19	social nets (*redes*), video (*vídeo*), publication (*publicação*), instagram (*instagram*), twitter (*twitter*), wrote (*escreveu*), shared (*compartilhada*), quarantine (*quarentena*), world (*mundo*)	Stories
20	masks (*máscaras*), protection (*proteção*), coronavirus (*coronavirus*), passengers (*passageiros*), american (*americano*), local cases (*casos*), week (*semana*), final (*final*), transmission (*transmissão*)	Prevention and control

XSL•FO
RenderX

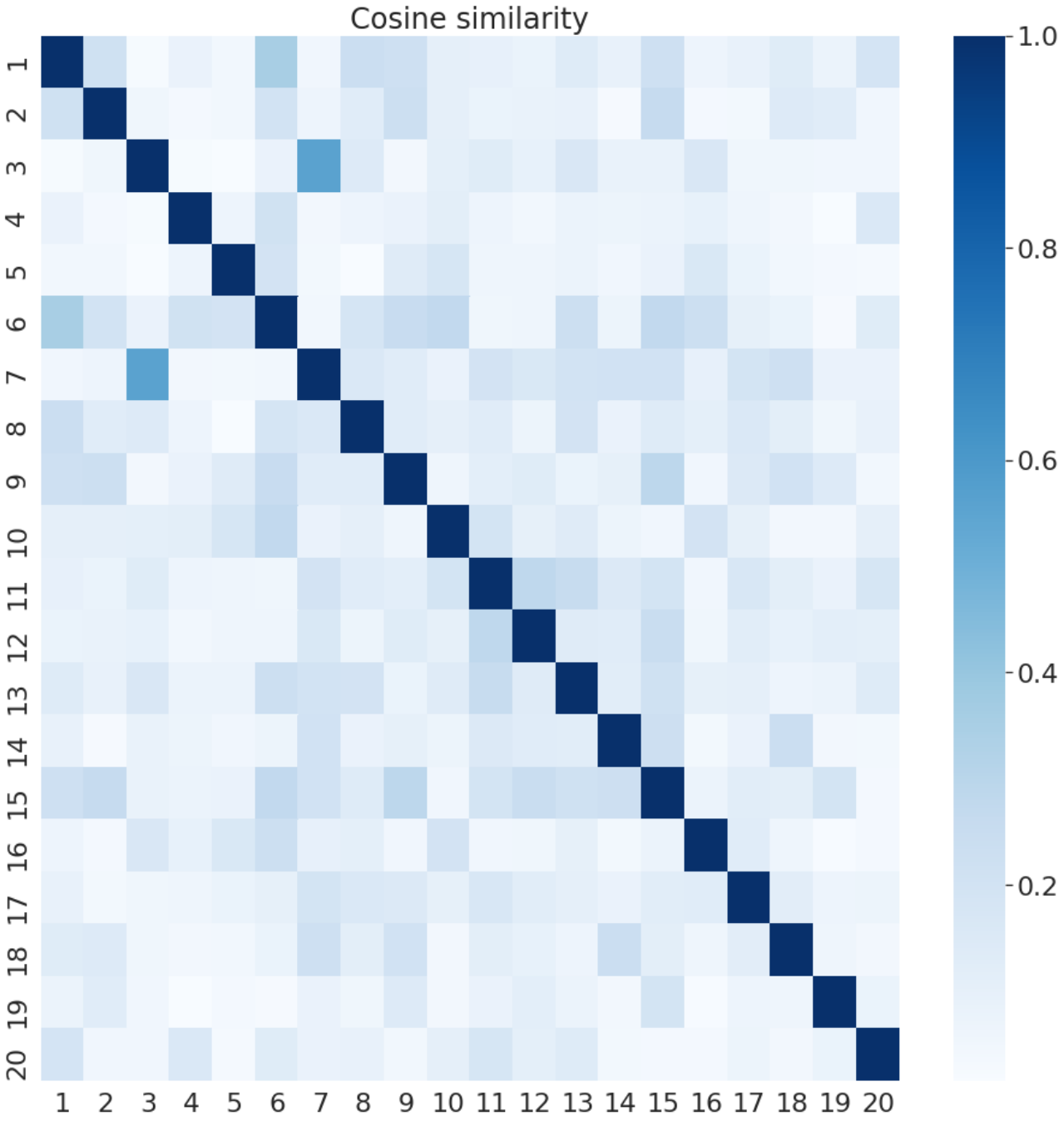
Table 5. Topics and themes for Twitter.

ID	Topic	Theme
1	president (*presidente*), bolsonaro (*bolsonaro*), minister (*ministro*), governors (*governadores*), sir (*senhor*), mayors (*prefeitos*), jairbolsonaro (*jairbolsonaro*), blame (*culpa*), mandetta (*mandetta*), meeting (*reunião*)	Political
2	instagram (*instagram*), igshid (*igshid*), covid (*covid*), twitter (*twitter*), mask (*máscara*), masks (*mascaras*), stay (*fique*), important (*importante*), prevention (*prevenção*), attention (*atenção*)	Medical Supplies
3	deaths (*mortes*), number (*número*), dead (*morte*), bigger (*maior*), covid (*covid*), numbers (*números*), countries (*países*), infected (*infectados*), weeks (*semanas*), months (*meses*)	Confirmed Cases
4	cases (*casos*), state (*estado*), confirmed (*confirmados*), tests (*testes*), city (*cidade*), twitter (*twitter*), deaths (*óbitos*), new (*novos*), coronavirus (*coronavirus*), total (*total*)	Confirmed Cases
5	health (*saúde*), hospitals (*hospitais*), combat (*combate*), professionals (*profissionais*), measures (*medidas*), public (*público*), actions (*ações*), beds (*leitos*), campaign (*campanha*), system (*sistema*)	Medical Supplies
6	twitter (*twitter*), pandemic (*pandemia*), covid (*covid*), lives (*vidas*), work (*trabalho*), moment (*momento*), video (*video*), congratulations (*parabéns*), big (*grande*), save (*salvar)*	Stories
7	people (*gente*), quarantine (*quarentena*), doing (*fazendo*), do (*fazer*), stay (*ficar*), friends (*amigos*), really (*sério*), damn (*porra*), finish (*acabar*), seeing (*vendo*)	Stories
8	pandemic (*pandemia*), time (*tempo*), quarantine (*quarentena*), things (*coisas*), moment (*momento*), time (*tempos*), difficult (*difícil*), do (*fazer*), pass (*passar*), expect (*espero*)	Stories
9	pandemic (*pandemia*), world (*mundo*), economy (*economia*), worldwide (*mundial*), general (*geral*), war (*guerra*), finish (*acabar*), ended (*acabou*), history (*história*), can (*podem*)	Economic Influences
10	true (*verdade*), policy (*política*), press (*imprensa*), left (*esquerda*), tell (*dizer*), state (*estado*), political (*político*), said (*falou*), media (*mídia*), shame (*vergonha*)	Political
11	people (*pessoas*), risk (*risco*), lack (*falta*), covid (*covid*), group (*grupo*), dying (*morrendo*), cause (*causa*), can (*podem*), died (*morreram*), diseases (*doenças*)	Confirmed Cases
12	social isolation (*isolamento*), detachment (*distanciamento*), measures (*medidas*), keep (*manter*), governor (*governador*), required (*necessário*), need (*necessidade*), services (*serviços*), commerce (*comércio*)	Prevention and Control
13	government (*governo*), population (*população*), combat (*combate*), money (*dinheiro*), pandemic (*pandemia*), fight (*combater*), want (*querem*), assist (*ajudar*), federal (*federal*), help (*ajuda*)	Political
14	death (*morte*), covid (*covid*), person (*pessoa*), disease (*doença*), positive (*positivo*), symptoms (*sintomas*), hospital died (*morreu*), week (*semana*), result (*resultado*)	Confirmed Cases
15	do (*fazer*), take (*tomar*), need (*precisa*), stay (*ficar*), die (*morrer*), work (*trabalhar*), want (*quero*), take (*pegar*), back (*voltar*), know (*saber*)	Economic Influences
16	corona (*corona*), virus (*vírus*), thing (*coisa*), speak (*falar*), state (*disse*), speaking (*falando*), cause (*causa*), buy (*comprar*), account (*voltar*), looks (*parece*)	Stories
17	coronavirus (*coronavirus*), twitter (*twitter*), covid (*covid*), vaccine (*vacina*), news (*notícias*), health (*saúde*), research (*pesquisa*), globo (*globo*), coronavirus (*coronavirus*)	Medical Treatment and Research
18	situation (*situação*), form (*forma*), problem (*problema*), exist (*existe*), best (*melhor*), done (*feito*), question (*questão*), possible (*possível*), looks (*parece*), example (*exemplo*)	Stories
19	pandemic (*pandemia*), crisis (*crise*), twitter (*twitter*), coronavirus (*coronavirus*), account (*conta*), companies (*empresas*), soccer (*futebol*), company (*empresa*), big (*grandes*), activities (*atividades*)	Entertainment
20	chloroquine (*cloroquina*), treatment (*tratamento*), patients (*pacientes*), doctor (*médicos*), medicine (*remédio*), study (*estudo*), doctors (*médicos*), medicine (*medicamento*), effects (*efeitos*), studies (*estudos*)	Medical Treatment and Research

Figure 6 and Figure 7 present the cosine similarities among the achieved topics of the UOL and Twitter collections, respectively. The goal was to observe the subject coverage obtained with these topics. We can note that the UOL news presents only two topics, which are very correlated (eg, 3 and 7 in Table 4 have many terms in common), while many other topics present lower correlations. This indicates that UOL topics cover more diverse subjects. Twitter, in contrast, presents more topics with stronger correlations (eg, 6, 9, 17, and 19) but also presents more topics with very low similarities. This is due to the fact that UOL documents are both formally written and longer than Twitter documents (which have a character limit). Thus, its documents naturally relate different subjects, and common terms can be found among different posts. Additionally, on social media, it is common for a few subjects to be concentrating the attention of the users, while people also talk about aleatory things.

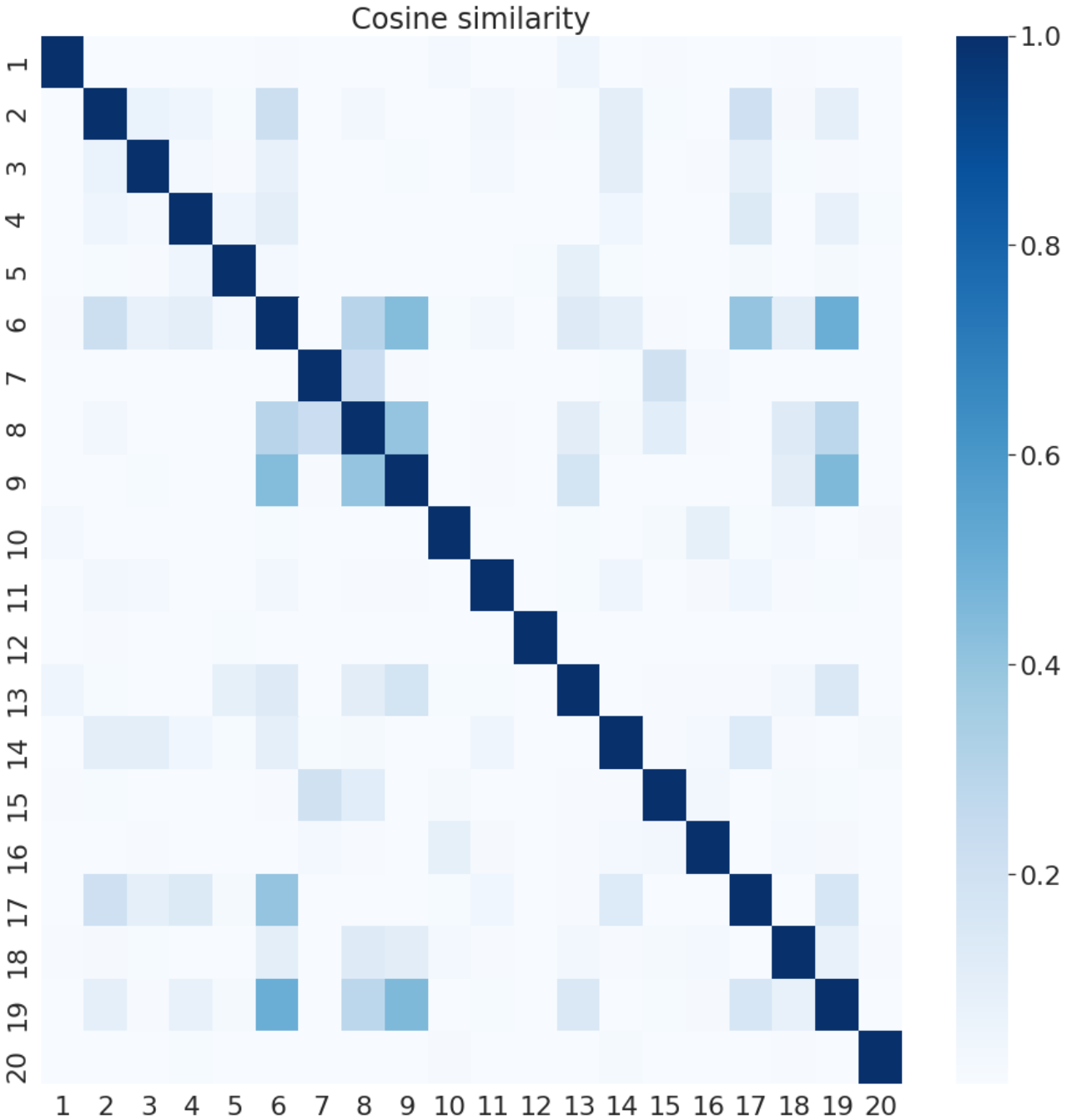
XSL•FO
RenderX

Figure 6. Similarity among Universo Online topics.



XSL•FO
RenderX

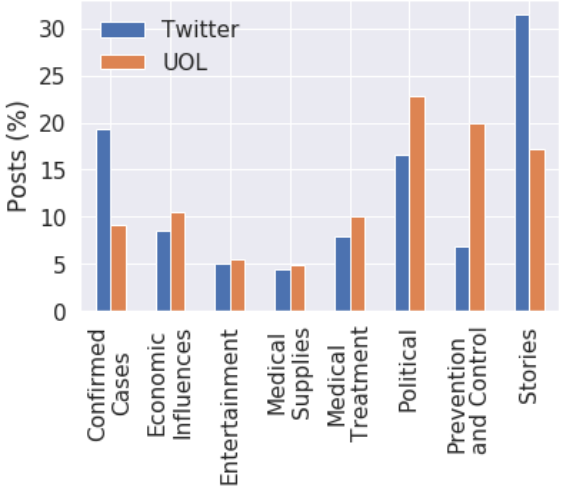
Figure 7. Similarity among Twitter topics.



The theme distributions between the UOL and Twitter collections are compared in Figure 8. In the UOL media, we can see that the preferred subjects are Politics, Prevention and Control, and Stories. One theme that was less common than expected is Confirmed Cases; however, this can be explained by the fact that this information was concentrated on fixed dashboards rather than in new documents. On Twitter, people posted frequently about the impact of COVID-19 on their lives (Stories), followed by concerns about Confirmed Cases and Political subjects, especially among supporters and critics of Brazil's president.

XSL•FO
RenderX

Figure 8. Distribution of themes. UOL: Universo Online.

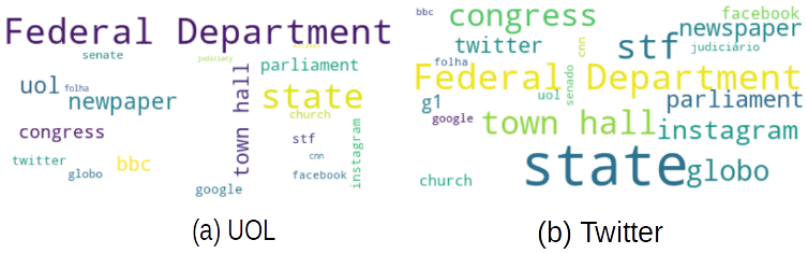


Entity Distribution and Network

According to the NER analysis method described in the last section, we compared the main mentions for each entity using word clouds, as this popular text analysis tool provides a visualization of word frequency in a source text while giving more prominence to words that occur more often. To facilitate the understanding of the most representative words by entity, we decided to show the 20 most frequently mentioned words in each entity. Words that were incorrectly extracted as belonging to an entity were manually removed. We assessed that our NER analysis method had an average accuracy of approximately 85% among the 20 most frequently mentioned terms. In Figure 9, we present the most frequent person entity mentions for the UOL and Twitter data collections. The entities of Bolsonaro (the president of Brazil) and Trump (the president of the United States) are the most frequently mentioned persons in both collections. Other frequently mentioned persons in both collections are governors and mayors (eg, Doria and Crivella) and other political personalities (eg, Maia and Moro). Figure 10 presents the most frequent organizations (ORG), and it can be observed that the State Department and Federal Department are the most commonly mentioned entities, followed by media companies. It is interesting to note that social media posts refer frequently to formal media (eg, Globo, which is the main television network in Brazil), and UOL news refers frequently to social networks (eg, Instagram and Twitter). Regarding the Disease entity (DIS), we can see in Figure 11 that the main terms are COVID and coronavirus, as expected, and the terms cancer and Dengue (a common tropical disease) are representative. In Figure 12, the Symptoms (SYMP) entity shows the most common COVID-19 symptom terms, namely *pain*, *fever*, and *cough*, as expected. Finally, Figure 13 shows that the Drugs entity (DRUG) is very polarized to the *chloroquine* discussion in both collections; however, UOL media seems to contain more information about vaccines.

Figure 9. Word clouds showing the most frequent entity mentions in the Persons category: (a) Universo Online; (b) Twitter.

Figure 10. Word clouds showing the most frequent entity mentions in the Organizations category: (a) Universo Online; (b) Twitter.



XSL•FO
RenderX

Figure 11. Word clouds showing the most frequent entity mentions in the Disease category: (a) Universo Online; (b) Twitter.

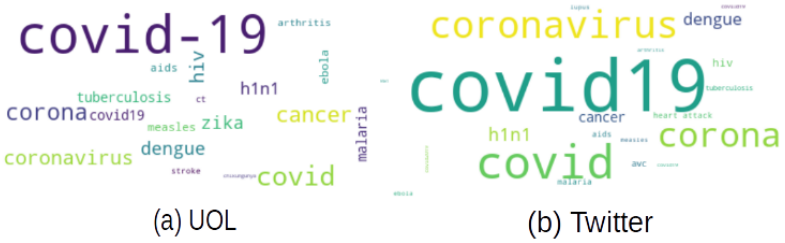


Figure 12. Word clouds showing the most frequent entity mentions in the Symptoms category: (a) Universo Online; (b) Twitter.

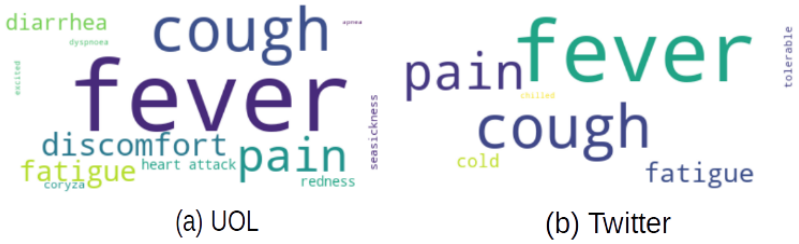
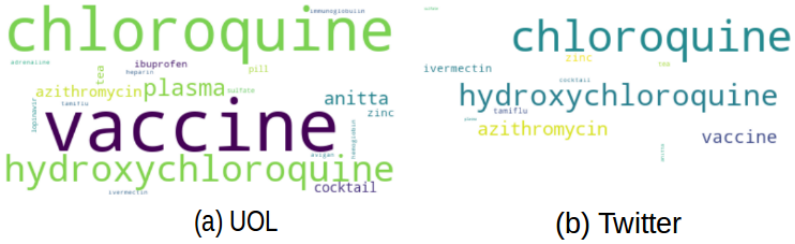
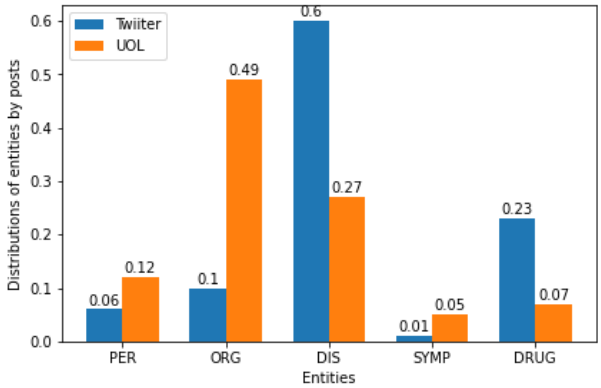


Figure 13. Word clouds showing the most frequent entity mentions in the Drugs category: (a) Universo Online; (b) Twitter.



From the word clouds for all these entities, it is important to mention that the found terms are very coherent with their respective entity categories. This fact reinforces that the adopted NER method is valid for the Portuguese language and that this study reflects the Brazilian perception of the COVID-19 pandemic. By comparing both formal and social media, it can be noted that there is no substantial difference regarding the main terms. However, people's discussions on Twitter have much sparser terms than those on UOL, while the terms in latter seem to be more diverse. Another important difference between the collections can be seen in the entity distribution graph in Figure 14. Both the UOL and Twitter texts obviously contain the main Disease terms (DIS) frequently, with a higher proportion in the size-limited Twitter posts. UOL news articles refer more to official sources of information (ORG entities), while people on Twitter talk more about drug treatments (DRUG). In fact, the administration of chloroquine was the cause of polemic and controversial debate in Brazil, with high politicization [51].

Figure 14. Entity distributions by data set. DIS: Disease; DRUG: Drugs; ORG: Organization; PER: Person; SYMP: Symptoms.



Topic and Sentiment Changes

Once topics were obtained for all posts in a collection, we classified every document by its topic with highest probability and applied the previously described sentiment analysis. We then grouped all posts by weekly intervals of time, summing the number of documents in each theme and calculating the sentiment averages.

Figure 15 and Figure 16 show how UOL and Twitter sentiments changed over time according to the defined themes. The y-axis

XSL•FO
RenderX

presents the sentiment mean for all documents in a given theme, the x-axis presents its evolution grouped by week of the year, and the dot size is proportional to the number of documents for a given theme and week. In both figures, we can see that the number of posts related to COVID-19 was very small during the initial weeks of the pandemic, and the posts presented high sentiment variation due to the impact of some specific posts. The number of posts began to grow considerably after mid-March, when Brazil registered its first death from COVID-19 (March 12, 2020). From the UOL sentiment analysis shown in Figure 15, we can see that all themes are more distributed around the neutral polarity (0.0). The themes of Entertainment and Stories have more positive averages (around 0.25), while Confirmed Cases (involving the number of cases and deaths) and Political are more negative themes (–0.25). Confirmed Cases reached a minimum representative polarity point (with more than 180 posts) by mid-April, just when the curves of confirmed cases and deaths started to scale exponentially. By the end of May, several themes presented a sentiment improvement; this coincides with the plateau of cases and deaths in several Brazilian capitals, such as Manaus, São Paulo, and Rio de Janeiro. For the Twitter collection, as shown in Figure 16, all the themes are positioned lower on the sentiment scale. Political, Confirmed Cases, Prevention and Control, and Economic Influences are more negative (near –0.2), while other themes are close to neutral polarity (0.0). For Twitter posts, it is possible to see that Economic Influences and Prevention and Control are positioned lower on the sentiment scale than UOL news. In fact, much discussion occurred regarding the need for quarantine or social distancing and the impact of these measures on the unemployment rate. This finding is reinforced by the observation that the Economic Influences sentiment increased on average in the beginning of April, when the government announced financial aid for autonomous workers [52]. Unlike UOL articles, Twitter had not yet shown any positivity by the end of May, and Economic Influences showed an additional decrease. We evaluated the standard deviations of the sentiment means, and we noted that they did not change greatly over time or among the themes; therefore, we omitted these data from the graphs. However, slight differences were observed between news (SD ~0.7) and social media (SD ~0.5). These standard deviations show that both sources present a high diversity of sentiments.

Figure 15. Universo Online sentiment analysis over time.

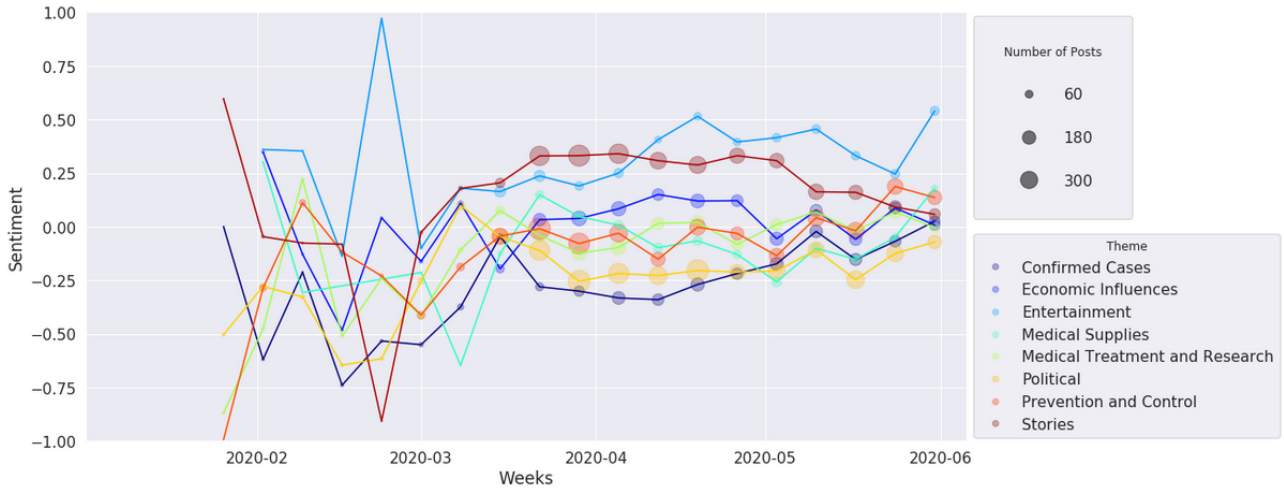
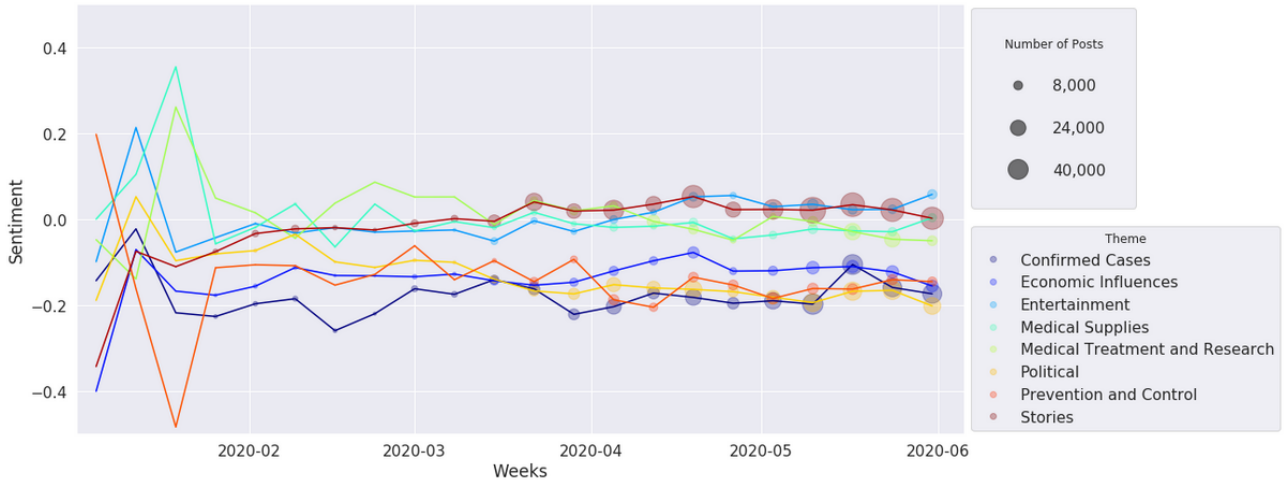


Figure 16. Twitter sentiment analysis over time.



XSL•FO
RenderX

Discussion

Principal Results

From a general point of view, we can observe that UOL articles and Twitter posts were concerned about the same main COVID-19 topics and themes. For instance, the topics and themes were very similar for both types of media, and this was reflected in the most common entity mentions. This study suggests that formal news media and social media influence each other; we found a representative cross-reference in the Organization entity graph.

The main differences found between UOL and Twitter pertain to the distribution of the main themes, diversity of entities, and overall sentiment about subjects related to COVID-19. Formal media naturally refers more to official entities and their recommendations. This can be seen in its top themes (Political and Prevention and Control), top entity groups (Organization and Disease), and diversity of entity mentions. Twitter, in contrast, is very focused on personal opinions and cases, as demonstrated by its top theme (Stories) and entity groups (Disease and Drugs). Additionally, social media tended to have a more negative polarity for all themes, while formal media seemed to present almost neutral polarity on average. Together with the very high number of collected tweets during the period, which shows that discussion about the disease was very active, we can observe the severity of the pandemic in Brazil and people's concerns about it.

It is remarkable how the subject of COVID-19 was the target of political polarization in Brazil. This theme was frequently discussed on both formal and social media, with higher negative sentiments over time. Drugs was the second most common entity in social media discussions, and it was very focused on the use of chloroquine to treat patients with COVID-19. A suggested hypothesis to explain this finding is that Brazil's government stated many times that this drug could help treat COVID-19 while minimizing the severity of the disease. In fact, in all the periods examined in this research, the government and formal media positioned themselves in opposite fields in this discussion, which is reflected in the high number of citations to political organization entities and in the disproportional reference to this specific drug.

Finally, by applying the proposed methodology, it was possible to observe the main information being conveyed and how people were reacting to it. This provides a way to monitor the evolution of a pandemic and its effects. Moreover, we believe this information can be useful for researchers and authorities to identify potentially controversial aspects, address possible misinformation, and establish better public policies for action and communication with the population.

Limitations

We discuss some limitations that can be attributed to this study as follows.

We retrieved data using a set of keywords; therefore, our data may have excluded tweets from users who wrote about the COVID-19 pandemic using different target keywords. A further limitation is that Twitter and UOL do not publish data about the profiles of their users, such as age, gender, or social class. Therefore, it was not possible to perform a stratified analysis of the users, and the results thus may not reflect the entire Brazilian population. A possible hypothesis is that different media reach different segments of society (eg, news media sites are accessed more frequently by more educated people); therefore, these differences may be reflected in the discovered topic distributions and sentiments. Thus, our findings may not be generalizable to other social media platforms or other communication media, such as television or radio. Moreover, the presented results for the selected vehicles may present some bias. For instance, a specific news media source may present a political leaning that can affect the sentiment about some themes. Therefore, while it is not our focus to explore possible bias and its impact on the results, caution is advised before assuming their generalization.

Conclusions

People rely on data published on the web to better understand recent global crises, and this is also occurring during the COVID-19 pandemic. News media web sites and social media are two distinguished channels of timely information. In this paper, we have proposed a methodological approach to analyze this type of media and to answer some questions regarding the COVID-19 pandemic in Brazil. The results presented and discussed in this study are particularly important because they make it possible to understand the difference between two data sources in how they cover global crises. In addition, this paper provides a method that uses several computational techniques to process textual social media in a language other than English. As the main contribution, this method resulted in observations that can aid understanding of the COVID-19 pandemic, with a better and more meaningful sentiment timeline.

In future work, we intend to extend this study to include data from longer periods of time, even after the pandemic ends. The idea is to understand how existing media platforms and people will react when they return to a normal situation and whether some trauma will remain. Additionally, we think that the proposed methodology is useful for studying other events of interest, such as other catastrophes and elections. Therefore, we intend to improve it by implementing a tool and applying it to new study cases.

Acknowledgments

This work was funded by Samsung Ocean Center, a research and development project at the State University of Amazonas.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Malta M, Rimoin A, Strathdee S. The coronavirus 2019-nCoV epidemic: is hindsight 20/20? EClinicalMedicine 2020 Mar;20:100289 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100289] [Medline: 32154505]
2. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280 [FREE Full text] [doi: 10.1111/tmi.13383] [Medline: 32052514]
3. Coronavirus disease 2019 (COVID-19): situation report 132. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/332280 [accessed 2021-02-05]
4. Yousuf H, Corbin J, Sweep G, Hofstra M, Scherder E, van Gorp E, et al. Association of a Public Health Campaign About Coronavirus Disease 2019 Promoted by News Media and a Social Influencer With Self-reported Personal Hygiene and Physical Distancing in the Netherlands. JAMA Netw Open 2020 Jul 01;3(7):e2014323 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.14323] [Medline: 32639569]
5. Boberg S, Quandt T, Schatto-Eckrodt T, Frischlich L. Pandemic populism: Facebook pages of alternative news media and the corona crisis—a computational content analysis. ArXiv Preprint posted online on April 6, 2020 [FREE Full text]
6. Henry M, Hauck G. Coronavirus is scary, but the flu is deadlier, more widespread. USA Today. 2020 Feb 01. URL: https://www.usatoday.com/story/news/health/2020/02/01/coronavirus-flu-deadlier-more-widespread-than-wuhan-china-virus/4632508002 [accessed 2020-08-03]
7. Quand T, Boberg S, Schatto-Eckrodt T, Frischlich L. Pandemic news: Facebook pages of mainstream news media and the coronavirus crisis—a computational content analysis. ArXiv. Preprint posted online on May 27, 2020 [FREE Full text]
8. Roser M, Ritchie H, Ortiz-Ospina E. Internet: the rise of social media. Our World in Data. URL: https://ourworldindata.org/internet [accessed 2020-08-03]
9. Hitlin P, Olmstead K. The science people see on social media. Pew Research Center. 2018 Mar 21. URL: https://www.pewresearch.org/science/2018/03/21/the-science-people-see-on-social-media [accessed 2020-06-01]
10. Shahi G, Dirkson A, Majchrzak T. An exploratory study of COVID-19 misinformation on Twitter. ArXiv. Preprint posted online on May 12, 2020 [FREE Full text]
11. Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. COVID-19 on social media: analyzing misinformation in Twitter conversations. ArXiv Preprint posted online on March 26, 2020 [FREE Full text]
12. Navarro JG. Statista. 2020 Aug 07. URL: https://www.statista.com/topics/2045/internet-usage-in-brazil/ [accessed 2020-10-26]
13. Washer P. Representations of SARS in the British newspapers. Soc Sci Med 2004 Dec;59(12):2561-2571 [FREE Full text] [doi: 10.1016/j.socscimed.2004.03.038] [Medline: 15474209]
14. Dias Neto D, Figueiras MJ, Campos S, Tavares P. Impact of economic crisis on the social representation of mental health: Analysis of a decade of newspaper coverage. Int J Soc Psychiatry 2017 Dec 23;63(8):736-743. [doi: 10.1177/0020764017737102] [Medline: 29058959]
15. Ribeiro B, Hartley S, Nerlich B, Jaspal R. Media coverage of the Zika crisis in Brazil: The construction of a 'war' frame that masked social and gender inequalities. Soc Sci Med 2018 Mar;200:137-144 [FREE Full text] [doi: 10.1016/j.socscimed.2018.01.023] [Medline: 29421460]
16. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
17. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
18. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
19. Comito C, Forestiero A, Pizzuti C. Twitter-based influenza surveillance: an analysis of the 2016-2017 and 2017-2018 seasons in Italy. In: IDEAS 2018: Proceedings of the 22nd International Database Engineering & Applications Symposium. 2018 Jun Presented at: 22nd International Database Engineering & Applications Symposium; June 18-20, 2018; Villa San Giovanni, Italy p. 175-183. [doi: 10.1145/3216122.3216128]
20. Ahmed W, Bath P, Sbaffi L, Demartini G. Zika outbreak of 2016: insights from Twitter. In: Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing. 2020 Presented at: International Conference on Human-Computer Interaction; July 19-24, 2020; Copenhagen, Denmark p. 447-458. [doi: 10.1007/978-3-030-49576-3_32]
21. 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. [doi: 10.2196/19447]

XSL•FO
RenderX

22. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
23. Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the characteristics of COVID-19 patients in China: analysis of social media posts. J Med Internet Res 2020 May 17;22(5):e19087 [FREE Full text] [doi: 10.2196/19087] [Medline: 32401210]
24. Du H, Nguyen L, Yang Z, Abu-Gellban H, Zhou X, Xing W, et al. Twitter vs news: concern analysis of the 2018 California wildfire event. 2019 Presented at: IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC); July 15-19, 2019; Milwaukee, WI p. 207-212. [doi: 10.1109/compsac.2019.10208]
25. Orellana-Rodriguez C, Keane MT. Attention to news and its dissemination on Twitter: a survey. Computer Science Review 2018 Aug;29:74-94. [doi: 10.1016/j.cosrev.2018.07.001]
26. Zhao W, Jiang J, Weng J, He J, Lim E, Yan H, et al. Comparing Twitter and traditional media using topic models. In: Advances in Information Retrieval. 2011 Presented at: European Conference on Information Retrieval; April 18-21, 2011; Dublin, Ireland p. 338-349. [doi: 10.1007/978-3-642-20161-5_34]
27. Valenzuela S, Puente S, Flores PM. Comparing disaster news on Twitter and television: an intermedia agenda setting perspective. J Broadcast Electron Media 2017 Nov 20;61(4):615-637. [doi: 10.1080/08838151.2017.1344673]
28. Aoki T, Suzuki T, Yagahara A, Hasegawa S, Tsuji S, Ogasawara K. Analysis of the regionality of the number of tweets related to the 2011 Fukushima Nuclear Power Station disaster: content analysis. JMIR Public Health Surveill 2018 Dec 18;4(4):e70 [FREE Full text] [doi: 10.2196/publichealth.7496] [Medline: 30563815]
29. Kim EH, Jeong YK, Kim Y, Kang KY, Song M. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. J Inf Sci 2016 Jul 11;42(6):763-781. [doi: 10.1177/0165551515608733]
30. Mondragon NI, Gil de Montes L, Valencia J. Ebola in the public sphere. Sci Commun 2017 Feb 01;39(1):101-124. [doi: 10.1177/1075547016688908]
31. Jónsson E, Stolee J. An evaluation of topic modelling techniques for Twitter. University of Toronto. 2015. URL: https://www.cs.toronto.edu/~jstolee/projects/topic.pdf [accessed 2021-02-05]
32. Araújo M, Pereira A, Benevenuto F. A comparative study of machine translation for multilingual sentence-level sentiment analysis. Inf Sci 2020 Feb;512:1078-1102. [doi: 10.1016/j.ins.2019.10.031]
33. TwitterScraper. GitHub. 2017. URL: https://github.com/taspinar/twitterscraper [accessed 2021-02-05]
34. García-Perdomo V, Salaverría R, Kilgo DK, Harlow S. To share or not to share. Journal Stud 2017 Jan 10;19(8):1180-1201. [doi: 10.1080/1461670x.2016.1265896]
35. Smola A, Narayanamurthy S. An architecture for parallel topic models. Proceedings VLDB Endowment 2010 Sep;3(1-2):703-710. [doi: 10.14778/1920841.1920931]
36. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022.
37. Mallet: MAchine Learning for LanguagE Toolkit. University of Massachusetts Amherst. 2002. URL: http://mallet.cs.umass.edu/ [accessed 2020-02-03]
38. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: WWW '13: Proceedings of the 22nd International Conference on World Wide Web. 3013 May Presented at: 22nd International Conference on World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil p. 1445-1456. [doi: 10.1145/2488388.2488514]
39. Sridhar V. Unsupervised topic modeling for short texts using distributed representations of words. In: Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. 2015 Jun Presented at: 1st Workshop on Vector Space Modeling for Natural Language Processing; June 2015; Denver, Colorado p. 192-200. [doi: 10.3115/v1/w15-1526]
40. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
41. Pereira DA. A survey of sentiment analysis in the Portuguese language. Artif Intell Rev 2020 Jul 6. [doi: 10.1007/s10462-020-09870-1]
42. Cirqueira D, Jacob A, Lobato F, de SA, Pinheiro M. Performance evaluation of sentiment analysis methods for Brazilian Portuguese. In: Business Information Systems Workshops. 2017 Jan 24 Presented at: BIS 2016: International Conference on Business Information Systems; July 6-8, 2016; Leipzig, Germany p. 245-251. [doi: 10.1007/978-3-319-52464-1_22]
43. Farias D, Matsuno I, Marcacini R, Rezende S. Opinion-meter: a framework for aspect-based sentiment analysis. In: Webmedia '16: Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web. 2016 Nov Presented at: 22nd Brazilian Symposium on Multimedia and the Web; November 8-11, 2016; Teresina, Brazil p. 351-354. [doi: 10.1145/2976796.2988214]
44. Balahur A, Turchi M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Comput Speech Lang 2014 Jan;28(1):56-75. [doi: 10.1016/j.csl.2013.03.004]
45. googletrans 3.0.0. Python Package Index. URL: https://pypi.org/project/googletrans [accessed 2021-02-05]
46. Laestadius L, Wang Y, Ben Taleb Z, Kalan ME, Cho Y, Manganello J. Online national health agency mask guidance for the public in light of COVID-19: content analysis. JMIR Public Health Surveill 2020 May 26;6(2):e19501 [FREE Full text] [doi: 10.2196/19501] [Medline: 32427103]

XSL•FO
RenderX

47. Hutto E, Gilbert C. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: Eighth International Conference on Weblogs and Social Media (ICWSM-14); June 1-4, 2014; Ann Arbor, MI URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399
48. Lopes F, Teixeira C, Gonçalo Oliveira H. Comparing different methods for named entity recognition in Portuguese neurology text. J Med Syst 2020 Feb 28;44(4):77. [doi: 10.1007/s10916-020-1542-8] [Medline: 32112285]
49. Marrero M, Urbano J, Sánchez-Cuadrado S, Morato J, Gómez-Berbís JM. Named Entity Recognition: Fallacies, challenges and opportunities. Comput Stand Inter 2013 Sep;35(5):482-489. [doi: 10.1016/j.csi.2012.09.004]
50. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 1997 Mar 19 Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2016; San Diego, CA p. 927-934. [doi: 10.18653/v1/N16-1174]
51. Wessel L. 'It's a nightmare.' How Brazilian scientists became ensnared in chloroquine politics. Science. 2020 Jun 22. URL: https://www.sciencemag.org/news/2020/06/it-s-nightmare-how-brazilian-scientists-became-ensnared-chloroquine-politics [accessed 2021-02-05]
52. Brazil: Government and institution measures in response to COVID-19. KPMG. URL: https://home.kpmg/xx/en/home/insights/2020/04/brazil-government-and-institution-measures-in-response-to-covid.html [accessed 2021-02-05]

Abbreviations

LDA: latent Dirichlet allocation
MALLET: Machine Learning for Language Toolkit
NER: named entity recognition
SARS: severe acute respiratory syndrome
UOL: Universo Online
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by CR Frances, B Sousa-Pinto; comments to author 13.10.20; revised version received 02.11.20; accepted 15.01.21; published 10.02.21.

Please cite as:
de Melo T, Figueiredo CMS
Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach
JMIR Public Health Surveill 2021;7(2):e24585
URL: http://publichealth.jmir.org/2021/2/e24585/
doi:10.2196/24585
PMID:33480853

©Tiago de Melo, Carlos M S Figueiredo. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 10.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Promoting Mask Use on TikTok: Descriptive, Cross-sectional Study

Corey H Basch[1], MPH, EdD; Joseph Fera[2], PhD; Isabela Pierce[1], AS; Charles E Basch[3], PhD

[1]Department of Public Health, William Paterson University, Wayne, NJ, United States

[2]Department of Mathematics, Lehman College, The City University of New York, Bronx, NY, United States

[3]Department of Health and Behavior Studies, Teachers College, Columbia University, New York, NY, United States

Corresponding Author:
Corey H Basch, MPH, EdD
Department of Public Health
William Paterson University
300 Pompton Rd
University Hall
Wayne, NJ, 07470
United States
Phone: 1 9737202603
Email: baschc@wpunj.edu

Abstract

Background: Over the past decade, there has been an increasing secular trend in the number of studies on social media and health.

Objective: The purpose of this cross-sectional study was to examine the content and characteristics of TikTok videos that are related to an important aspect of community mitigation—the use of masks as a method for interrupting the transmission of SARS-CoV-2.

Methods: In total, 100 trending videos with the hashtag #WearAMask (ie, a campaign on TikTok), along with 32 videos that were posted by the World Health Organization (WHO) and involved masks in any way (ie, all related WHO videos at the time of this study), were included in our sample. We collected the metadata of each post, and created content categories based on fact sheets that were provided by the WHO and the US Centers for Disease Control and Prevention. We used these fact sheets to code the characteristics of mask use.

Results: Videos that were posted on TikTok and had the hashtag #WearAMask garnered almost 500 million views, and videos that were posted by the WHO garnered almost 57 million views. Although the ratio of the number of trending #WearAMask videos to the number of WHO videos was around 3:1, the #WearAMask videos received almost 10 times as many cumulative views as the WHO videos. In total, 68% (68/100) of the trending #WearAMask videos involved humor and garnered over 355 million cumulative views. However, only 9% (3/32) of the WHO videos involved humor. Furthermore, 27% (27/100) of the trending #WearAMask videos involved dance and garnered over 130 million cumulative views, whereas none of the WHO videos involved dance.

Conclusions: This study is one of the first to describe how TikTok is being used to mitigate the community spread of COVID-19 by promoting mask use. Due to the platform's incredible reach, TikTok has great potential in conveying important public health messages to various segments of the population.

(JMIR Public Health Surveill 2021;7(2):e26392) doi:10.2196/26392

KEYWORDS

TikTok; COVID-19; social media; infodemiology; infoveillance; mask use; prevention; promotion; communication; public health; cross-sectional; content analysis; transmission

XSL•FO
RenderX

Introduction

As of November 30, 2020, there have been 62,363,527 COVID-19 cases and 1,456,687 deaths from COVID-19, and 13,082,877 cases and 263,946 deaths have been reported in the United States [1]. As with all emerging infectious disease outbreaks, public health messages have become centrally important. In fact, researchers have created infectious disease transmission models that are based on media influence–related data [2]. The sheer volume of information that is being generated throughout the COVID-19 pandemic has been classified as an "infodemic" by the World Health Organization (WHO) [3]. This influx of information (ie, correct information and incorrect/misleading information) can cause confusion, negate mitigation efforts [4], and result in serious negative consequences [5,6]. Members of the general public rely on different outlets for obtaining information. Recent studies have indicated that those who mostly rely on national news outlets believe that media coverage is largely accurate [7]. However, national news varies in content and focus; often presents information that is based on the different contexts of pandemics [8,9]; and at times, fails to capitalize on the opportunity to promote infection prevention techniques and coping strategies [8]. During the time when newspaper content was filled with information on COVID-19, popular social media outlets also began to provide information about the disease [10-14].

The majority of Americans who use the internet have, in part, done so to find web-based health information [15]. In a survey of US adults, 18% of participants (ie, in a sample of younger people who were generally unconcerned with political news) relied on social media as their customary source of news [15]. However, only 37% of these participants reported that they monitored news about the COVID-19 pandemic very closely, and over half (57%) of all participants reported that they have seen information about COVID-19 that seemed to be "completely made up" [15]. Social media is incredibly popular in the United States, as roughly 70% of the public has reported that they use some type of social media platform during their typical daily routines [16]. Prior to the COVID-19 pandemic, researchers raised questions about the proliferation of misinformation during public health emergencies [17-20], and one study on social media concluded that falsities spread more rapidly than the truth [21]. Several social media platforms have taken measures (eg, filtering posts from untrustworthy websites) for addressing concerns about the potentially harmful effects of misleading or incorrect information [22,23].

Public health professionals, agencies, and organizations have made efforts to use social media platforms to share information, or have partnered with popular members of these platforms to disseminate messages. One such effort has been promoting mask use on TikTok [24]. TikTok is a popular social media platform with approximately 800 million users worldwide and over 30 million users in the United States [25]. TikTok allows users to post 15-second or 60-second videos that are generally oriented toward entertainment. Roughly 42% of TikTok users are aged 18-24 years, and about 27% of frequent users are aged 13-17 years. As of November 12, 2020, 410 people in the United States aged between 15 and 24 years have died due to COVID-19 [26]. Although the incidence rate of such deaths is higher among people aged ≥25 years, studies have confirmed that younger individuals can transmit SARS-CoV-2 [27,28].

Over the past decade, there has been an increasing secular trend in the number of studies on social media and health, as indexed by the National Library of Medicine. Yet on December 5, 2020, a web-based search for the keyword "TikTok" only yielded 13 results, of which 4 were studies that focused on COVID-19. The purpose of this study was to examine the content and characteristics of TikTok videos that relate to an important aspect of community mitigation—the use of masks as a method for interrupting the transmission of SARS-CoV-2.

Methods

The methods that we used in this study mirror those that were used in a prior study on TikTok and COVID-19 [10]. However, our study focused exclusively on the use of TikTok as an infection prevention tool for promoting mask use and mitigating the airborne transmission of SARS-CoV-2. On November 3, 2020, we identified 100 trending TikTok videos that used the hashtag #WearAMask, which is a campaign on TikTok that encourages mask use [29,30]. In addition, our sample included all WHO videos that mentioned masks (n=32). The WHO has approximately 2.7 million followers on TikTok; however, unlike the TikTok users in the #WearAMask campaign, the WHO's followers are not represented by a hashtag. The reason our study was structured in this way was so that we could compare TikTok posts that used the most popular mask-promoting hashtag at the time of this study (ie, consumer-driven posts), with those from the most popular nongovernmental organization at the time of this study [31] (ie, professionally driven posts). Thus, the videos that we analyzed in this study represented consumer and professional interests for the same topic.

We conducted various quantitative data collection and analysis methods. Content categories were created by using fact sheets about mask use, which were provided by the WHO and the US Centers for Disease Control and Prevention [32,33]. Furthermore, we collected the metadata of each post, which included the date of the post; the number of views, likes, and comments; whether the post was in English or Spanish; the gender of the individual(s) in the video; whether the video included consumers (ie, members of the general public) or professionals (ie, doctors, registered nurses, public health professionals, etc); and whether the video involved dance, music, or humor. It is important to note that videos with the hashtag #WearAMask reported the exact number of views, while WHO videos reported a rounded number of views. Although these view counts were nuanced, we chose to report them as they appeared on TikTok, so as to not lose data. We also noted whether metadata characteristics or a mask appeared within the first 15 seconds of a video, whether the specific hashtag #WearAMask was mentioned or displayed in a video's description, and whether an image of a mask or the word "mask" appeared in a video's thumbnail. With regard to coding, we determined whether a video mentioned the following information: the hashtag #WearAMask, the correct use of a mask, the ramifications of improper mask use, guidelines for

XSL•FO
RenderX

using fabric masks, the necessity of masks for people who do not practice social distancing, the importance of using a mask to prevent infection, three-layer masks, and comparisons among different types of masks. To establish interrater readability, one author (ie, IP) recorded the content of all included videos, and another author (ie, CHB) coded the content of a random sample of 20 videos. The agreement of the coded data was excellent (κ=0.98). Data analysis involved the calculation of descriptive statistics, which included frequencies and percentages. Since this study did not involve human subjects, it was not reviewed by the institutional review board at William Paterson University, as per their protocol. This study was also deemed exempt for review by the Teachers College, Columbia University Institutional Review Board.

Results

A total of 132 TikTok videos were analyzed in this study. Of these 132 videos, 100 (75.8%) used the hashtag #WearAMask and garnered 494,824,395 views, and 32 (24.2%) were posted by the WHO and garnered almost 57 million views. Although the ratio of the number of trending #WearAMask videos to the number of WHO videos was 3:1, the #WearAMask videos received almost 10 times as many cumulative views as the WHO videos.

A total of 68% (68/100) of the trending #WearAMask videos involved humor and garnered over 355 million cumulative views, but only 9% (3/32) of the WHO videos involved humor. Additionally, while 27% (27/100) of the trending #WearAMask videos involved dance and garnered over 130 million cumulative views, none of the WHO videos involved dance. With regard to the trending #WearAMask videos, the proportion of videos was generally consistent with the proportion of cumulative views. With regard to the WHO videos, there were several notable differences. For example, while proper mask use was shown in 8 of the 32 WHO videos (25%), these videos received more than 75% (43,148,900/56,874,200, 75.87%) of the cumulative views. Additionally, while 15 of the 32 WHO videos (47%) mentioned wearing a mask, these videos received over 82% (47,049,700/56,874,200, 82.73%) of the cumulative views. In contrast, while 22 of the 32 WHO videos (69%) used music, these videos received less than 20% (10,534,400/56,874,200, 18.52%) of the cumulative views. Furthermore, although almost half of the WHO videos (15/32, 47%) mentioned that masks were essential infection prevention tools and garnered over 44 million views, this was not mentioned in a great majority of trending #WearAMask videos (Table 1).

Although our observations show how the presence of certain characteristics was associated with the number of cumulative views, independent 1-tailed *t* tests (α=.05) did not indicate whether these characteristics (ie, the use of dance, the use of music, mentions of wearing a mask, mentions of proper mask use, and mentions of masks as an essential infection prevention tool) led to higher view count averages. The tests we conducted for all videos and video categories returned *P* values of >.05. The presence of humor however had a marginal statistical effect (*P*=.052); videos that involved humor had an average view count of 5,010,849, while those that did not involve humor had an average view count of 3,211,939.

Of the 132 videos, 2 (1.5%) were recorded in English, 1 (0.8%) was recorded in Spanish, and 1 (0.8%) was recorded in both English and Spanish. The video that was recorded in Spanish had 4,600,000 views, and the video that was recorded in both English and Spanish had 3,500,000 views. Overall, 40.9% (54/132) of the videos only featured females, 32.6% (43/132) only featured males, and 16.7% (22/132) featured both.

Videos that used humor collectively received over 70% (66,329,979/92,922,611; 71.38%) of the total likes and over 75% (683,822/897,252; 76.21%) of the total comments.

Furthermore, although almost half of the WHO videos (15/32, 47%) mentioned that masks were essential infection prevention tools and garnered over 44 million views, this was not mentioned in a great majority (16/100, 16%) of trending #WearAMask videos (Table 1.)

XSL•FO
RenderX

Table 1. Observed characteristics, content, and view count of 132 TikTok videos (ie, 100 videos with the #WearAMask hashtag and 32 videos created by the World Health Organization).

Variables	#WearAMask videos (n=100)		World Health Organization videos (n=32)	
	Number of videos, n	Number of views, n (%)	Number of videos, n	Number of views, n (%)
Total	100	494,824,395 (89.69)[a]	32	56,874,200 (10.31)[a]
Video characteristics				
Used dance	27	130,480,395 (26.37)	0	0 (0)
Used music	58	284,671,795 (57.53)	22	10,534,400 (18.52)
Used humor	68	355,304,595 (71.8)	3	465,700 (0.82)
Mentioned or showed a mask in the first 15 seconds	59	240,836,695 (48.67)	27	55,412,700 (97.43)
Mentioned #WearAMask in the video's description	100	494,824,395 (100)	8	938,100 (1.65)
Contained the word "mask," an image of a mask, or an image of a person wearing a mask in the thumbnail	42	161,267,595 (32.59)	21	46,288,200 (81.39)
Video content				
Used the term "wear a mask"	10	31,583,695 (6.38)	15	47,049,700 (82.73)
Proper mask use	11	58,226,495 (11.77)	8	43,148,900 (75.87)
Improper mask use	4	9,974,400 (2.02)	1	84,000 (0.15)
Fabric mask use guidelines	0	0 (0)	6	1,265,100 (2.22)
Wearing a mask when not social distancing	0	0 (0)	7	1,698,900 (2.99)
Wearing a mask as an essential infection prevention tool	16	52,325,695 (10.57)	15	44,652,500 (78.51)
Mentioned three-layer fabric masks	0	0 (0)	5	1,098,800 (1.93)
Compared different types of masks	1	10,300,000 (2.09)	1	247,100 (0.43)

[a]This percentage refers to the number of views out of the total number of views for all 132 videos (N=551,998,595).

Discussion

The 100 trending TikTok videos on mask use that were reviewed in this study were collectively viewed almost 500 million times, which indicates the immense popularity of this emerging social media platform. Although the ratio of the number of #WearAMask videos to the number of WHO videos was 3:1, the former received almost 10 times as many views as the latter. A relatively small number of videos addressed proper versus improper mask use, the necessity of wearing a mask when not practicing social distancing, or the importance of masks as an infection prevention tool.

The differences between the #WearAMask and WHO videos in terms of video format (ie, the use of dance, music, and humor) and the commensurate differences in number of views suggest that these video formats warrant consideration as an effective method for disseminating up-to-date, mask use–related messages to the younger population.

To date, there is a paucity of research on public health information that is conveyed on TikTok. Although younger populations do not seem to be the most vulnerable to developing COVID-19 or experiencing the harmful consequences of the disease, they play an important role in disease transmission [27,28]. It is therefore imperative that public health professionals find methods for communicating effectively with younger populations when it comes to disseminating information about community mitigation (eg, promoting mask use).

Despite the large, relative differences between #WearAMask and WHO videos in terms of the number of views, the fact that the WHO videos on mask use garnered almost 57 million views is encouraging. The #WearAMask videos clearly used a different approach from that of the WHO videos; the #WearAMask videos were more likely to use dance, music and humor. They were also far more likely to attract viewers. Research is needed to help inform the WHO and other public health agencies about methods for adapting their communications, which are generally about very serious subjects; and to appeal to their intended audiences. For example, the presence of entertainers has been associated with the popularity of messages on other social media platforms [34], and the presence of influencers in messages related to health has been increasing [35,36]. This suggests that social media can be an effective platform for disseminating public health messages [37]. However, to date, TikTok videos on mask use do not, for the most part, highlight masks as an essential infection prevention tool or demonstrate proper mask use. Given the short length of TikTok videos, further research is needed to determine the types of messages that can effectively be conveyed on this platform.

XSL•FO
RenderX

This study had several limitations and delimitations. First, the content on TikTok fluctuates on a constant basis, and our study was cross-sectional in design. Therefore, we could not generalize our findings to different time periods. Second, the sample of videos that we analyzed was relatively small, given the large volume of posts on TikTok. Third, we could not distinguish the number views from the number of viewers, as we did not know whether the same users viewed a video multiple times. Fourth, we could not determine whether the videos were viewed in their entirety. Furthermore, this study was delimited to focus specifically on masks, which are an effective infection prevention tool [38,39] given the airborne nature of SARS-CoV-2 (ie, the virus that causes COVID-19) [40,41]. However, there are other hashtags that also promote mask use. We selected #WearAMask because it had the largest number of collective views (ie, about 4.5 billion views at the time the data was coded) [29]. There are also other organizations that have posted videos on TikTok; however, the WHO has been noted as the most popular [31].

Despite these limitations and delimitations, this study is one of the first to describe how TikTok is being used to mitigate the community spread of COVID-19 by promoting mask use. The nature of this medium presents challenges for conveying complex information. Nevertheless, because of its widespread reach, TikTok has great potential in conveying important public health messages to various segments of the population. Future research is needed to develop methods for applying the characteristics of highly viewed videos to public health messages. This is necessary for helping people make informed decisions about health promotion and disease prevention in general, and COVID-19–related mask use in particular.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-11-29]
2. Kim L, Fast SM, Markuzon N. Incorporating media data into a model of infectious disease transmission. PLoS One 2019 Feb 04;14(2):e0197646. [doi: 10.1371/journal.pone.0197646] [Medline: 30716139]
3. Novel coronavirus (2019-nCoV): Situation report 13. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf [accessed 2020-11-07]
4. Katella K. A COVID-19 'infodemic'? How to make sense of what you're reading. Yale Medicine. URL: https://www.yalemedicine.org/stories/covid-19-infodemic/ [accessed 2020-11-24]
5. Barua Z, Barua S, Aktar S, Kabir N, Li M. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. Progress in Disaster Science 2020 Dec;8:100119 [FREE Full text] [doi: 10.1016/j.pdisas.2020.100119]
6. Merchant RM, Asch DA. Protecting the value of medical science in the age of social media and "fake news". JAMA 2018 Dec 18;320(23):2415-2416. [doi: 10.1001/jama.2018.18416] [Medline: 30453315]
7. Jurkowitz M, Mitchell A, Oliphant JB, Shearer E. Americans who rely most on White House for COVID-19 news more likely to downplay the pandemic. Pew Research Center. 2020 May 20. URL: https://www.journalism.org/2020/05/20/americans-who-rely-most-on-white-house-for-covid-19-news-more-likely-to-downplay-the-pandemic/ [accessed 2020-12-02]
8. Basch CH, Hillyer GC, Erwin ZM, Mohlman J, Cosgrove A, Quinones N. News coverage of the COVID-19 pandemic: Missed opportunities to promote health sustaining behaviors. Infect Dis Health 2020 Aug;25(3):205-209 [FREE Full text] [doi: 10.1016/j.idh.2020.05.001] [Medline: 32426559]
9. Basch CH, Kecojevic A, Wagner VH. Coverage of the COVID-19 Pandemic in the Online Versions of Highly Circulated U.S. Daily Newspapers. J Community Health 2020 Dec;45(6):1089-1097 [FREE Full text] [doi: 10.1007/s10900-020-00913-w] [Medline: 32902813]
10. Basch CH, 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. Epub ahead of print 2020 Aug 10. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
11. Basch CH, Hillyer GC, Meleo-Erwin ZC, Jaime CJ, Mohlman J, Basch CE. Preventive behaviors conveyed on YouTube to mitigate transmission of COVID-19: Cross-sectional study. JMIR Public Health Surveill 2020 Apr 02;6(2):e18807 [FREE Full text] [doi: 10.2196/18807] [Medline: 32240096]
12. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
13. González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 pandemic. Int Braz J Urol 2020 Jul;46(suppl.1):120-124 [FREE Full text] [doi: 10.1590/S1677-5538.IBJU.2020.S121] [Medline: 32550706]
14. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of Chinese social media platform Weibo during early COVID-19 Outbreak: A retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]

XSL•FO
RenderX

15. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-11-16]
16. Social media fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-12-04]
17. Sell TK, Hosangadi D, Trotochaud M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 2020 May 07;20(1):550 [FREE Full text] [doi: 10.1186/s12889-020-08697-3] [Medline: 32375715]
18. Basch CH, Basch CE, Ruggles KV, Hammond R. Coverage of the Ebola virus disease epidemic on YouTube. Disaster Med Public Health Prep 2015 Oct;9(5):531-535. [doi: 10.1017/dmp.2015.77] [Medline: 26088275]
19. Larson HJ. The biggest pandemic risk? Viral misinformation. Nature Magazine. 2018 Oct 18. URL: https://media.nature.com/original/magazine-assets/d41586-018-07034-4/d41586-018-07034-4.pdf [accessed 2020-12-04]
20. Coren MJ. How to communicate in a postmodern pandemic. Quartz. 2020 May 07. URL: https://qz.com/1852552/past-pandemics-show-the-danger-of-misinformation-during-covid-19/ [accessed 2020-12-04]
21. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Dec 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
22. Pennycook G, Rand D. The right way to fight fake news. The New York Times. 2020 Mar 24. URL: https://www.nytimes.com/2020/03/24/opinion/fake-news-social-media.html [accessed 2020-12-04]
23. Fighting Fake News workshop report. Yale Law School. URL: https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf [accessed 2020-12-04]
24. Marks M. How state officials are trying to use TikTok to stop the spread of COVID-19. KUT Radio. 2020 Aug 20. URL: https://www.kut.org/post/how-state-officials-are-trying-use-tiktok-stop-spread-covid-19 [accessed 2020-12-04]
25. Sehl K. Everything brands need to know about TikTok in 2020. Hootsuite. 2020 Mar 02. URL: https://blog.hootsuite.com/what-is-tiktok/ [accessed 2020-12-04]
26. COVID-19 provisional counts - Weekly updates by select demographic and geographic characteristics. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm#AgeAndSex [accessed 2020-12-04]
27. Schwartz NG, Moorman AC, Makaretz A, Chang KT, Chu VT, Szablewski CM, et al. Adolescent with COVID-19 as the source of an outbreak at a 3-week family gathering — four states, June–July 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 09;69(40):1457-1459. [doi: 10.15585/mmwr.mm6940e2] [Medline: 33031365]
28. Park YJ, Choe YJ, Park O, Park SY, Kim YM, Kim J, COVID-19 National Emergency Response Center, Epidemiology and Case Management Team. Contact tracing during coronavirus disease outbreak, South Korea, 2020. Emerg Infect Dis 2020 Oct;26(10):2465-2468. [doi: 10.3201/eid2610.201315] [Medline: 32673193]
29. #wearamask hashtag videos on TikTok. TikTok. URL: https://www.tiktok.com/tag/wearamask [accessed 2020-12-04]
30. A guide to hashtags on TikTok. Boosted. URL: https://boosted.lightricks.com/a-guide-to-hashtags-on-tiktok/ [accessed 2020-12-04]
31. Lozan T. [Case Study] Governmental and non-governmental organizations are using TikTok for COVID-19 campaigns. SocialInsider. 2020 Apr 14. URL: https://www.socialinsider.io/blog/ngo-tiktok-campaigns/ [accessed 2020-12-04]
32. Infographic - Understanding the difference. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/niosh/npptl/pdfs/UnderstandDifferenceInfographic-508.pdf [accessed 2020-12-04]
33. Coronavirus disease (COVID-19) advice for the public: When and how to use masks. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks [accessed 2020-12-04]
34. Basch CE, Basch CH, Hillyer GC, Jaime C. The role of YouTube and the entertainment industry in saving lives by educating and mobilizing the public to adopt behaviors for community mitigation of COVID-19: Successive sampling design study. JMIR Public Health Surveill 2020 Apr 21;6(2):e19145. [doi: 10.2196/19145] [Medline: 32297593]
35. Vallone D, Smith A, Kenney T, Greenberg M, Hair E, Cantrell J, et al. Agents of social change: A model for targeting and engaging Generation Z across platforms - How a nonprofit rebuilt an advertising campaign to curb smoking by teens and young adults. J Advert Res 2016 Dec 01;56(4):414-425. [doi: 10.2501/jar-2016-046]
36. Hair E, Pitzer L, Bennett M, Halenar M, Rath J, Cantrell J, et al. Harnessing youth and young adult culture: Improving the reach and engagement of the truth® campaign. J Health Commun 2017 May 26;22(7):568-575 [FREE Full text] [doi: 10.1080/10810730.2017.1325420]
37. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: Using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
38. Scientific brief: Community use of cloth masks to control the spread of SARS-CoV-2. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html#:~:text=SARS%2DCoV%2D2%20infection%20is,of%20SARS%2DCoV%2D2 [accessed 2020-12-04]

XSL•FO
RenderX

39. Van Dyke ME, Rogers TM, Pevzner E, Satterwhite CL, Shah HB, Beckman WJ, et al. Trends in county-level COVID-19 incidence in counties with and without a mask mandate - Kansas, June 1-August 23, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 27;69(47):1777-1781 [FREE Full text] [doi: 10.15585/mmwr.mm6947e2] [Medline: 33237889]
40. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
41. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Mar;579(7798):265-269 [FREE Full text] [doi: 10.1038/s41586-020-2008-3] [Medline: 32015508]

Abbreviations

WHO: World Health Organization

Edited by T Sanchez; submitted 10.12.20; peer-reviewed by T Dowdle, P Kumar, N Noreen; comments to author 06.01.21; revised version received 08.01.21; accepted 27.01.21; published 12.02.21.

Please cite as:
Basch CH, Fera J, Pierce I, Basch CE
Promoting Mask Use on TikTok: Descriptive, Cross-sectional Study
JMIR Public Health Surveill 2021;7(2):e26392
URL: http://publichealth.jmir.org/2021/2/e26392/
doi:10.2196/26392
PMID:33523823

©Corey H Basch, Joseph Fera, Isabela Pierce, Charles E Basch. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 12.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Apple Inc Mobility Trend Data Related to the COVID-19 Outbreak in Japan: Statistical Analysis

Junko Kurita[1], MN; Yoshiyuki Sugishita[2], PhD, MPD, MD; Tamie Sugawara[2], PhD; Yasushi Ohkusa[2], PhD

[1]Department of Nursing, Tokiwa University, Mito, Ibraki, Japan

[2]National Institute of Infectious Diseases, Tokyo, Japan

Corresponding Author:
Junko Kurita, MN
Department of Nursing
Tokiwa University
1-430-1 Miwa
Mito, Ibraki, 3108585
Japan
Phone: 81 29 232 2511
Email: kuritaj@tokiwa.ac.jp

Abstract

Background: In Japan, as a countermeasure against the COVID-19 outbreak, both the national and local governments issued voluntary restrictions against going out from residences at the end of March 2020 in preference to the lockdowns instituted in European and North American countries. The effect of such measures can be studied with mobility data, such as data which is generated by counting the number of requests made to Apple Maps for directions in select countries/regions, sub-regions, and cities.

Objective: We investigate the associations of mobility data provided by Apple Inc and an estimate an an effective reproduction number R(t).

Methods: We regressed R(t) on a polynomial function of daily Apple data, estimated using the whole period, and analyzed subperiods delimited by March 10, 2020.

Results: In the estimation results, R(t) was 1.72 when voluntary restrictions against going out ceased and mobility reverted to a normal level. However, the critical level of reducing R(t) to <1 was obtained at 89.3% of normal mobility.

Conclusions: We demonstrated that Apple mobility data are useful for short-term prediction of R(t). The results indicate that the number of trips should decrease by 10% until herd immunity is achieved and that higher voluntary restrictions against going out might not be necessary for avoiding a re-emergence of the outbreak.

(JMIR Public Health Surveill 2021;7(2):e20335) doi:10.2196/20335

KEYWORDS

peak; COVID-19; effective reproduction number; mobility trend data; Apple; countermeasure

Introduction

The rapid spread of COVID-19, the disease caused by SARS-CoV-2, continues to have a substantial impact. At the end of October 2020, the World Health Organization reported approximately 46 million COVID-19 cases and more than one million fatalities. Japan, however, reported only approximately 100,000 cases with 1700 deaths, and at 769.2 per one million residents, the country's incidence rate was considerably lower than the global average of 5974 per million. Japan's case fertility rate, however, was comparable with the worldwide rate (1.7% vs 2.2%); that is, the country showed fewer patients but a moderate case fertility rate. Understanding the dynamics of Japan's COVID-19 outbreak has been important for planning and evaluating countermeasures.

As a countermeasure against the COVID-19 outbreak in Japan, both the national and local governments issued voluntary restrictions against going out at the end of March 2020 in preference to lockdowns such as those instituted in European and North American countries [1]. However, it is unclear exactly how voluntary restriction against going out requirements affected the outbreak compared with the country's post–voluntary restriction against going out statistics. The outbreak in Japan reached its first peak at the beginning of April

XSL•FO
RenderX

2020, but it was unclear why the peak occurred at that time. Thus, understanding the associations between the voluntary restriction against going out implementation and the first peak is urgent for controlling subsequent waves in Japan.

Susceptible-infected-recovered (SIR) models for COVID-19 incorporating countermeasures have emphasized the date that countermeasures were initiated [2,3]. However, at least in Japan, voluntary restrictions against going out expanded gradually; therefore, all-or-nothing approaches such as SIR models might not be appropriate. A more continuous variable is expected to be necessary to represent voluntary restriction against going out compliance over time.

Several companies, including Apple Inc and Alphabet Inc (hereinafter Apple and Google, respectively) worldwide, and Nippon Telegraph and Telephone (NTT) and East and West Japan Railway (JR) companies in Japan, have made available relevant data on such variables. Apple, the front-runner of this service, began providing data from January 13, 2020, specifically the daily ratio of the number of trips from homes to activities by transportation type (driving, transit, and walking) [4]. This data is generated by counting the number of requests made to Apple Maps for directions in select countries/regions, sub-regions, and cities. For privcy protection, data that is sent from users' devices to the Apple Maps service is associated with random, rotating identifiers; thus Apple claims to not have a profile of the movements and searches of individual users. Using trip information provided by Apple, we examined various associations and estimated the effective reproduction number R(t).

Methods

Applying a simple SIR model [2,3,5] to the epidemic curve in Japan, with its 120 million population, we assumed an incubation period that conformed to the empirical distribution in Japan; we used the number of patients who were symptomatic reported by the Ministry of Labour, Health and Welfare (MLHW) for February 10 to April 30, 2020, published on May 13, 2020 [6]. We excluded some patients from the data: persons presumed to have been infected abroad or as passengers on the Diamond Princess (the Ministry presumed that these patients did not represent a community-acquired infection in Japan). For some patients who were symptomatic, their onset dates were unknown; for these patients, we estimated the dates from an empirical distribution with duration extending from the onset to the report date among patients for whom the onset date had been reported.

In detail, we estimated the unknown onset dates as follows. Letting $f(k)$ represent this empirical distribution and N_t denote the number of patients for whom onset dates were not available published at date t, the number of patients for whom the onset date was known is $t - 1$. The number of patients for whom onset dates were not available was estimated as $f(1)N_t$. Similarly, the number of patients with onset date $t - 2$ for whom onset dates were not available was estimated as $f(2)N_t$. Hence, we estimated the total number of patients for whom the onset date was not available, considering an onset date of s, as $\Sigma_{k=1}f(k)N_s + k$ for the long duration extending from s.

Moreover, reporting delays were possible for the MLHW's published data; that is, if $s + k$ was larger than that in the current period t, then $s + k$ represented the future for period t, and for this reason, $Ns + k$ was not observable. This reporting delay engendered the underestimation of the number of patients, and thus, it had to be adjusted as $\Sigma_{k=1}^{t-s}f(k)N_s + k/\Sigma_{k=1}^{t-s}f(k)$. Similarly, we expected that patients for whom the onset dates were available would be affected by the reporting delay, leading to $M_s|_t/\Sigma_{k=1}^{t-s}f(k)$, where $M_s|t$ represents the reported number of patients for whom onset dates were within period s, extending until current period t.

We defined R(t) as the number of patients who were infected on day t divided by the number of patients who were presumed to be infectious. We calculated the number of patients who were infected from the epidemic curve by the onset date using a distribution of the incubation period. The distribution of infectiousness in symptomatic and asymptomatic cases was assumed to be 30% on the onset day, 20% on the following day, and 10% for the subsequent 5 days [7].

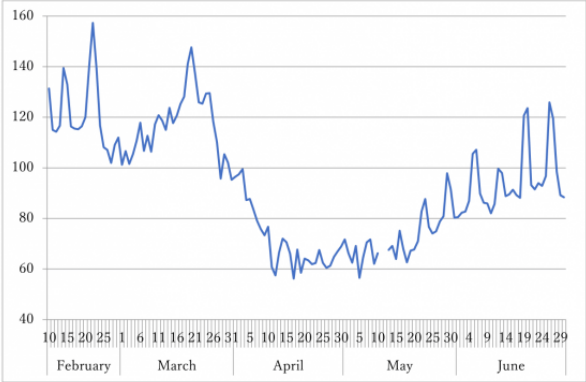
For analyses, we used Apple's average daily information on the three types of trips as well as their ratios using a root search function on Apple Maps on each day compared with that on January 13, 2020. Apple provides no information regarding how many users or proportions of Apple users used this function.

To elucidate the associations between R(t) and Apple data, we regressed R(t) on a polynomial function of daily data. We determined the order of the polynomial function stepwise from the model including only a linear term to the model, including a higher order if all estimated coefficients were significant. Besides the estimation using the whole period, we analyzed subperiods before and after March 10, 2020.

To intuitively understand the estimation result, we predicted R(t) when voluntary restriction against going out ceased and Apple data reached 100. We also sought the highest value of Apple data that implied R(t)<1 by a grid search from 100. To confirm the predictive power of the model, we predicted R(t) based on available Apple data on day t prospectively for the subsequent 2 months until the end of June 2020 (Figure 1); then, we compared the result with the observed R(t) on day $s>t + 40$ to avoid uncertainty because of reporting delays. We inferred significance at the 5% level.

XSL•FO
RenderX

Figure 1. Apple data in Japan from February 10 to the end of June 2020. The line represents the ratio of the number of trips from homes to activities by transportation type (driving, transit, and walking) on January 13, 2020. Apple data for May 12 and 13, 2020, were missing. The prediction in the period was also missing.

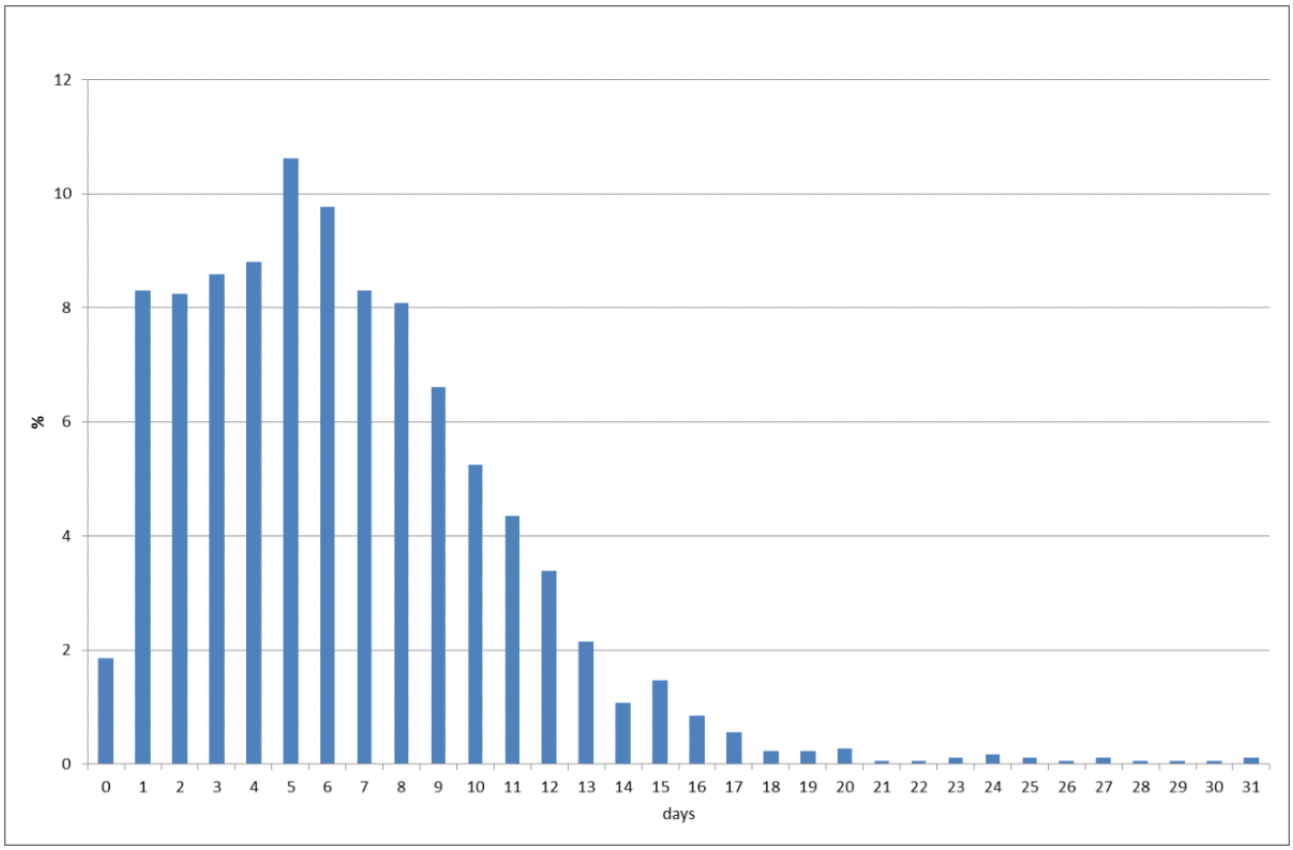


Results

From February 10 to April 30, 2020, in Japan, 13,967 community-acquired cases were identified, excluding asymptomatic cases. Figure 2 presents an empirical distribution of the duration of onset to a report in Japan, showing a maximum delay of 30 days. Figure 3 depicts the empirical distribution of incubation periods among 91 cases for which the MLHW had published an exposed date and an onset date. The mode was 6 days, and the average was 6.6 days.

Table 1 presents the estimation result. During the whole period and after the March 10, 2020, period, we selected up to cubic terms; however, before the March 10 period, we selected the quadratic term. The estimated coefficients in the whole period and after the March 10 period were similar, although they differed greatly from the subperiod before March 10.

Figure 2. Empirical distribution of duration from onset to report by Ministry of Labour, Health and Welfare in Japan. Bars represent the probability of duration from onset to report based on 657 patients for whom the onset date was available.



XSL•FO
RenderX

Figure 3. Empirical distribution of the incubation period published by Ministry of Labour, Health and Welfare (MLHW) in Japan. Bars show the distribution of incubation periods for 91 cases for which MLHW Japan had published the exposure date and the onset date. The patients for whom incubation was longer than 14 days are included in the bar shown for day 14.

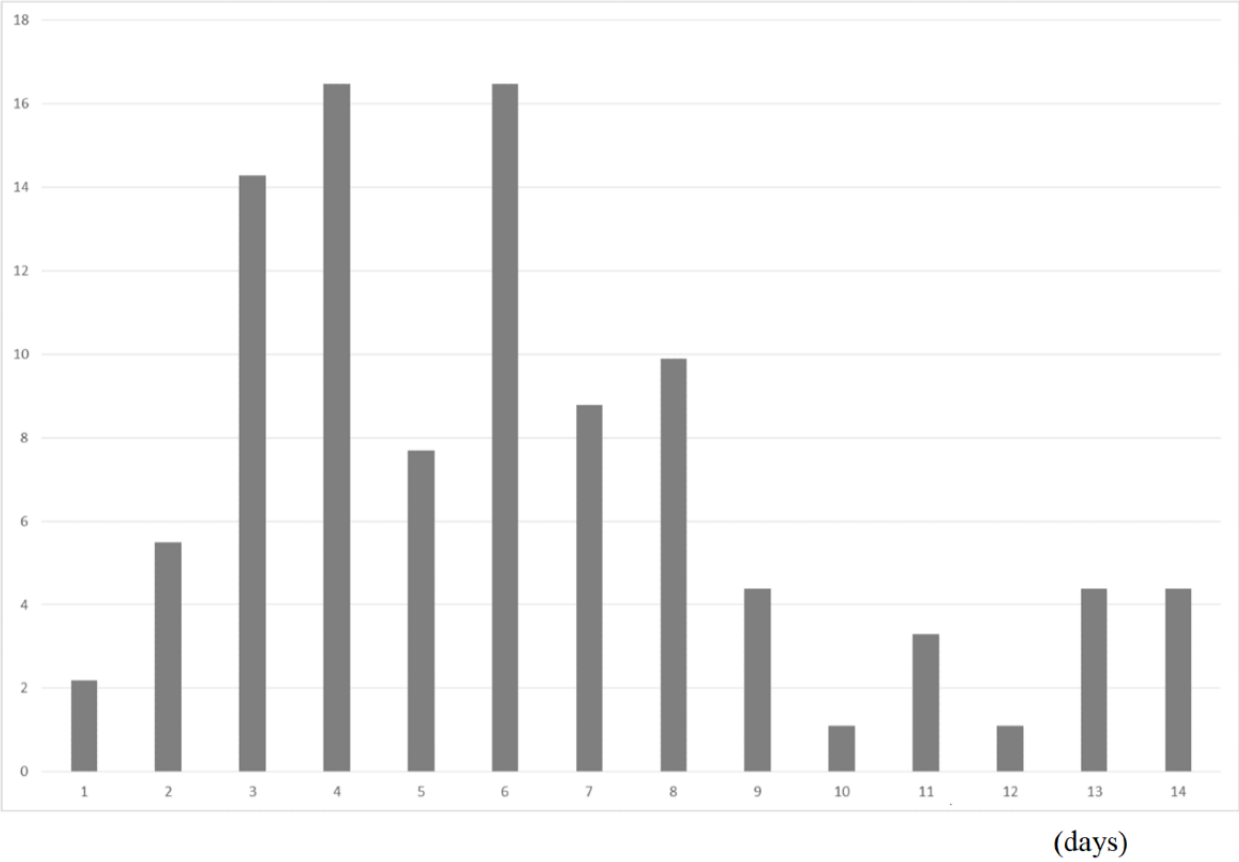


Table 1. Results of estimate R(*t*) for Apple data.[a]

Terms	Whole period (N=81)		Before March 10, 2020 (n=29)		After March 10, 2020 (n=52)	
	Estimate coefficient	*P* value	Estimate coefficient	*P* value	Estimate coefficient	*P* value
Linear term	−0.705	<.001	0.451	.004	−0.408	.008
Quadratic term	0.007	<.001	−0.002	.006	0.004	.008
Cubic term	−0.228*10−4	<.001	N/A[b]	N/A	−0.126*10−4	.02
Constant	21.55	<.001	−27.26	.005	13.51	.008
Adjusted R^2	0.701	N/A	0.349	N/A	0.891	N/A

[a]The dependent variable was R(*t*), and the explanatory variables were the polynomial function of Apple data.

[b]N/A: not applicable.

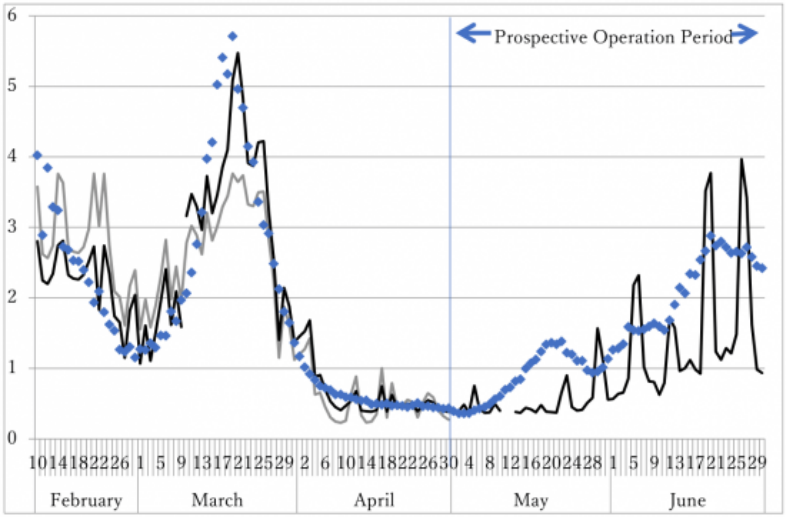
Figure 4 depicts the R(*t*) and the predicted line based on Apple data in the whole period and the two subperiods. The model clearly shows poor fit in the whole period, especially around the peak at the end of March 2020. Conversely, the prediction in the post–March 10 subperiod shows a clearly good fit around the peak.

The estimation result indicates that R(*t*) was 1.72 when voluntary restriction against going out ceased and mobility reverted to a normal level, which was 100 in the Apple data. Therefore, complete cessation of voluntary restrictions against going out would reinitiate the outbreak. We found in the Apple data that the critical level of R(*t*)<1 was 89.3.

Prospective operation using Apple data from the beginning of May 2020 for 2 months is shown on the right-hand side of Figure 3. The correlation coefficient for the actual observed R(*t*) and prospective prediction from Apple data for this period was 0.6433, which was significant.

XSL•FO
RenderX

Figure 4. R(*t*) and prediction from Apple data. The two black lines represent the prediction from Apple data in the subperiods before and after March 10, 2020; the gray line represents the prediction in the whole period. Dots represent the observed R(*t*). Apple data on May 12 and 13 were missing, and the prediction in the period was also missing. Until the end of April 2020, the prediction from Apple data was retrospective; from the beginning of May 2020, the prediction was prospective, although observed R(*t*) was calculated as of the date on t + 40.



Discussion

Principal Findings

Here, we demonstrated that Apple data are useful for short-term prediction of R(*t*). Figure 3 suggests that a model with two subperiods might be preferable to the whole model. Such a model might be consistent with the apparent change in the dominant virus among SARS-CoV-2 that occurred in March 2020 [8]. The estimations seem to indicate that the Wuhan-originated strain was dominant before the end of March but that a mutated virus subsequently became dominant in European countries and the United States. These two virus subtypes showed different associations with the mobility data and might have different infectiousness.

In the prospective operation period, the prediction using Apple data correlated significantly with the ultimately observed R(*t*), although the correlation among them was only slightly higher than the finding from the retrospective analysis until the end of April 2020; this finding is likely attributable to prospective operation. Furthermore, because the state of emergency declaration was canceled at the end of May 2020, the behavior of the population could be expected to have changed around then; for example, the intensity of meetings might have been lower or more mask wearing and longer distancing might be used for conversations. Under such circumstances, Apple data might lose their predictive power.

Current results obtained using Apple data indicate a critical level of R(*t*)<1 as 89.3. Consequently, a 10% reduction in trips from home, likely for more than 1 year, will be necessary until the so-called herd immunity is achieved; if this decrease is achieved, full voluntary restrictions against going out might not be necessary to avoid another outbreak. Current voluntary restriction against going out data indicate a 40% reduction from Apple data. Therefore, restrictions against excursions can be relaxed by 30 percentage points; such relaxation might include restarting schools or private events and sports that involve small numbers of players, little player contact, and large outside spaces, but large entertainment events and professional sports events with numerous participants might continue to be risky. In other words, it is possible to monitor a partial relaxation of voluntary restrictions against going out consistent with a controlled outbreak including monitoring economic activity to maintain R(*t*)<1 but not extremely low. Apple data might support fine control of the outbreak and acceptable levels of social interaction.

In an unpublished study using Google data [9], a 70% reduction in going out was necessary to control the outbreak in Spain, which is different from the degree required in Japan. That in Spain was seven times higher than that in Japan. In fact, the outbreak in Japan was controlled using voluntary restrictions against going out without lockdown, yet the country's necessary degree of reduction in going out was seven times lower than Spain's.

A COVID-19 advisory council in Japan has required an 80% reduction since the emergency declaration on April 8, 2020 [10], and trips outside the home have decreased 40% from normal levels. However, the outbreak has been decreasing; it has become apparent that the requirement was stricter than necessary, which has at least produced some evidence of what a request of 80% reduction is likely to elicit from citizens. Our study results indicate that a mere 11% reduction might be sufficient to control an outbreak. In the early stage of the outbreak, R_0 of 0.6 was inferred [11], which indicated that an outbreak would never have occurred in Japan. Such estimates might also have led to misguided countermeasures in Japan that necessitated adherence to contact tracing for cluster detection.

Earlier researchers [12] who used worldwide Apple and Google data examined only linear terms, whereas results show that higher terms might be necessary for predicting R(*t*). As such, the higher predictive power from Apple data suggests that R(*t*) predicted using Apple data might be more reliable.

XSL•FO
RenderX

Particularly, the estimated R(*t*) over a few days was less precise because the virus incubation period is approximately 6 days on average, and increasing R(*t*) over time up to 6 days was a widely observed phenomenon. Conversely, Apple data are available the next day, and thus, the latest R(*t*) from Apple data might be predicted.

We used Apple data exclusively, rather than data from Google or other sites, because Apple began providing the earliest data, and it has also continued to provide data on a timely basis with only 2 days of delay. Google data are delayed longer than the Apple data, and although the NTT and JR are likely delayed by only 1 day or half a day, they began compiling their data later. Additionally, NTT and JR data were not published systematically or were not made available for public use. Therefore, Apple's data were most appropriate for evaluating the outbreak dynamics and predicting them prospectively.

This study has some limitations. First, R(*t*) was not determined for the number of contacts only. It depends on other circumstances such as the proportion of people who are susceptible and the infectiousness of people who are asymptomatic, and we did not incorporate these into our calculations. Considering such data would have required an SIR model that included asymptomatic cases, which we anticipate to be a challenge for future research.

Second, we examined this problem only for the entirety of Japan, but Apple also provides this information by prefecture; therefore, we could have extended our analyses to affected prefectures such as Tokyo, Osaka, and Hokkaido.

Third, it should be highlighted that Apple data show the proportions of users leaving their residences, but the data do not directly reflect numbers or rates of contacts; in other words, Apple data show no intensity of respective contacts. In fact, the measurements of contact intensity are difficult, but we consider meeting this challenge as a future research objective.

Fourth, although Apple data were better than those from other sources, Apple users might be limited to a particular population of young or healthy people, whereas information from NTT or JR might not be limited to those users. Thus, combining Apple data with data from these other sources might yield better results than those from this study, although these data have some disadvantages compared with the Apple data. Assessing and overcoming such limitations are challenges for future research efforts.

Conclusions

We demonstrated that mobility data from Apple Inc are useful for short-term prediction of R(*t*). Specifically, we determined that reducing trips from home by 11% was sufficient to maintain R(*t*)<1. We should examine the Apple data carefully to evaluate the effects of countermeasures.

Acknowledgments

We acknowledge the great efforts of all staff at public health centers, medical institutions, and other facilities who are fighting the spread associated with COVID-19. This study is based on the authors' opinions. It does not reflect any stance or policy of their professionally affiliated bodies.

Conflicts of Interest

None declared.

References

1. Tokyo governor urges people to stay indoors over weekend as virus cases spike. The Japan Times. 2020 Mar 25. URL: https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-40-coronavirus-cases/#.Xr4TV2eP604 [accessed 2020-05-14]
2. Sugishita Y, Kurita J, Sugawara T, Ohkusa Y. Preliminary evaluation of voluntary event cancellation as a countermeasure against the COVID-19 outbreak in Japan as of 11 March, 2020. medRxiv. Preprint posted online March 16, 2020. [doi: 10.1101/2020.03.12.20035220]
3. Sugishita Y, Kurita J, Sugawara T, Ohkusa Y. Forecast of the COVID-19 outbreak, collapse of medical facilities, and lockdown effects in Tokyo, Japan. medRxiv. Preprint posted online April 6, 2020. [doi: 10.1101/2020.04.02.20051490]
4. Mobility trends reports. Apple and CDC. URL: https://www.apple.com/covid19/mobility [accessed 2020-05-13]
5. Ohkusa Y, Sugawara T, Taniguchi K, Okabe N. Real-time estimation and prediction for pandemic A/H1N1(2009) in Japan. J Infect Chemother 2011 Aug;17(4):468-472. [doi: 10.1007/s10156-010-0200-3] [Medline: 21387184]
6. Press releases. Japan Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/newpage_10723.html [accessed 2020-05-14]
7. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, CDC COVID-19 Investigation Team. Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility - King County, Washington, March 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 03;69(13):377-381. [doi: 10.15585/mmwr.mm6913e1] [Medline: 32240128]
8. Survey of genomic molecular epidemiology survey for SARS-CoV-2. National Institute of Infectious Diseases. URL: https://www.niid.go.jp/niid/ja/basic-science/467-genome/9586-genome-2020-1.html [accessed 2020-05-14]

XSL•FO
RenderX

9. Arenas A, Cota W, Gomez-Gardenes J, Gomez S, Granell C, Matamalas JT, et al. Derivation of the effective reproduction number for COVID-19 in relation to mobility restrictions and confinement. medRxiv. Preprint posted online April 8, 2020. [doi: 10.1101/2020.04.06.20054320]
10. Sugiyama S. Japan state of emergency to cover Tokyo, Osaka and five other prefectures. The Japan Times. 2020 Apr 06. URL: https://www.japantimes.co.jp/news/2020/04/06/national/japan-state-of-emergency-covid-19/#.XpPANv37RaQ [accessed 2020-05-15]
11. Nishiura H, Oshitani H, Kobayashi T, Saito T, Sunagawa T, Matsui T, MHLW COVID-19 Response Team, et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). medRxiv. Preprint posted online March 3, 2020. [doi: 10.1101/2020.02.28.20029272]
12. Bergman NK, Fishman R. Mobility reduction and Covid-19 transmission rates. medRxiv. Preprint posted online May 10, 2020. [doi: 10.1101/2020.05.06.20093039]

Abbreviations

JR: Japan Railway
MLHW: Ministry of Labour, Health and Welfare
NTT: Nippon Telegraph and Telephone
SIR: susceptible-infected-recovered

Edited by G Eysenbach, R Kukafka; submitted 16.05.20; peer-reviewed by Y Nakamura, M Popovic; comments to author 30.09.20; revised version received 19.10.20; accepted 19.01.21; published 15.02.21.

Please cite as:
Kurita J, Sugishita Y, Sugawara T, Ohkusa Y
Evaluating Apple Inc Mobility Trend Data Related to the COVID-19 Outbreak in Japan: Statistical Analysis
JMIR Public Health Surveill 2021;7(2):e20335
URL: http://publichealth.jmir.org/2021/2/e20335/
doi:10.2196/20335
PMID:33481755

©Junko Kurita, Yoshiyuki Sugishita, Tamie Sugawara, Yasushi Ohkusa. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 15.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation

Mark Yu Zheng Wong[1,2], BA; Dinesh Visva Gunasekeran[2,3], MD; Simon Nusinovici[2,4], PhD; Charumathi Sabanayagam[2,4], PhD; Khung Keong Yeo[4,5], MD; Ching-Yu Cheng[2,3,4], MD, PhD; Yih-Chung Tham[2,4], PhD

[1]School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

[2]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

[3]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[4]Duke-NUS Medical School, National University of Singapore, Singapore, Singapore

[5]National Heart Centre Singapore, Singapore, Singapore

Corresponding Author:
Yih-Chung Tham, PhD
Singapore Eye Research Institute
Singapore National Eye Centre
20 College Road Discovery Tower
Level 6, The Academia
Singapore, 169856
Singapore
Phone: 65 6576 7200
Email: tham.yih.chung@seri.com.sg

Abstract

Background: The COVID-19 pandemic has led to urgent calls for the adoption of telehealth solutions. However, public interest and demand for telehealth during the pandemic remain unknown.

Objective: We used an infodemiological approach to estimate the worldwide demand for telehealth services during COVID-19, focusing on the 50 most affected countries and comparing the demand for such services with the level of information and communications technology (ICT) infrastructure available.

Methods: We used Google Trends, the Baidu Index (China), and Yandex Keyword Statistics (Russia) to extract data on worldwide and individual countries' telehealth-related internet searches from January 1 to July 7, 2020, presented as relative search volumes (RSV; range 0-100). Daily COVID-19 cases and deaths were retrieved from the World Health Organization. Individual countries' ICT infrastructure profiles were retrieved from the World Economic Forum Report.

Results: Across the 50 countries, the mean RSV was 18.5 (SD 23.2), and the mean ICT index was 62.1 (SD 15.0). An overall spike in worldwide telehealth-related RSVs was observed from March 11, 2020 (RSV peaked to 76.0), which then tailed off in June-July 2020 (mean RSV for the period was 25.8), but remained higher than pre-March RSVs (mean 7.29). By country, 42 (84%) manifested increased RSVs over the evaluation period, with the highest observed in Canada (RSV=100) and the United States (RSV=96). When evaluating associations between RSV and the ICT index, both the United States and Canada demonstrated high RSVs and ICT scores (≥70.3). In contrast, European countries had relatively lower RSVs (range 3.4-19.5) despite high ICT index scores (mean 70.3). Several Latin American (Brazil, Chile, Colombia) and South Asian (India, Bangladesh, Pakistan) countries demonstrated relatively higher RSVs (range 13.8-73.3) but low ICT index scores (mean 44.6), indicating that the telehealth demand outstrips the current ICT infrastructure.

Conclusions: There is generally increased interest and demand for telehealth services across the 50 countries most affected by COVID-19, highlighting the need to scale up telehealth capabilities, during and beyond the pandemic.

(JMIR Public Health Surveill 2021;7(2):e24445) doi:10.2196/24445

KEYWORDS

COVID-19; infodemiology; telehealth; telemedicine; internet

XSL•FO
RenderX

Introduction

COVID-19 was formally declared a pandemic by the World Health Organization (WHO) on March 11, 2020. As of September 7, the WHO has reported over 27 million cases, with a cumulative mortality rate of 3.26% [1]. In the context of infectious disease outbreaks such as the current COVID-19 pandemic, concerns regarding the overloading of health care facilties, coupled with the need to minimize patient and health care provider exposure in hospital care settings have led to calls for a shift from the traditional patient-physician face-to-face physical consultations to telehealth-based remote clinical services [2-6]. However, the magnitude of this major shift in health care management has yet to be evaluated. Public interest in and potential demand for telehealth services are relatively unknown [7,8]. This information gap poses challenges for health care providers to redesign strategies, institute new policies, and restructure manpower and infrastructure to address a potential "new wave" of clinical needs.

Infodemiology is a rapidly growing field of methodology in health informatics, which study trends in online search behavior and internet activity [9,10]. These methods provide new insights on population behavior and health-related phenomena, particularly during infectious disease outbreaks [11-13]. Google Trends (GT) and the Baidu Index are examples of infodemiological tools that researchers have used to analyze temporal and geographical trends in relative search volume (RSV) on the internet, with GT being the most prolific among published reports [14-16]. These tools have the additional advantage of providing real-time data, reflecting the immediate changes in population behavior in response to real-world events [9,10]. In the current climate of the COVID-19 pandemic, these tools have recently been used to investigate the overall public interest in COVID-19 [17], public fear of COVID-19 symptoms [18], and changes in behavioral attitudes toward activies such as social distancing and hand washing [19].

A recent paper by Hong et al [2] utilized a similar infodemiological approach to describe the increase in telehealth-related search volumes in the United States, up to March 2020. Building upon this work, we further broadened our current investigation toward a global perspective and extended the evaluation period beyond the initial wave of COVID-19 to include postlockdown periods and the reopening of countries' economy, society, and health care systems.

To provide a broader understanding of the current global interest and demand for telehealth, we used an infodemiological approach to investigate internet RSV as a proxy for public interest and demand for telehealth services in the 50 countries most affected by COVID-19. We described trends in telehealth-related RSVs across these countries over a 6-month period, spanning from the start of the pandemic to each country's lockdowns and their subsequent reopening. Finally, we compared the demand for telehealth with the level of information and communications technology (ICT) infrastructure for each country. These findings may provide valuable information for policy makers and health care providers to better cater to the new demands for telehealth services during COVID-19, and into the post–COVID-19 new normal.

Methods

Retrieving COVID-19 Key Dates and Confirmed Case Numbers

Real-world data on daily confirmed COVID-19 cases and deaths were retrieved on July 9, 2020, from the WHO's COVID-19 dashboard from January 1, 2020, until July 7, 2020 [1]. Worldwide data, as well as individual country-level data for the 50 countries with the highest cumulative confirmed COVID-19 case numbers (as of July 7, 2020), were also retrieved. Key dates of the COVID-19 pandemic were retrieved from the WHO's COVID-19 timeline and news reports of regional COVID-19–related events [20,21].

Retrieving Data From GT and Other Country-Specific Search Query Databases

GT provides data on volumes and patterns in online search behaviors of internet users [15]. It tracks keyword search queries that users enter into the Google search engine and presents information on the search query according to the specified time period and geographical location [22]. The search volume results are normalized and presented as an RSV index, wherein each data point is divided by the total searches performed in a specified geography within a given time range to provide relative comparisons [22]. The resulting output ranges from 0 to 100, with 100 indicating the maximum search interest in the selected time period and location. To comprehensively capture trends in online search behavior and infectious burden over an extended period, daily worldwide and country-specific GT data were retrieved over the first 6 months of the outbreak, from January 1 to July 7, 2020.

In addition to GT, the Baidu Index and the Yandex Wordstat Keyword Statistics Service were used to retrieve data on search queries in China and Russia, respectively. Baidu and Yandex are the predominant search engines used in China and Russia, respectively [14,23]. To facilitate direct visualization and comparisons with the RSV index obtained from GT, data from the Baidu Index and Yandex were similarly scaled to range from 0 to 100 [18]. In this work, total-RSV denotes the cumulative RSVs over the entire evaluation period, while average RSV represents an RSV value that was averaged over a specified period (eg, pre– or post–COVID-19 period; further descriptions are provided below).

Keyword Selection

For the GT analysis, we followed the detailed methodology described by Mavragani et al [10] for our keyword selection. First, different permutations of search terms and topics related to "telehealth" and "telemedicine" were searched to understand overarching trends in worldwide interest and to optimize keyword search combinations. We then conducted worldwide and country-specific GT searches using a baseline combination of English, Spanish, Russian, and French translations (chosen from the list of most commonly spoken languages worldwide) [24], in addition to translations in the native or official language

XSL•FO
RenderX

of that particular country. Mandarin Chinese is the second most spoken language in the world [24] but was not included in the baseline search combination since the majority of native Mandarin Chinese speakers reside in China and do not use Google as their main internet search tool. For the Baidu and Yandex search indexes, a combination of keywords in Mandarin Chinese (both traditional and simplified) and Russian, respectively, were used. The detailed keyword search strategy can be found in Table S1 in Multimedia Appendix 1.

Retrieval of Additional Telehealth-Related National Indicators

Data regarding the key dates of major public health responses such as lockdowns for each country were obtained from internet sources and news reports [25-27]. In addition, ICT data from the ICT adoption pillar of the Global Competitiveness Index 4.0 framework were obtained from the World Economic Forum Global Competitiveness Report 2019 and used to compare ICT level across countries [28]. The extracted ICT adoption index (ICT index) scores range from 0 to 100 (highest), wherein a higher score represents greater levels of networked infrastructure and higher regional usage and access to such infrastructure. Additional country-specific indices including GDP (gross domestic product) per capita ($US), literacy rates, and World Bank regional income groups were obtained from the World Bank DataBank [29].

As a proxy measure of the existing telehealth capacity of the respective countries, the Crunchbase (CB) [30] and CB Insights (CBI) [31] business analytic platforms were used to search for the prevalence of prominent telehealth providers within each of the 50 countries, and was used to define a CB/CBI score.

Analysis

First, to provide a more accurate observation of underlying trends and eliminate short-term fluctuations in data, the time trends for telehealth-related RSVs were smoothened by 7-day rolling intervals [32]. For countries with lower total-RSVs (<5), 14-day interval smoothing was used instead, as these countries are more susceptible to daily fluctuations (thus more "noise" in the trend data). Having a longer smoothing interval helps to minimize errors in estimations caused by these fluctuations over a short period. The RSVs were then plotted against daily COVID-19 confirmed cases and deaths (two separate Y axes), both worldwide and for each of the 50 countries.

Mean (SD) and median (IQR) values were used to provide summary statistics for key country parameters including total-RSVs, COVID-19 cases and deaths, and ICT indexes. The Pearson and Spearman correlation tests (when applicable) were used to further investigate associations between key parameters. The kurtosis/skewness and Shapiro-Wilk tests were used to determine normality of data distribution. P values of <.05 were considered as statistically significant.

The lockdown periods for each country were incorporated along with two key dates for reference—January 23, 2020, when China first imposed a lockdown in Hubei Province, and March 11, 2020, when the WHO declared COVID-19 a global pandemic. When evaluating changes in average RSV levels in the pre– and post–COVID-19 periods (ie, RSVs averaged over the period before and after COVID-19, respectively), we defined these periods based on the landmark date of March 11, 2020 (the WHO's COVID-19 pandemic declaration), except for China, where the pre–COVID-19 period was defined as before January 23, 2020. The ratios of the average pre– and post–COVID-19 RSV levels for each country were then calculated.

Lastly, bubble plots were used to illustrate the relationships between the total-RSVs of individual countries and various telehealth-related national indicators (GDP per capita [$US], literacy rates, the ICT index, and the CB/CBI score). Countries were grouped and color-coded according to World Bank regional classifications. All analyses and visualizations were conducted using Python (Python Software Foundation, version 3.7.4).

Results

Characteristics of the 50 countries, including the number of COVID-19 cases and deaths, telehealth-related RSVs, and ICT index values, are presented in Table 1. Across the 50 countries, the mean total-RSV was 18.5 (SD 23.2; median 9.20, IQR 5.75–18.68), and the mean ICT index score was 62.1 (SD 15.0; median 64.5, IQR 51.2–72.5). Figure 1 (top) shows a geographic choropleth map of the telehealth-related GT RSVs. North American countries had the highest total-RSVs (RSV=100 in Canada and RSV=96.6 in the United States). Within Europe, Switzerland (RSV=19.5) and Portugal (RSV=16.1) had the highest total-RSVs. For Latin America and the Carribean region, Chile (RSV=74.7) and Ecuador (RSV=69.0) had the highest total-RSVs. Likewise, the United Arab Emirates (RSV=40.2) scored the highest for the Middle East, South Africa (RSV=12.6) for sub-Saharan Africa, Bangladesh (RSV=41.4) for South Asia, and Singapore (RSV=41.4) for East Asia. Similarly, Figure 1 (bottom row) demonstrates geographic choropleth maps by the number of COVID-19 confirmed cases and deaths, respectively. Overall, among the evaluated countries, there were fair correlations between total-RSVs and COVID-19 cases (Pearson r=0.46, P<.001; Spearman ρ=0.29, P=.04) and deaths (r=0.39, P=.005; ρ=0.17, P=.25) (Figure 2).

XSL•FO
RenderX

Table 1. Key COVID-19 and telehealth-related parameters for the top 50 countries most affected by the pandemic.

Rank	Country	COVID-19 cases[a], n	COVID-19 deaths[a], n	Total-RSVs[b]	ICT[c] adoption index
1	United States	2,877,238	129,643	96.6	74.35
2	Brazil	1,603,055	64,867	29.9	58.06
3	India	719,665	20,160	13.8	32.11
4	Russia	694,230	10,494	10.3[d]	77.03
5	Peru	302,718	10,589	46.0	45.70
6	Chile	298,557	6384	74.7	63.13
7	United Kingdom	285,772	44,236	9.2	72.99
8	Mexico	256,848	30,639	6.9	55.03
9	Spain	251,789	28,388	6.9	78.21
10	Iran	243,051	11,731	4.6	50.85
11	Italy	241,819	34,869	9.2	64.49
12	Pakistan	234,509	4839	34.5	25.21
13	Saudi Arabia	213,716	1968	13.8	69.30
14	Turkey	206,844	5241	5.7	57.82
15	South Africa	205,721	3310	12.6	49.67
16	Germany	196,944	9024	5.7	69.98
17	Bangladesh	165,618	2096	41.4	39.14
18	France	159,568	29,831	5.7	73.66
19	Colombia	117,110	4064	32.2	49.89
20	Canada	105,536	8684	100.0	70.29
21	Qatar	100,345	133	13.8	83.78
22	China	85,345	4648	52.9[e]	78.49
23	Argentina	77,815	1523	11.5	57.99
24	Egypt	76,222	3422	9.2	40.57
25	Sweden	73,061	5433	3.42	87.78
26	Indonesia	64,958	3241	5.7	55.37
27	Belarus	63,804	429	4.6	__[f]
28	Ecuador	62,380	4821	69.0	47.62
29	Iraq	62,275	2567	0	—
30	Belgium	62,058	9774	3.4	67.02
31	United Arab Emirates	52,068	324	40.2	91.87
32	Netherlands	50,602	6119	8.0	76.29
33	Kuwait	50,644	373	0	69.57
34	Ukraine	49,607	1283	5.7	51.85
35	Kazakhstan	49,683	264	6.9	67.99
36	Oman	47,735	218	0	58.11
37	Philippines	46,333	1303	31.0	49.69
38	Singapore	44,983	26	41.4	87.11
39	Portugal	44,129	1620	16.1	71.24
40	Panama	38,149	747	13.8	50.06
41	Bolivia	39,297	1434	12.6	51.45

XSL•FO
RenderX

Rank	Country	COVID-19 cases[a], n	COVID-19 deaths[a], n	Total-RSVs[b]	ICT[c] adoption index
42	Dominican Republic	38,128	804	10.3	51.79
43	Poland	36,155	1521	9.2	65.41
44	Afghanistan	33,384	920	0	—
45	Switzerland	32,230	1685	19.5	78.58
46	Israel	30,055	331	9.2	67.56
47	Bahrain	29,821	98	0	67.19
48	Nigeria	29,286	654	10.3	33.39
49	Armenia	29,285	503	0	62.02
50	Romania	29,223	1768	9.2	72.01

[a]COVID-19 case and death numbers as of July 7, 2020.

[b]RSV: relative search volume. Total-RSVs were calculated over the evaluation period from January 1 to July 7, 2020.

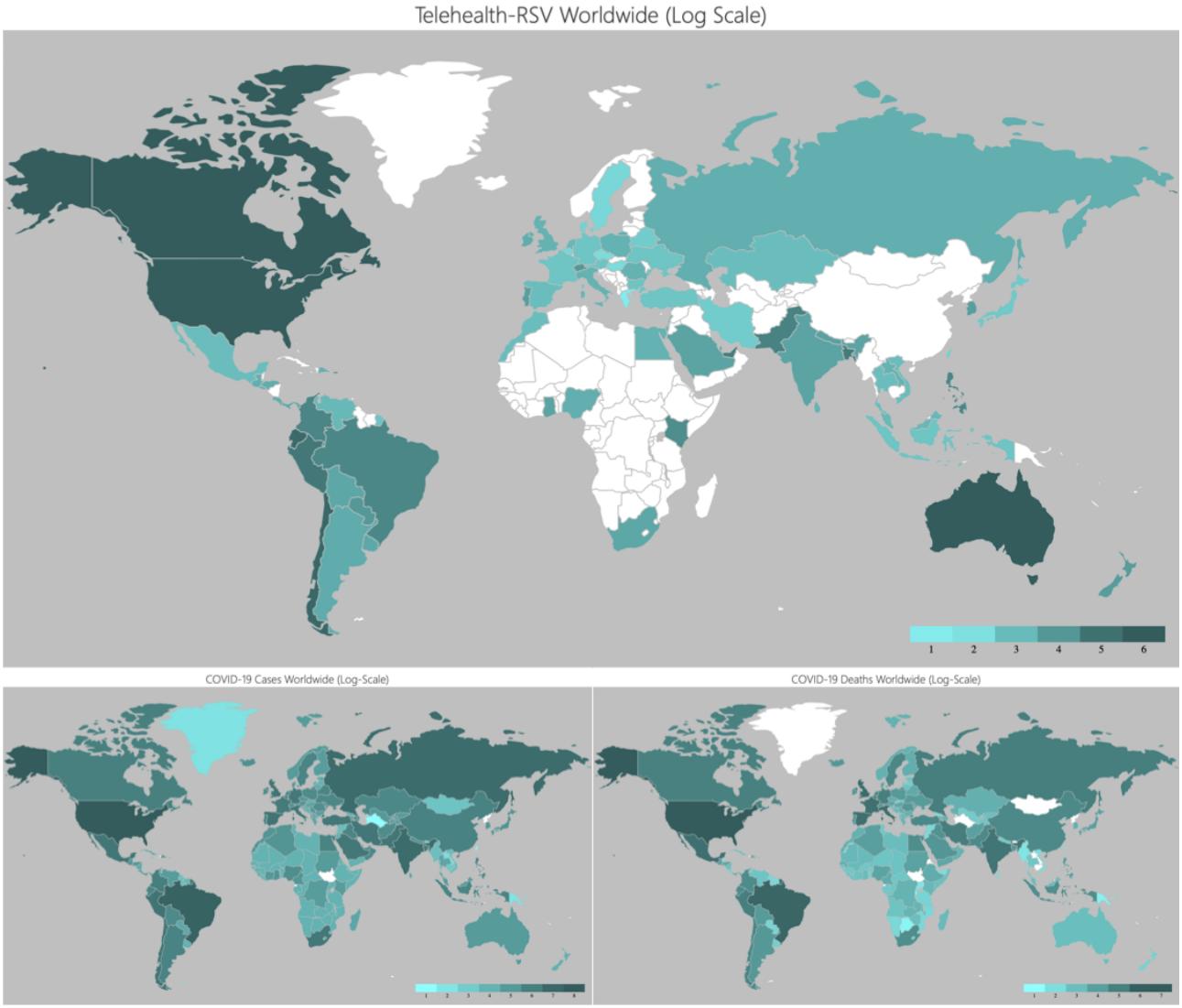
[c]ICT: information and communications technology.

[d]For Russia, the RSV value reflects search volumes as measured by GT only. Mean search volumes based on Yandex was 44.3 but was not listed in the table.

[e]For China, the RSV value reflects the mean RSV as measured by the Baidu Index. GT was not applicable for China.

[f]Not available.

Figure 1. Global choropleth map comparing telehealth-related relative search volumes (RSVs) (top), real-world COVID-19 confirmed cases (bottom left), and real-world COVID-19 confirmed deaths (bottom right).



XSL•FO
RenderX

Figure 2. Correlation between total telehealth-related relative search volumes (RSVs) against total COVID-19 cases (left) and deaths (right) per country.

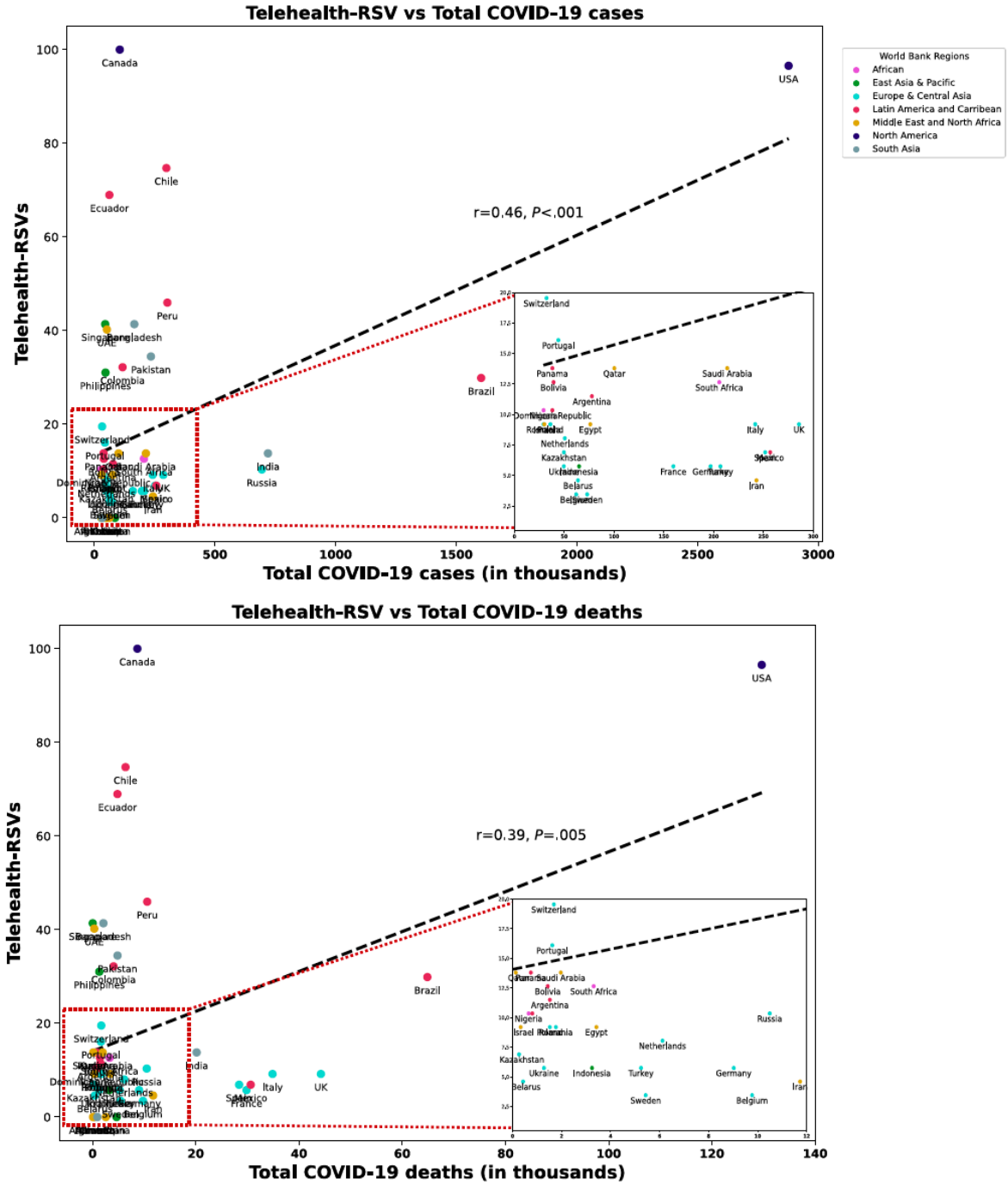


Figure 3 depicts the overall worldwide trends in telehealth-related RSVs during the study period, on a backdrop of accumulated COVID-19 cases and deaths. A surge in RSV levels can be observed from March 11, 2020, the date when the WHO officially declared the COVID-19 outbreak a pandemic, and culminates with an observed peak (RSV=76.0) in telehealth-RSV levels on March 24, 2020, with more than 10 times the pre–COVID-19 levels (mean 7.29). RSV levels then tailed off in June-July 2020 (mean 25.8) but still remained higher than pre–COVID-19 levels. For comparison, another curve representing RSVs for a search of "coronavirus" (in blue) was plotted alongside the telehealth-RSV curve (in red). The coronavirus-related RSVs demonstrated a similar trend—a sharp spike near March 11 and peaking on March 18.

Country-specific telehealth-RSV levels during the study period were similarly plotted in Multimedia Appendix 2. In total, 42 (84%) of the evaluated countries manifested spikes in telehealth-RSV levels over the evaluation period. When comparing the average pre– and post–COVID-19 RSV levels across these 42 countries, RSV levels increased by an average of 4.07 (SD 3.23) times post COVID-19 (Table S2, Multimedia Appendix 1). The remaining 8 countries were from Central Asia (Armenia) and the Middle East (Iran, Belarus, Iraq, Kuwait, Oman, Afghanistan, Bahrain), and had either no RSVs or no observable increased trends.

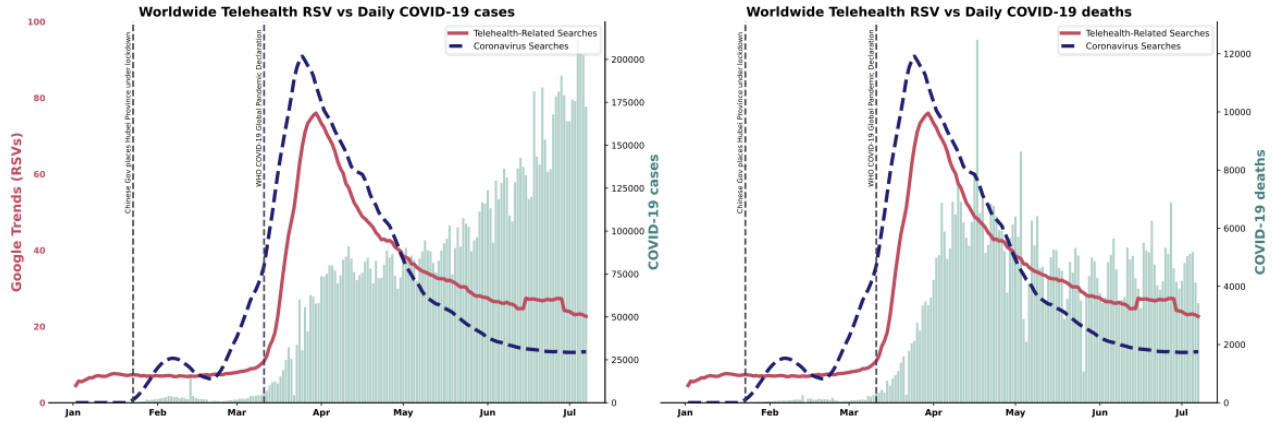
Of the 42 countries with increases in RSVs, the United States and Canada had the highest total-RSVs, showing sharp increases

XSL•FO
RenderX

in telehealth-RSV levels during early March, peaking in mid to late March, before decreasing and eventually plateauing to levels that were still higher than pre–COVID-19 levels (Multimedia Appendix 2). Compared to pre–COVID-19 levels, the average RSV levels increased by 5.83 times in the United States and by 2.69 times in Canada (Table S2, Multimedia Appendix 1). The remaining 40 countries also displayed increases in RSV levels over the evaluation period, albeit with less prominent spikes compared to the United States and Canada. These countries often approached peak RSVs more gradually over a longer period of months. These countries can be further subdivided into two categories based on the magnitude of RSV increases (Table S2, Multimedia Appendix 1). The first group observed large increases in average RSV levels (≥2.5 times increase compared to pre–COVID-19 levels; n=24). This group consisted largely of countries from Latin America (Brazil, Peru, Chile, Colombia, Argentina, Ecuador), South Asia (India, Bangladesh, Pakistan), East Asia (China, Indonesia, the Philippines), the Middle East (Saudi Arabia, Qatar, Egypt), and several European countries (Russia, Spain, Italy, Turkey, Romania, Ukraine). The second group of countries experienced smaller increases in average RSV levels compared to pre–COVID-19 levels (between 1 to 2.5 times; n=16), and largely comprised the United Kingdom, Germany, France, Netherlands, Portugal, Poland, Sweden, Belgium, Switzerland, Singapore, and Israel (Table S2, Multimedia Appendix 1). For most of these 40 countries, increases in RSV levels began in early March, either alongside or preceeding the rise in local COVID-19 case numbers, with the exception of China where the increase in Baidu RSVs was observed earlier during late January.

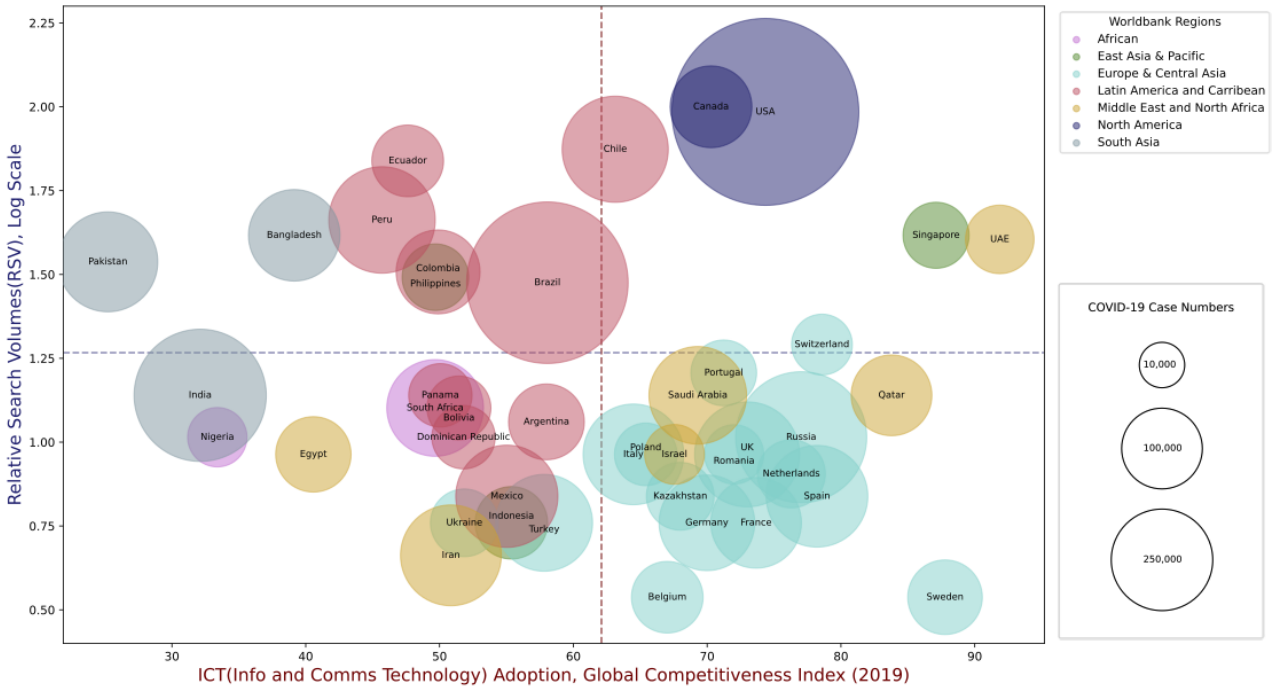
Figure 4 illustrates the relationship between telehealth total-RSVs and ICT index values across the evaluated countries. There was no significant correlation between RSVs and ICT scores (r=0.11, P=.45; ρ=–0.08, P=.59). However, broad clustering patterns among countries in similar regions can be observed. By using the mean RSV value (horizontal dashed line) and mean ICT index (vertical dashed line), we visually divided the plot into 4 quadrants and observed that the United States and Canada occupy the top right quadrant, with high total-RSVs (≥96.6) and ICTs (≥70.3). The United Arab Emirates and Singapore were also in the top right quadrant, with similarly high total-RSVs (40.2 and 41.4, respectively) and ICT scores (91.9 and 87.1, respectively). In contrast, while European countries generally had high ICT scores (range 51.9-87.8), they had lower total-RSVs (range 3.4-19.5). Latin American countries generally occupied a cluster near the middle of the bubble plot, with moderate levels of total-RSVs (6.9-74.7) and moderate ICT scores (45.7-63.1). On the other hand, South Asian countries generally had moderate to moderate-high total-RSVs (13.8-41.4), despite their low ICT scores (≤39.1). Similarly, Figures S2 and S3 in Multimedia Appendix 1 illustrate the relationship between telehealth-RSVs with (1) the relative literacy rate and (2) the GDP per capita across the 50 countries. Similar clustering patterns of countries (as in the ICT evaluation) were observed.

Figure 3. Worldwide time trends for telehealth-related relative search volumes (RSVs) (red) against daily COVID-19 cases (left) and deaths (right). The x axis represents time in individual days from January 1 to July 7, 2020. The left and right y axes represent Google Trends RSVs and COVID-19 cases or deaths, respectively. Blue and red trendlines represent "coronavirus" and telehealth-RSVs, respectively. Teal vertical bars represent daily COVID-10 cases or deaths. Black vertical lines represent two key dates: the start of the Hubei Province lockdown (January 23, 2020) and the declaration of COVID-19 as a pandemic by the World Health Organization (WHO) on March 11, 2020.



XSL•FO
RenderX

Figure 4. Relationship between telehealth-related relative search volumes (RSVs) versus information and communications technology (ICT) adoption index across the 50 countries most affected by COVID-19. Each country is represented as a data point and color-coded according to its World Bank region. The size of each plot reflects the accumulated total COVID-19 case numbers (as of July 7, 2020). The x axis represents the ICT adoption index, while the y axis represents the scaled (log) RSVs for each country. Vertical and horizontal dashed lines represent the mean values for the x and y axes, respectively. China and countries where RSV=0 (Google Trends) were not included in the plot.



Discussion

Principal Findings

In this study, we used data from GT, the Baidu Index, and Yandex Keyword Statistics to evaluate trends in telehealth demand during the first 6 months of the COVID-19 pandemic. To our knowledge, this is the first study to apply an infodemiological approach to investigate the potential public demand for telehealth in the 50 countries most affected by COVID-19. Our study further unraveled trends of demands alongside key COVID-19 events and varying levels of ICT infrastructure. Our findings suggest a general trend of increased demand for telehealth services across the evaluated countries during the COVID-19 pandemic. This trend was largely sustained beyond the initial wave, country lockdown periods, and subsequent reopenings. Our results suggest an ongoing and possible increased interest in telehealth services in the future as we enter a post–COVID-19 new normal phase. We also observed that current ICT infrastructure in several developing countries may lag behind this surging demand for telehealth. Our findings collectively indicate a pressing need to scale up telehealth capabalities in response to this growth in telehealth demand.

Our results demonstrate increased RSVs in most of the 50 evaluated countries. Among them, Canada and the United States had the highest total-RSVs and displayed a large spike in interest for telehealth services compared to most other countries. It is noteworthy that this pattern was also found in Australia (though not ranked within the top 50 countries), which had similarly high total-RSV values (RSV=98.9; data not shown in tables). Most other countries had less well-defined but observable increases in RSVs.

Interestingly, in many countries, RSV trends over time did not closely follow local COVID-19 case and death counts. However, RSVs often increased in early March, suggesting that the global announcement by the WHO on March 11, 2020, possibly had a great impact on the public's change in behavior to seek for telehealth options. Other studies investigating COVID-19–related search trends have reported that increases in RSVs often preceeded local COVID-19 cases and deaths [17,33]. It has been suggested that RSV trends tend to change in response to particular "index events" [18]. For instance, Rovetta et al [33] reported that significant increases in COVID-19–related RSVs were only observed after the WHO's COVID-19 pandemic announcement and the imposition of strict social distancing rules by governments.

Our study also found fair correlations between total-RSVs and a country's COVID-19 cases and deaths. We posit two possible explanations for this. First, larger numbers of COVID-19 cases and deaths could result in increased risk perception in the general population, thus driving greater interest in telehealth to minimize risk exposures. This is in line with other studies that have also reported that changes in search volumes toward other risk-minimizing activities, such as social distancing, hand washing, and face mask wearing, corresponded with increases in COVID-19 cases [19,33]. Another potential explanation would be that the increase in telehealth-related RSVs reflected actual needs of affected patients. The timing of the RSV spikes for each country often preceded the actual spike in COVID-19 cases, suggesting that changes in search activity were probably

XSL•FO
RenderX

more likely due to changes in public perception as the pandemic situation evolved. Nevertheless, further real-world investigations in the form of custom-designed questionnaires would be needed to further elucidate this aspect.

The bubble plots presented in this report further investigated the relationships between total-RSVs and various telehealth-related indices (ICT index, GDP per capita, literacy rates). These indices were evaluated as proxies for the capacity of a country to adopt and operationalize new telehealth services [34-36]. The clustering patterns observed in Figure 4 and Figures S2-S4 in Multimedia Appendix 1 may enable the classification of regions or countries based on their relative demands and capacity for telehealth services. For example, regions with the potential for rapid growth and adoption of telehealth are those with higher literacy levels, better ICT infrastructure, and higher RSVs (Figure S4, Multimedia Appendix 1). These countries include Argentina, Chile, Qatar, the United Arab Emirates, Saudi Arabia, and Singapore [37]. On the other hand, countries such as Colombia, Peru, Ecuador, Bangladesh, Pakistan, and the Philippines demonstrated growing interest in telehealth (moderate to high RSVs) but are limited by existing ICT capability (low to moderate ICT index).

Despite high ICT indexes and strong existing telehealth capacity (high CB/CBI scores; Figure S4, Multimedia Appendix 1) in a majority of European countries [38], the RSVs of these countries were relatively low compared to other developed western countries such as Canada and the United States. This observation may be explained by the availability and easy accessibility of existing telehealth services, which may also impact the population's information-seeking behavior. For instance, in European countries that are at a more advanced stage of telehealth adoption, telehealth awareness and literacy may already be present, and patients may be directly seeking telehealth services from providers instead of conducting online queries in search engines [38]. Other factors that may also influence search volumes include different models of health care systems. For example, patients in the United Kingdom may contact their general practitioner or seek online consultations readily and directly through the NHS App [39]. Public health communications play a role as well. Sweden, for instance, contrarily downplayed the significance of the pandemic [40], eschewing lockdown measures, which may also explain why Sweden had the lowest RSVs among the evaluated European countries [40,41].

Our study offers useful insights into the short- and medium-term trends in telehealth demand in response to the COVID-19 pandemic. The trend of sustained increase observed in the Baidu Index RSV in China provides preliminary indications that telehealth demand will likely remain higher than in the pre–COVID-19 era. This trend is especially likely given that China can be considered to have entered its post–COVID-19 recovery phase, with minimal new cases reported over the last 3 months of our study period. On the other hand, given the resurgence of second or third waves of COVID-19 in several countries in recent months [1,42], it is foreseeable that telehealth demand may surge yet again or remain higher than the pre–COVID-19 period. Nonetheless, to better ascertain the long-term impact of COVID-19 on telehealth interest, further evaluation over a longer period is required.

Strengths and Limitations

The key strengths of our study include our unique infoveillance approach to evaluate the public demand for telehealth services, capturing real-time responses to key COVID-19 pandemic events [11,16,43]. Second, our study provided extensive coverage of the 50 most affected countries worldwide, and evaluated these countries over a long duration (6 months), thus providing more concrete insights on trends. Third, we also included the additional use of the Baidu Index and Yandex Keyword Statistics to further investigate RSVs in China and Russia. The high degree of correlation between GT and Yandex Keyword Statistics for RSVs in Russia (r=0.875, P<.001) further confers a degree of reliability to our results.

This study also has a few limitations. First, it should be acknowledged that infodemiological approaches can only serve as a proxy for estimating the true demand for telehealth services. Second, as suggested earlier, although search engines provide a good catchment of overall interest, there may be alternative channels for the public to seek health care–related information, including their general practitioners, insurance services, or directly from telehealth service providers. In addition, it should also be acknowledgd that it is not possible to include all terms utilized by internet users to search for information on telehealth. Third, not all countries use Google as their primary search engine. Hence, using GT as a main proxy for overall demand when comparing across countries may be subjected to bias for some countries. Fourth, overall RSV trends may be confounded by different sampling profiles between countries, wherein factors including education and internet access may skew the representativeness of the RSV samples for each country. Additionally, infodemiology platforms such as GT and the Baidu Index present data as normalized RSVs rather than as absolute search counts, thus limiting direct comparisons of RSV data extracted from different sources of search engines [10]. Fifth, to reduce subjectivity, across all included countries (except for China), we standardized the definition of the post–COVID-19 period based on the date when the WHO declared COVID-19 as a global pandemic. Although it might have been ideal to individually quantify each country's pre– and post–COVID-19 periods, it was difficult to do so, considering each country had a different community transmission trajectory. Furthermore, changes in search volumes were often in response to not just local events (lockdowns, local increases in COVID-19 cases, etc) but also global events such as the WHO's pandemic announcement. Lastly, to better ascertain the long-term impact of COVID-19 on telehealth interest, further evaluation over a longer period is required.

Conclusion

Telehealth is a major health technology solution that has gained further traction during the COVID-19 pandemic. We identified increased demand for telehealth services across the 50 countries most affected by COVID-19. We also found indications that several developing countries may still have suboptimal ICT infrastructure to cope with this surge in telehealth demand. These findings underscore a pressing need for policy makers

XSL•FO
RenderX

and health care providers to scale up telehealth infrastructure and capability, during and beyond COVID-19.

Conflicts of Interest

DVG reports equity investment in AskDr, Doctorbell (acquired by MaNaDr, Mobile Health), VISRE, and Shyfts. He also reports appointments as Physician leader (Telemedicine) at Raffles Medical Group, Senior Lecturer (Medical Innovation) at the National University of Singapore (NUS).

Multimedia Appendix 1
Supplementary Information.
[PDF File (Adobe PDF File), 1608 KB - publichealth_v7i2e24445_app1.pdf]

Multimedia Appendix 2
Time-trends for Telehealth-RSVs against daily COVID-19 cases in the top 50 countries most-affected by COVID-19. Countries arranged from left to right in order of total number of reported COVID-19 cases. x-axis represents time in individual days from January 1 to July 7, 2020. Left and right y-axes represent GT-RSVs and COVID-19 case numbers respectively. Red trendlines represent telehealth-related RSVs as measured by Google trends. Vertical bars in teal represent daily COVID-19 cases. Black vertical lines represent 2 key dates: the start of the Hubei Province lockdown (January 23, 2020) and the declaration of COVID-19 as a Pandemic by the World Health Organization (March 11, 2020). Shaded yellow regions represent country-specific lockdown or restriction periods. *Blue trendlines for China and Russia represent telehealth-related RSVs as measured by Baidu and Yandex respectively.
[PDF File (Adobe PDF File), 763 KB - publichealth_v7i2e24445_app2.pdf]

References

1. WHO Covid-19 Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-07-07]
2. Hong Y, Lawrence J, Williams D, Mainous I. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. JMIR Public Health Surveill 2020 Apr 07;6(2):e18961 [FREE Full text] [doi: 10.2196/18961] [Medline: 32250963]
3. Young JD, Abdel-Massih R, Herchline T, McCurdy L, Moyer KJ, Scott JD, et al. Infectious Diseases Society of America Position Statement on Telehealth and Telemedicine as Applied to the Practice of Infectious Diseases. Clin Infect Dis 2019 Apr 24;68(9):1437-1443. [doi: 10.1093/cid/ciy907] [Medline: 30851042]
4. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
5. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
6. Webster P. Virtual health care in the era of COVID-19. The Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
7. Blandford A, Wesson J, Amalberti R, AlHazme R, Allwihan R. Opportunities and challenges for telehealth within, and beyond, a pandemic. The Lancet Global Health 2020 Nov;8(11):e1364-e1365 [FREE Full text] [doi: 10.1016/S2214-109X(20)30362-4] [Medline: 32791119]
8. Pappot N, Taarnhøj GA, Pappot H. Telemedicine and e-Health Solutions for COVID-19: Patients' Perspective. Telemed J E Health 2020 Jul 24;26(7):847-849. [doi: 10.1089/tmj.2020.0099] [Medline: 32329654]
9. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
10. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
11. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
12. Eysenbach G. SARS and population health technology. J Med Internet Res 2003;5(2):e14 [FREE Full text] [doi: 10.2196/jmir.5.2.e14] [Medline: 12857670]
13. Zeraatkar K, Ahmadi M. Trends of infodemiology studies: a scoping review. Health Info Libr J 2018 Jun;35(2):91-120. [doi: 10.1111/hir.12216] [Medline: 29729073]
14. Baidu Index. Baidu. URL: http://zhishu.baidu.com/v2/index.html#/ [accessed 2020-07-07]
15. Google Trends. Google LLC. URL: https://trends.google.com/ [accessed 2020-07-07]
16. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]

XSL•FO
RenderX

17. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
18. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. JMIR Public Health Surveill 2020 May 21;6(2):e19702 [FREE Full text] [doi: 10.2196/19702] [Medline: 32401211]
19. Springer S, Menzel LM, Zieger M. Google Trends provides a tool to monitor population concerns and information needs during COVID-19 pandemic. Brain Behav Immun 2020 Jul;87:109-110 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.073] [Medline: 32360607]
20. WHO Response Timeline to COVID-19. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline [accessed 2020-07-07]
21. Muccari R, Chow D, Murphy J. Coronavirus timeline: Tracking the critical moments of Covid-19. NBC. 2020 Mar 10. URL: https://www.nbcnews.com/health/health-news/coronavirus-timeline-tracking-critical-moments-covid-19-n1154341 [accessed 2020-07-08]
22. FAQ about Google Trends Data. Google LLC. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2021-02-02]
23. Yandex Keyword Statistics. Yandex. URL: https://wordstat.yandex.com/ [accessed 2020-07-07]
24. Eberhard DM, Simons GF, Fennig CD. Ethnologue: Languages of the World. Dallas, Texas: SIL International; 2020.
25. Coronavirus: What are the lockdown measures across Europe? DW. 2020 Apr 14. URL: https://www.dw.com/en/coronavirus-what-are-the-lockdown-measures-across-europe/a-52905137 [accessed 2020-07-16]
26. Global Covid-19 Lockdown Tracker. Aura Vision. URL: https://auravision.ai/covid19-lockdown-tracker/ [accessed 2020-07-16]
27. Mourad M, Lewis A. Egypt declares two-week curfew to counter coronavirus. Reuters. 2020 Mar 24. URL: https://www.reuters.com/article/us-health-coronavirus-egypt/egypt-declares-two-week-curfew-to-counter-coronavirus-idUSKBN21B1MR [accessed 2020-07-18]
28. Schwab K. The Global Competitiveness Report 2019. World Economic Forum. 2019. URL: http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf [accessed 2021-02-02]
29. DataBank. The World Bank. URL: https://databank.worldbank.org/home.aspx [accessed 2020-07-19]
30. Crunchbase. URL: https://www.crunchbase.com/ [accessed 2020-07-20]
31. CBI Insights. URL: https://www.cbinsights.com/ [accessed 2020-07-20]
32. Arce GR. Nonlinear Signal Processing: A Statistical Approach. Hoboken, NJ: John Wiley & Sons, Inc; 2004.
33. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
34. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
35. Liu Z, Ng M, Gunasekeran DV, Li H, Ponampalam K, Ponampalam R. Mobile technology: Usage and perspective of patients and caregivers presenting to a tertiary care emergency department. World J Emerg Med 2020;11(1):5-11 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2020.01.001] [Medline: 31892997]
36. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
37. Gutierrez G. Medicare, the Internet, and the future of telemedicine. Crit Care Med 2001 Aug;29(8 Suppl):N144-N150. [doi: 10.1097/00003246-200108001-00002] [Medline: 11496036]
38. Global diffusion of eHealth: Making universal health coverage achievable. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf [accessed 2021-02-02]
39. Online consultations. NHS Digital. 2020. URL: https://digital.nhs.uk/services/nhs-app/prepare-your-practice-for-connection-to-the-nhs-app/online-consultations-in-the-nhs-app [accessed 2020-08-09]
40. Giesecke J. The invisible pandemic. The Lancet 2020 May;395(10238):e98. [doi: 10.1016/s0140-6736(20)31035-7]
41. King C, Einhorn L, Brusselaers N, Carlsson M, Einhorn S, Elgh F, et al. COVID-19—a very visible pandemic. The Lancet 2020 Aug;396(10248):e15. [doi: 10.1016/s0140-6736(20)31672-x]
42. Stevis-Gridneff M. France and Germany Lock Down as Second Coronavirus Wave Grows. The New York Times. 2020 Oct 28. URL: https://www.nytimes.com/2020/10/28/world/europe/france-germany-coronavirus-covid.html#:~:text=BRUSSELS%20%E2%80%94%20France%20announced%20a%20second,undoing%20hopes%20of%20economic%20recovery [accessed 2021-02-02]
43. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]

XSL•FO
RenderX

Abbreviations

CB: Crunchbase
CBI: CB Insights
GDP: gross domestic product
GT: Google Trends
ICT: information and communications technology
RSV: relative search volume
WHO: World Health Organization

Edited by T Sanchez; submitted 20.09.20; peer-reviewed by CW Ong, A Rovetta; comments to author 25.10.20; revised version received 10.11.20; accepted 14.12.20; published 19.02.21.

Please cite as:
Wong MYZ, Gunasekeran DV, Nusinovici S, Sabanayagam C, Yeo KK, Cheng CY, Tham YC
Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation
JMIR Public Health Surveill 2021;7(2):e24445
URL: http://publichealth.jmir.org/2021/2/e24445/
doi:10.2196/24445
PMID:33605883

©Mark Yu Zheng Wong, Dinesh Visva Gunasekeran, Simon Nusinovici, Charumathi Sabanayagam, Khung Keong Yeo, Ching-Yu Cheng, Yih-Chung Tham. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 19.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Knowledge, Attitudes, and Practices Related to COVID-19 in Iran: Questionnaire Study

Mohsen Abbasi-Kangevari[1], MD, MPH; Ali-Asghar Kolahi[1], MD, MPH; Seyyed-Hadi Ghamari[1], MD, MPH; Hossein Hassanian-Moghaddam[1], MD, FACMT

Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Corresponding Author:
Ali-Asghar Kolahi, MD, MPH
Social Determinants of Health Research Center
Shahid Beheshti University of Medical Sciences
Koodakyar Ave and Daneshju Blvd
Tehran
Iran
Phone: 98 2122220980
Email: a.kolahi@sbmu.ac.ir

Abstract

Background: The COVID-19 pandemic is a rapidly growing outbreak, the future course of which is strongly determined by people's adherence to social distancing measures.

Objective: The objective of this study was to determine the knowledge level, attitudes, and practices of the Iranian population in the context of COVID-19.

Methods: A nationwide study was conducted from March 24 to April 3, 2020, whereby data were collected via an online self-administered questionnaire.

Results: Responses from 12,332 participants were analyzed. Participants' mean knowledge score was 23.2 (SD 4.3) out of 30. Most participants recognized the cause of COVID-19, its routes of transmission, its symptoms and signs, predisposing factors, and prevention measures. Social media was the leading source of information. Participants recognized the dangers of the situation and felt responsible for following social distancing protocols, as well as isolating themselves upon symptom presentation. Participants' mean practice score was 20.7 (SD 2.2) out of 24. Nearly none of the respondents went on a trip, and 92% (n=11,342) washed their hands before touching their faces.

Conclusions: Knowledge of COVID-19 among people in Iran was nearly sufficient, their attitudes were mainly positive, and their practices were satisfactory. There is still room for improvement in correcting misinformation and protecting people from deception. Iranians appear to support government actions like social distancing and care for their and others' safety.

(JMIR Public Health Surveill 2021;7(2):e21415) doi:10.2196/21415

KEYWORDS

COVID-19; SARS-CoV-2; attitudes; coronavirus; knowledge; perceptions; practices; Iran

Introduction

The COVID-19 pandemic is a rapidly growing outbreak that has resulted in more than 13 million reported cases worldwide, as of June 2020 [1]. In response to the evolving situation, physical distancing, school closures, and lockdown measures have been implemented in many countries to contain the outbreak and prevent the overburdening of health care systems. COVID-19 has short-term and long-term socioeconomic consequences affecting the health of communities, which are not yet fully understood [2].

Since the beginning, the spread of misinformation primarily via social media became a significant concern among health authorities. Therefore, the World Health Organization launched a webpage addressing misbeliefs via sharable media [3]. There have been unsafe or stigmatizing attitudes toward COVID-19 or patients affected by it. In addition, there have been reports on dangerous or irresponsible practices among people, which could put communities at risk.

Given the high transmissibility of COVID-19, many control measures have been implemented worldwide. Nevertheless, the future course of the pandemic is strongly determined by people's

XSL•FO
RenderX

adherence to control measures [4], which highlights the essential role of proper knowledge, responsible attitudes, and safe practices [5].

Outbreak management requires people to follow social distancing measures, which occurs when people defer to health authorities due to proper knowledge, positive attitudes, and safe practices. Therefore, the educational needs of people should be addressed to provide training programs across the community, and their current knowledge level, attitudes, and practices need to be determined quickly, considering the rapid spread of COVID-19 [1]. The objective of this study was therefore to determine the level of knowledge, attitudes, and practices related to COVID-19 among the Iranian population via a rapid online survey.

Methods

This community-based national study was approved by the Ethical Committee of Shahid Beheshti University of Medical Sciences (IR.SBMU.RETECH.REC.1399.020). Participation was anonymous and voluntary.

Setting and Sampling

This survey was conducted from March 24 to April 3, 2020, in Iran. We collected data via an online self-administered questionnaire using Google Docs and circulated an online invitation post with a link to the questionnaire on popular social networks in Iran, including Telegram, Instagram, WhatsApp, Twitter, and LinkedIn. In the early days of the online circulation of the questionnaire, we requested medical students from all 31 provinces of Iran to send the invitation to others. Participants were Iranian adults who agreed to participate in the study and were currently living in Iran.

Variables and the Questionnaire

Variables included sociodemographic characteristics, knowledge about COVID-19, attitudes toward COVID-19, and practices during the pandemic. Sociodemographic characteristics included participants' age, sex, ethnicity, literacy, marital status, number of living children, smoking status, knowing someone with COVID-19, and province of residence.

The knowledge section included questions about symptoms of COVID-19, signs, predisposing factors related to severe outcomes of COVID-19, statements on general knowledge, a self-evaluation of one's level of knowledge, information sources, and one's need for further training. The attitudes section included statements about participants' attitudes toward COVID-19. In response to these statements, participants chose an item using a 5-point Likert scale. The practices section included questions about self-isolation and care seeking, preventive measures, use of face masks when leaving one's home, handwashing, leaving town, and general practices related to COVID-19.

A panel of experts evaluated the content validity of the questionnaire. An item discrimination analysis was conducted for each scale to eliminate items deemed too tricky or easy. Factor analysis was performed for factor structure. Separate test-retest analyses were conducted over 2 weeks for the three scales of the questionnaire. The test-retest correlation for the knowledge scale was 0.86. The Kuder-Richardson-20 was used to prevent an overestimation of internal consistency and resulted in a coefficient of 0.85. The test-retest correlation for the attitudes scale was 0.87, and the coefficient alpha was .89. The test-retest correlation for the practice scale was 0.91, and the coefficient alpha was .92. The pilot survey was conducted on 50 men and 50 women, who were recruited online via the convenience sampling method.

Data Analysis

One point was awarded to each correct answer to score participants' knowledge about COVID-19. To analyze participants' attitudes, responses "I strongly agree" and "I agree" were combined to signify agreement; responses "I strongly disagree" and "I disagree" were combined to indicate disagreement. To analyze participants' practices, 1 point was awarded for undertaking each correct measure or for not undertaking each wrong measure. For the key proportions using the exact binomial distribution, the 95% CI was reported. Categorical variables were analyzed via the chi-square test. For analyzing the differences between the means of two groups and or those of three or more groups, independent-sample *t* tests and one-way analysis of variance (ANOVA) were used, respectively. Statistical analyses were performed using SPSS (version 21, IBM Corp). A probability level of less than .05 was considered significant.

Results

Responses from 12,332 participants were analyzed. The mean age was 33.2 (SD 9.9) years and ranged from 18-88 years (women: 32.9 [SD 9.5] years; men: 33.5 [SD 10.8] years). As many as 6403 (51.9%) participants reported that they knew someone with COVID-19. Other sociodemographic characteristics are presented in Table 1.

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of participants.

Variable	Participants, n (%)
Sex	
Female	7629 (61.9)
Male	4703 (38.1)
Marital status	
Never married	4502 (36.5)
Engaged	282 (2.3)
Married	7183 (58.2)
Divorced	296 (2.4)
Widowed	69 (0.6)
Ethnicity	
Fars	7575 (61.4)
Turk/Azari	2417 (19.6)
Kurd	723 (5.8)
Lor	696 (5.6)
Other	921 (7.6)
Literacy	
Illiterate	18 (0.1)
Primary school	449 (3.6)
Secondary school	518 (4.2)
High school diploma	2195 (17.8)
Associate degree	964 (7.8)
Bachelor	4962 (40.3)
Master	2262 (18.3)
PhD and higher	945 (7.7)
Hawzeh	19 (0.2)

Knowledge

General Knowledge About COVID-19

The majority of participants identified the cause of COVID-19, its routes of transmission, and the role of self-isolation (Table 2).

The mean knowledge score was 23.2 (SD 4.3; range 0-30), out of a total of 30 (77%). The mean knowledge score of those with high school diplomas or higher degrees was significantly higher compared to those with less education. No significant differences were observed between knowledge scores and participants' age, sex, province of residence, marital status, number of children, smoking status, knowing someone with COVID-19, having a household member with COVID-19, having the disease themselves, or having a family member passing away due to COVID-19 (Table 3).

XSL•FO
RenderX

Table 2. Participants' responses to questions about general knowledge related to COVID-19.

Statement	Response		
	"It's true," n (%)	"It's false," n (%)	"I don't know," n (%)
COVID-19 is transmitted via respiratory droplets and infected surfaces.	12,140 (98.4)	66 (0.5)	126 (1.1)
Patients with COVID-19 and their close contacts should be quarantined >14 days.	12,035 (97.6)	61 (0.5)	236 (1.9)
COVID-19 could be transmitted from an infected travel companion.	11,991 (97.2)	167 (1.4)	176 (1.4)
COVID-19 is caused by a virus.	11,959 (97.0)	88 (0.7)	285 (2.3)
There are still no medications available for COVID-19 prevention.	11,074 (89.8)	357 (2.9)	901 (7.3)
There is currently no specific treatment available for COVID-19.	11,009 (89.3)	377 (3.1)	946 (7.6)
Handwashing with water alone is not enough for disinfection.	9916 (80.4)	2115 (17.2)	301 (2.4)
Solely using the hand dryer cannot kill the pathogen.	9093 (73.7)	2018 (16.4)	1221 (9.9)
Vaccination against pneumonia or influenza imposes no protection against COVID-19.	8956 (72.6)	720 (5.9)	2656 (21.5)
Eating garlic does not kill the pathogen.	8269 (67.1)	1812 (14.7)	2251 (18.2)
COVID-19 is not transmitted via insects.	6762 (54.8)	1584 (12.9)	3986 (32.3)
Viruses in the Coronaviridae family can cause zoonotic diseases.	6242 (50.6)	3284 (26.6)	2806 (22.8)
Regularly rinsing nostrils with saline has a protective effect against COVID-19.	5789 (46.9)	4014 (32.6)	2529 (20.5)
Mouth washing has a protective effect against COVID-19.	3705 (30.0)	5436 (44.1)	3191 (25.9)
COVID-19 always has mild symptoms.	2763 (22.4)	7332 (59.5)	2237 (18.1)
Using air humidifiers could kill the agent at home.	2543 (21.4)	7553 (61.3)	2136 (17.3)
Hand rubbing with alcohol-based solutions is more effective than handwashing with soap.	1936 (15.7)	9400 (76.2)	996 (8.1)
The cause of COVID-19 is the influenza virus.	1319 (10.7)	9449 (76.6)	1564 (12.7)
The severity of COVID-19 is always comparable to the common cold.	1265 (10.3)	10,238 (83.0)	829 (6.7)
People who recovered from COVID-19 will never get the disease again.	1129 (9.2)	8496 (68.8)	2707 (22.0)
Patients with COVID-19 cannot transmit the disease unless they have a fever.	1107 (8.9)	10,033 (81.4)	1192 (9.7)
All patients with COVID-19 require hospital admission and treatment.	885 (7.1)	10,809 (87.7)	638 (5.2)
Taking a shower with hot water can kill the agent inside the body.	858 (7.0)	10,265 (83.2)	1209 (9.8)
A person with no cough, fever, or dyspnea cannot carry the pathogen.	801 (6.5)	10,999 (89.2)	532 (4.3)
Children do not get COVID-19.	766 (6.2)	10,832 (87.8)	734 (6.0)
Going to a steam room or sauna can kill the agent inside the body.	599 (4.9)	10,357 (83.9)	1376 (11.2)
Warming the upper airways with a blow dryer can kill the agent inside the body.	446 (3.6)	11,017 (89.3)	869 (7.1)
Drinking alcoholic beverages could kill the pathogen inside the body.	209 (1.7)	11,420 (92.6)	703 (5.7)
A person who recovered from COVID-19 does not need to follow prevention protocols.	78 (0.6)	11,923 (96.7)	331 (2.7)

XSL•FO
RenderX

Table 3. Determinants of the COVID-19 general knowledge score among participants.

Variable	Score (SD)	95% CI
Sex		
Female	23.4 (4.1)	23.3-23.5
Male	22.8 (4.5)	22.7-22.9
Age		
<40 years	23.2 (4.2)	23.1-23.3
≥40 years	23.1 (4.4)	22.9-23.3
Marital status		
Married	23.2 (4.2)	23.1-23.3
Never married	23.2 (4.3)	23.1-23.4
Divorced	23.1 (4.6)	22.6-23.6
Engaged	22.5 (5.0)	22.1-23.2
Widowed	21.5 (5.9)	19.8-22.6
High school diploma		
Yes	23.3 (4.2)	23.2-23.4
No	20.4 (5.4)	19.9-20.9
Number of children		
0	23.3 (4.3)	23.3-23.4
1	23.4 (4.1)	23.2-23.5
2	23.0 (4.2)	22.9-23.2
≥3	22.5 (4.9)	22.1-22.9
Smoking		
Yes	22.8 (4.6)	22.6-23.2
No	23.3 (4.2)	23.1-23.3
Former smoker	22.9 (4.4)	22.5-23.4
Knowing someone with COVID-19		
Yes	23.6 (4.1)	23.5-23.7
No	23.4 (4.4)	23.3-23.5
Household member with COVID-19		
Yes	23.4 (4.4)	23.0-23.9
No	23.2 (4.3)	23.1-23.3
Has or has had COVID-19		
Yes	22.8 (5.0)	22.1-23.5
No	23.2 (4.2)	23.2-23.3
Not sure	22.9 (4.6)	22.7-23.2
Deceased family member due to COVID-19		
Yes	23.5 (4.2)	23.3-23.6
No	23.2 (4.3)	23.1-23.3

Symptoms of COVID-19

More than 90% of participants recognized the most common symptoms of COVID-19: fever, dyspnea, and cough. In addition, 98% (n=11,846) knew that worsening dyspnea could be a red flag for severe disease outcomes. However, 221 (1.8%) reported no knowledge of the general symptoms, and 280 (2.3%) stated they did not know what the symptoms of severe disease were (Table 4).

XSL•FO
RenderX

Table 4. Participants' knowledge of symptoms of COVID-19 and severe course of disease.

Symptoms	Response	
	Yes, n (%)	No, n (%)
Symptoms (n=12,111)		
Fever	11,567 (95.5)	544 (4.5)
Dyspnea	11,403 (94.2)	708 (5.8)
Cough	11,030 (91.1)	1081 (8.9)
Loss of smell or taste	8707 (71.9)	3404 (28.1)
Myalgia	6928 (57.2)	5183 (42.8)
Malaise	6678 (55.1)	5433 (44.9)
Sore throat	6478 (53.5)	5633 (46.5)
Diarrhea	4839 (40.0)	7272 (60.0)
Loss of appetite	4435 (36.6)	7676 (63.4)
Sneeze	2613 (21.6)	9498 (78.4)
Rhinorrhea	2590 (21.4)	9521 (78.6)
Loss of consciousness	2075 (17.1)	10,036 (82.9)
Confusion	1968 (16.2)	10,143 (83.8)
Signs (n=12,052)		
Worsening dyspnea	11,846 (98.3)	206 (1.7)
Worsening coughs	8564 (71.1)	3488 (28.9)
Fever >5 days	8252 (68.5)	3800 (31.5)
Loss of consciousness	2605 (21.6)	9447 (78.4)
Confusion	1046 (8.7)	11,006 (91.3)
Productive coughs	608 (5.0)	11,444 (95.0)

High-Risk Settings and Predisposing Factors of Severe COVID-19

Most participants recognized that COVID-19 could be transmitted via hospitals, religious facilities, sports complexes, and public transportation. The majority considered respiratory diseases, age >60 years, and diabetes mellitus as predisposing factors for COVID-19 complications (Table 5). However, almost half (n=5253, 44.1%) did not recognize hypertension as a predisposing factor, and 411 (3.3%) reported that they did not know what the predisposing factors were.

Table 5. Participants' knowledge of predisposing factors of severe COVID-19 (n=11,921).

Factors	Response	
	Yes, n (%)	No, n (%)
Respiratory disease	10,843 (91.0)	1078 (9.0)
Age >60 years	10,674 (89.5)	1247 (10.5)
Diabetes mellitus	9659 (81.0)	2262 (19.0)
Cardiovascular disease	9277 (77.8)	2644 (22.2)
Chemotherapy	9193 (77.1)	2728 (22.9)
Extensive corticosteroids use	7993 (67.0)	3928 (33.0)
Organ transplant	7896 (66.2)	4025 (33.8)
Hypertension	6668 (55.9)	5253 (44.1)
Severe obesity	3022 (25.4)	8899 (74.6)

XSL•FO
RenderX

Estimation of Pathogenicity, Hospitalization Rate, Case Fatality, and Reproduction Number

Participants were asked about their perception of the percentage of people exposed to the virus who may contract, be hospitalized, or pass away due to COVID-19 infection. The median estimates were 60 (IQR 30-80), 25 (IQR 17-50), and 6 (IQR 3-20), respectively. In addition, participants reported the median number of people that an individual with COVID-19 could infect to be 10 (IQR 5-30).

Self-Evaluation on the Level of Knowledge About COVID-19

Participants self-reported varying levels of knowledge: sufficient (n=2542, 20.6%), fair (n=8336, 67.6%), and insufficient (n=1454, 11.8%). Subsequently, when asked whether they felt the need for further training, 8892 (72.1%) responded in the affirmative. As indicated in Table 6, Telegram and audiovisual media were the leading sources of information about COVID-19 in Iran.

Table 6. Participants' sources of information about COVID-19.

Source of information	Participants, n (%)
Telegram	8740 (70.9)
Audiovisual media	8631 (70.0)
Internet search	7657 (62.1)
Instagram	5762 (46.7)
News agencies	4367 (35.4)
Family and friends	4184 (33.9)
Health care professionals	4199 (34.0)
Scientific articles	3296 (26.7)
Work	1381 (11.2)
Print media	1048 (8.5)
Pamphlets and posters	1045 (8.5)
Twitter	870 (7.1)

Attitudes

When asked about participants' reaction to receiving a positive diagnosis for COVID-19, 11,452 (92.9%) reported self-isolation and resting at home, 10,671 (86.5%) mentioned seeing a doctor in case symptoms worsen, and 332 (2.7%) reported continuing with daily life.

In response to a question about the likelihood of COVID-19 being manufactured for bioterrorism purposes, 2971 (24.1%) reported a very high likelihood. The remaining participants answered as follows: high (n=2928, 23.7%), fair (n=3225, 26.2%), low (n=1178, 9.5%), and very low (n=2030, 16.5%). When asked about the dangers of the current situation, responses were as follows: very high (n=5817, 47.2%), high (n=5146, 41.7%), fair (n=1238, 10.1%), low (n=103, 0.8%), and very low (n=28, 0.2%). When asked whether they thought that countries would eventually defeat the disease, 11,212 (90.9%) responded in the affirmative, while 1120 (9.1%) said no.

In response to a question asking how long it would take to control the pandemic, 131 (1.1%) reported <1 month, 2762 (22.4%) reported 1-3 months, 3917 (31.8%) reported 3-6 months, 1496 (12.1%) reported 6-9 months, 1548 (12.6%) reported 9-12 months, 2141 (17.4) reported >1 year, and 337 (2.7%) had no idea.

Most respondents had positive attitudes toward COVID-19. They were on board with the government in terms of maintaining a safe physical distance, city lockdowns, and self-isolation upon symptom onset. However, participants were mostly unsure when asked how comfortable they would be when encountering someone who has recently recovered from COVID-19 (Table 7).

XSL•FO
RenderX

Table 7. Attitudes of participants toward COVID-19.

Statement	Respondents, n	Response	
		"I agree," n (%)	"I disagree," n (%)
Maintaining a safe physical distance is a duty for all citizens.	12,264	12,219 (99.6)	45 (0.4)
Only people at risk of the severe form of COVID-19 need to take preventive measures.	12,184	459 (3.8)	11,725 (96.2)
As an employee, I would notify my employer upon the emergence of COVID-19 symptoms.	12,118	12,070 (99.6)	48 (0.4)
I would visit a close friend or family member with COVID-19.	12,061	75 (0.6)	11,986 (99.4)
Every household member should use their towel or facial tissue to dry their hands and face.	12,056	11,933 (99.0)	123 (1.0)
As an employer, I would agree with my employees' sick leave if they have COVID-19 symptoms.	12,019	11,959 (99.5)	60 (0.5)
Closure of businesses, schools, and universities is an opportunity to visit family and friends.	11,958	706 (5.9)	11,252 (94.1)
Relatives of a person deceased from COVID-19 should not feel ashamed.	11,902	11,673 (98.1)	229 (1.9)
All affected cities need to go under immediate lockdown.	11,573	11,215 (97.2)	322 (2.8)
Authorities need to prioritize pressing issues like road accidents or noncommunicable diseases.	11,184	442 (4.0)	10,742 (96.0)
Pet owners need to take special care of their pets' contact with people.	11,099	10,721 (96.6)	378 (3.4)
People with a suspected travel history or close contact should report themselves to health authorities.	11,001	10,615 (96.5)	386 (3.5)
I would rather shop online instead of shopping in person.	10,339	9809 (94.9)	530 (5.1)
I would notify health authorities if my employer insisted on the illegal reopening of business.	10,140	9830 (96.9)	310 (3.1)
In a resource-scarce setting, I would make a face mask at home to cover my nose and mouth.	9705	8278 (85.3)	1427 (14.7)
Despite keeping a safe distance, I feel uncomfortable around people who have recovered from COVID-19.	9691	5589 (57.7)	4102 (42.3)
This pandemic is divine retribution due to the sins of humankind.	8699	1553 (17.9)	7146 (82.1)
The pandemic will resolve on its own when the weather gets warmer.	7527	919 (7.5)	6608 (87.8)

Practices

General Practices Related to COVID-19

The mean score for general practices related to COVID-19 was 20.7 (SD 2.2, range 2-24) out of 24 (86.2%) (Table 8).

No significant differences were observed between practice scores of participants and their sex, age, province of residence, marital status, number of living children, smoking status, whether they knew someone with COVID-19, whether they had a household member with COVID-19, whether they caught the disease themselves, or whether a family member had passed away due to COVID-19 (Table 9).

XSL•FO
RenderX

Table 8. General statements regarding participants' practices pertaining to COVID-19.

Statement	Response	
	Yes, n (%)	No, n (%)
I will not go on a trip during Nowruz.	11,945 (96.9)	387 (3.1)
I will isolate myself at home if I have a symptom of COVID-19.	11,878 (96.3)	454 (3.7)
I cover my mouth and nose when sneezing or coughing.	11,694 (94.8)	638 (5.2)
I disinfect the packaging of the goods I buy from stores.	11,626 (94.3)	706 (5.7)
I did not go on a trip during Nowruz.	11,468 (93.0)	864 (7.0)
I always wash my hands or use hand sanitizers before touching my face.	11,342 (92.0)	990 (8.0)
As requested by authorities, I stayed at home and isolated myself immediately.	11,208 (90.9)	1124 (9.1)
I regularly disinfect my cell phone according to guidelines.	10,412 (84.4)	1920 (15.6)
I carry hand sanitizers or water-soap solutions when outside the home.	10,357 (84.1)	1957 (15.9)
I always use a face mask if I must leave the house for an emergency.	9519 (77.2)	2813 (22.8)
I heat bread before use to kill potential pathogens.	9512 (77.1)	2820 (22.9)
I regularly disinfect highly touched surfaces at home.	9499 (77.0)	2833 (23.0)
I bought some face masks for personal protection.	8638 (70.0)	3694 (30.0)
I keep my cell phone in my pocket to reduce the chance of infection.	8251 (66.9)	4081 (33.1)
I made homemade face masks for me/my family.	2392 (19.4)	9940 (80.6)
I bought some unprescribed drugs to prevent the disease.	2065 (16.7)	10,267 (83.3)
I let my children visit their grandparents.	1573 (12.8)	10,759 (87.2)
I went to the bazaar to shop on days before Nowruz.	1176 (9.5)	11,156 (90.5)
I drank alcoholic beverages for disinfection.	301 (2.4)	12,031 (97.6)
I attended a relative's funeral.	152 (1.2)	12,180 (98.8)
I visited a cemetary on the last Thursday before Nowruz.	138 (1.1)	12,194 (98.9)
As always, I visited my family and friends during Nowruz.	128 (1.0)	12,204 (99.0)
I let my children play with other children outside the house.	107 (0.9)	12,225 (99.1)
As always, I will visit my family and friends during Nowruz.	84 (0.7)	12,248 (99.3)

XSL•FO
RenderX

Table 9. Determinants of the COVID-19 general practices score among participants.

Variable	Score (SD)	95% CI
Sex		
Female	20.7 (1.9)	20.5-20.8
Male	20.4 (2.1)	20.3-20.6
Age		
<40 years	20.9 (1.9)	20.8-21.1
≥40 years	20.6 (2.2)	20.4-20.9
Marital status		
Married	20.8 (2.1)	20.8-20.9
Never married	20.5 (2.2)	20.4-20.5
Divorced	20.8 (2.2)	20.4-20.9
Engaged	20.3 (2.5)	20.1-20.6
Widowed	20.8 (2.2)	20.4-20.9
High school diploma		
Yes	20.6 (2.1)	20.4-20.6
No	20.2 (2.8)	20.1-20.4
Number of children		
0	20.5 (2.2)	20.4-20.7
1	20.8 (2.1)	20.6-20.9
2	20.9 (2.1)	20.8-20.9
≥3	20.8 (2.2)	20.7-20.9
Smoking		
Yes	20.4 (2.5)	20.3-20.7
No	20.7 (2.1)	20.5-20.8
Former smoker	20.6 (2.1)	20.4-20.8
Knowing someone with COVID-19		
Yes	20.7 (2.1)	20.6-20.8
No	20.7 (2.2)	20.6-20.7
Household member with COVID-19		
Yes	20.7 (2.1)	20.5-20.9
No	20.7 (2.2)	20.6-20.7
Has or has had COVID-19		
Yes	20.3 (2.9)	19.9-20.7
No	20.7 (2.1)	20.7-20.8
Not sure	20.5 (2.4)	20.3-20.7
Deceased family member due to COVID-19		
Yes	20.7 (2.2)	20.6-20.8
No	20.7 (2.2)	20.6-20.7

Self-isolation, Care Seeking, and Preventive Measures

In total, 200 (1.6%) participants reported they became COVID-19 positive, 10,453 (84.8%) stated they did not get the disease, and 1679 (13.6%) reported that they had experienced some symptoms but took no measures. Among those who indicated an infection, 174 (87.0%) reported self-isolation at home, 99 (49.5%) reported visiting a doctor when symptoms became worse, and 19 (9.5%) reported carrying on with daily life.

XSL•FO
RenderX

When asked about what they did to prevent COVID-19, 11,760 (95.4%) participants reported staying at home, 11,687 (94.8%) practiced hand hygiene, 9691 (78.6%) maintained a safe physical distance, 9391 (76.2%) wore gloves, 8907 (72.2%) wore a face mask, 5458 (44.3%) took vitamin supplements, 3405 (27.6%) consumed herbal teas, 252 (2.0%) took the Imam Kazem drug, 57 (0.5%) performed wet cupping, and 48 (0.4%) took violet oil.

Among 6156 respondents, 2914 (47.3%) reported wearing surgical masks, 1989 (32.3%) cloth masks, 1566 (25.4%) N95 masks, and 772 (12.5%) homemade masks. In total, 728 (11.8%) respondents reported not wearing a mask. The mean duration of wearing each mask was 5.1 (SD 7.9) hours (median 3.0, IQR 2-5).

The mean number of times respondents performed daily handwashing with water and soap was 13.1 (SD 12.6; median 10, IQR 6-15). The mean duration of handwashing was 22.5 (SD 14.7) seconds (median 20, IQR 20-25). The mean number of times of daily rubbing hands with alcohol-based hand sanitizers among participants was 4.9 (SD 9.8; median 2, IQR 1-5). The mean duration of hand rubbing was 9.9 (SD 10.1) seconds (median 10, IQR 3-15).

Leaving One's Home or Town

The mean number of times participants left their homes during the past week was 2.7 (SD 4.1; median 2, IQR 1-3). When questioned about the reason, 7303 (59.2%) reported going grocery shopping, 2103 (17.1%) went to work, 834 (6.8%) shopped for items other than groceries, 755 (6.1%) visited family and friends, 521 (4.2%) went sightseeing, 511 (4.1%) went to the hospital, and 460 (3.7%) left to take care of the elderly. As many as 2239 (18.2%) reported that they did not leave their home. Only 7.0% (n=327) of male participants stated they had not left town during the pandemic, which differed significantly from females (n=1912, 25.1%; $P<.001$).

In total, 9074 (73.6%) participants reported that they did not leave town since the official declaration of the outbreak. Among those who did leave town, the mean number of times of going outside the city was 6.5 (SD 14.7; median 2, IQR 1-4). When asked about their reason for leaving town, 1955 (60.0%) reported going to work, 906 (27.8%) visited family and friends, 338 (10.4%) went sightseeing, 24 (0.7%) visited holy shrines, and 22 (0.7%) reported other reasons.

Discussion

Knowledge

Participants' COVID-19 knowledge scores were adequate. More than 90% recognized the cause of the disease, its routes of transmission, and measures for prevention. Since the beginning of the COVID-19 pandemic, the media has focused on three symptoms—fever, dyspnea, and cough. Hence, more than 90% recognized the three most common symptoms, similar to people in the United Kingdom and the United States [6]. However, infrequently discussed symptoms like gastrointestinal symptoms were less known to the public. A large majority (98%) knew that worsening dyspnea could be a sign for the severe form of the disease. However, 70% considered fever alarming. Although more than 80% considered respiratory diseases, age >60 years, and diabetes as predisposing factors, almost half did not recognize hypertension, despite imposing significant risk [7], and only 25% knew severe obesity to be a prognostic factor. Nearly all adults in the United Kingdom and the United States recognized that adults with underlying diseases were more likely to experience a fatal disease course [6]. Participants' median estimate for the case fatality rate of COVID-19 was 6%, which was similar to results from the United Kingdom and the United States [6]. The observed case fatality rate is 4.9% in Iran, 5.8% in the United States, 14.1% in the United Kingdom, and 5.9% worldwide [1].

Regarding misbeliefs about the disease, about 15% thought that eating garlic would kill the pathogen. In comparison, 33% of Egyptians thought eating garlic was a proper preventive measure [8]. A study analyzing the trends in social media use among Iranians reported that the main misconceptions were about statistics, prevention and protection methods, treatment, dietary recommendations, and disease transmission [9]. Misbeliefs were not confined to people in Iran; 44% and 36% of the participants in the United States and the United Kingdom believed that using a hand dryer, rinsing the nose with saline, taking antibiotics, or gargling with mouthwash were effective prevention measures for COVID-19 [6]. In India, false claims about the use of herbal medicine, as well as religious and spiritual methods for prevention and treatment, became problematic and added to confusion [10].

The role of social media in delivering health information during the pandemic is significant. Telegram, the most popular social media platform in Iran, was the leading source of information among participants, followed by television and radio. When looking for a platform to deliver rapid health information to the public, social media appears to be promising [11]. There were no significant differences among participants' knowledge score and their sex, ethnicity, or province of residence. This could suggest that the information origin of popular media sources was the same, led by national audiovisual media sources. However, there are studies that express the concern that the freedom of expression in social media could result in the dissemination of misbeliefs, misinformation, and deception [10]; this can be corrected with public health information campaigns that target these misconceptions [12]. The active and effective presence of health professionals and authorities on the internet and social media is essential, especially in the early days of a crisis when quality information might be limited on the internet [13].

Attitudes

Participants' attitudes toward COVID-19 were mainly positive, and the majority believed that the COVID-19 pandemic would eventually be controlled, which is similar to the results of a study from Wuhan, China [5]. Most participants recognized the dangers of the current situation. They agreed that all people need to maintain a safe physical distance and stay at home and that all affected cities should implement immediate lockdown. While 86% of people in Egypt considered COVID-19 to be dangerous [8], only 20% of participants in a study in Pakistan believed that the disease is dangerous [14]. In our study,

XSL•FO
RenderX

participants felt inclined to wear face masks in public, as 85% reported that they would even make a face mask at home if they could not buy one. More than 90% of participants said they would self-isolate if they experienced symptoms. In comparison, 64% of participants in the United States reported they follow self-isolation protocols [6].

Nearly 50% of participants considered it highly likely that the COVID-19 is caused by a human-made pathogen for bioterrorism purposes, which is larger than the 27% of Egyptians who believe the virus was initially designed as a biological weapon [8].

Practices

The mean practice score of participants was adequate. Most participants had left their home 1 to 3 times during the past week, mostly for grocery shopping. Almost 75% of participants reported that they had not left town since the beginning of the outbreak.

The early days of social distancing implementation coincided with the Persian New Year holiday, Nowruz, which lasts for at least 2 weeks. People displayed satisfactory cooperation with the government in not leaving their homes. More than 90% reported self-isolation at home immediately after being requested to do so by the government. Iranians perform holiday customs before and during Nowruz. Before Nowruz, they shop for new clothes, shoes, sweets, and nuts, and pay their respects to deceased beloved ones in cemeteries on the last Thursday before Nowruz. They also visit family and friends and take trips during this holiday. More than 90% reported that they did not go shopping on the days before Nowruz, and 99% said they did not visit a cemetery. Moreover, almost all participants stated that they did not visit their family and friends, and 97% reported that they did not go on a trip this Nowruz.

Almost all participants reported that they would not let their children play with other children outside their home. Hence, the social and mental needs of children need attention during lockdowns [15].

People also followed hygiene protocols correctly. The majority of participants covered their faces when sneezing or coughing and washed their hands before touching their faces. In comparison, only 38% of participants in a study in Pakistan avoided touching their face with contaminated hands [14]. Most participants (77%) used face masks and carried hand sanitizers when outside, which was lower than a study from Wuhan, China, where 98% of participants used face masks [5]. It is worth mentioning that although supply shortages in personal protective equipment in Iran was resolved after the early days of the pandemic, it was severely compromised at the beginning of the outbreak, which may explain the lower adherence to face mask use seen in our study. About 20% reported the use of homemade face masks, which could ease the supply demand for surgical or N95 masks that are needed at health care facilities.

While 95% of participants reported staying home and practicing hand hygiene to prevent COVID-19, less than 2% were involved in some sorts of deception. The most challenging to control, trending practices included the oral use of an herbal combination called the Imam Kazem drug and a method that involved inserting a ball of cotton wool drenched in violet oil into the anus. The use of all unapproved medications is under question due to possible serious consequences [16].

About 2% of participants reported drinking alcoholic beverages for disinfection purposes, which needs particular attention, as a new methanol poisoning outbreak is being witnessed in Iran [17]. We also noted that participation was higher among women. Gender-based differences in participation have been reported in other studies as well [18,19].

Strengths and Limitations

This is the first nationwide study to assess the knowledge level, attitudes, and practices of people in Iran in the context of COVID-19. The strengths of this study lie in its large sample size and time of data gathering (ie, in the early days of the outbreak in Iran). Findings could empower public health authorities to precisely direct relevant information campaigns to eliminate misconceptions.

There are some limitations that must be acknowledged. Compared to the most recent national population statistics of Iran, our sample was overrepresentative of women and younger people. Since there was no representative online platform to connect researchers to people in Iran, we decided to conduct this study using the Google Forms platform. Despite the high penetration rate of the internet in Iran, some groups, especially the elderly and vulnerable groups in rural areas, might not have access to the platform. Participants were asked to read the questionnaire to their parents and grandparents to ensure higher participation of those groups; however, such voluntary measures are not guaranteed.

Conclusions

Knowledge pertaining to COVID-19 was almost adequate among Iranians, their attitudes were mainly positive, and their practices were satisfactory. There is still room for improvement in correcting misinformation and protecting people from deception, especially in terms of consuming alcohol. Iranians appear to agree with government decisions, like the implementation of social distancing measures, and care for their safety and that of others. The government and health care authorities need to continue to maintain public trust.

Acknowledgments

The authors would like to thank all the people who participated in the survey. We appreciate the invaluable contributions of medical students Mohammad-Reza Malekpour, Javad Khanali, Ali Akhoondi, and Mohammad-Mahdi Salar-Abedi. This study was supported by the Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences (#22975).

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. The Lancet Public Health. COVID-19 puts societies to the test. The Lancet Public Health 2020 May;5(5):e235. [doi: 10.1016/s2468-2667(20)30097-9]
3. Coronavirus disease (COVID-19) advice for the public: Mythbusters. World Health Organization. 2020 Nov 23. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-07-21]
4. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. The Lancet Public Health 2020 May;5(5):e249-e250. [doi: 10.1016/s2468-2667(20)30086-4]
5. 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]
6. 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]
7. 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. International Journal of Infectious Diseases 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
8. 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]
9. Bastani P, Bahrami MA. COVID-19 Related Misinformation on Social Media: A Qualitative Study from Iran. J Med Internet Res 2020 Apr 05 [FREE Full text] [doi: 10.2196/18932] [Medline: 32250961]
10. Kadam AB, Atre SR. Negative impact of social media panic during the COVID-19 outbreak in India. J Travel Med 2020 May 18;27(3) [FREE Full text] [doi: 10.1093/jtm/taaa057] [Medline: 32307545]
11. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec 30;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
12. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
13. 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]
14. Salman M, Mustafa ZU, Asif N, Zaidi HA, Hussain K, Shehzadi N, et al. Knowledge, attitude and preventive practices related to COVID-19: a cross-sectional study in two Pakistani university populations. Drugs Ther Perspect 2020 May 09:1-7 [FREE Full text] [doi: 10.1007/s40267-020-00737-7] [Medline: 32395069]
15. Golberstein E, Wen H, Miller BF. Coronavirus Disease 2019 (COVID-19) and Mental Health for Children and Adolescents. JAMA Pediatr 2020 Sep 01;174(9):819-820. [doi: 10.1001/jamapediatrics.2020.1456] [Medline: 32286618]
16. Yang Y. Use of herbal drugs to treat COVID-19 should be with caution. The Lancet 2020 May 30;395(10238):1689-1690 [FREE Full text] [doi: 10.1016/S0140-6736(20)31143-0] [Medline: 32422123]
17. Delirrad M, Mohammadi AB. New Methanol Poisoning Outbreaks in Iran Following COVID-19 Pandemic. Alcohol Alcohol 2020 Jun 25;55(4):347-348 [FREE Full text] [doi: 10.1093/alcalc/agaa036] [Medline: 32400874]
18. Aerny-Perreten N, Domínguez-Berjón MF, Esteban-Vasallo MD, García-Riolobos C. Participation and factors associated with late or non-response to an online survey in primary care. J Eval Clin Pract 2015 Aug;21(4):688-693. [doi: 10.1111/jep.12367] [Medline: 25929295]
19. Harrison JM, Jung M, Lennie TA, Moser DK, Smith DG, Dunbar SB, et al. Refusal to participate in heart failure studies: do age and gender matter? J Clin Nurs 2016 Apr;25(7-8):983-991 [FREE Full text] [doi: 10.1111/jocn.13135] [Medline: 26914834]

Abbreviations

ANOVA: analysis of variance

XSL•FO
RenderX

Edited by G Eysenbach; submitted 14.06.20; peer-reviewed by M Kwasny, H Ashrafi-rizi; comments to author 13.07.20; revised version received 28.07.20; accepted 07.12.20; published 23.02.21.

Please cite as:
Abbasi-Kangevari M, Kolahi AA, Ghamari SH, Hassanian-Moghaddam H
Public Knowledge, Attitudes, and Practices Related to COVID-19 in Iran: Questionnaire Study
JMIR Public Health Surveill 2021;7(2):e21415
URL: https://publichealth.jmir.org/2021/2/e21415
doi:10.2196/21415
PMID:33620326

©Mohsen Abbasi-Kangevari, Ali-Asghar Kolahi, Seyyed-Hadi Ghamari, Hossein Hassanian-Moghaddam. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 23.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Distance Management of Spinal Disorders During the COVID-19 Pandemic and Beyond: Evidence-Based Patient and Clinician Guides From the Global Spine Care Initiative

Scott Haldeman[1,2,3], MD, PhD, DC; Margareta Nordin[4,5], DrMedSci, PT; Patricia Tavares[6], BSc, DC; Rajani Mullerpatan[7], PT, MSc, PhD; Deborah Kopansky-Giles[8,9], DC, MSc; Vincent Setlhare[10], MD, MBA, MFamMed; Roger Chou[11,12], MD; Eric Hurwitz[13], DC, PhD; Caroline Treanor[14,15], PT, MSc; Jan Hartvigsen[16,17], DC, PhD; Michael Schneider[18,19], DC, PhD; Ralph Gay[20], MD, DC; Jean Moss[21], DC, MBA; Joan Haldeman[1], PT; David Gryfe[21], BSc, DC; Adam Wilkey[22], DC, MSc; Richard Brown[23], DC, LLM; Geoff Outerbridge[24], MSc, DC; Stefan Eberspaecher[24], DC; Linda Carroll[25], PhD; Reginald Engelbrecht[26], DC; Kait Graham[27], BHSc, DC; Nathan Cashion[28], MS, DC; Stefanie Ince[29]; Erin Moon[30]

[1]World Spine Care, Santa Ana, CA, United States
[2]Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa, ON, Canada
[3]Department of Neurology, University of California, Irvine, CA, United States
[4]Department of Orthopedic Surgery, New York University, New York, NY, United States
[5]Department of Environmental Medicine, New York University, New York, NY, United States
[6]Department of Clinical Education, Canadian Memorial Chiropractic College, Toronto, ON, Canada
[7]MGM School of Physiotherapy, MGM Institute of Health Sciences, Navi, Mumbai, India
[8]Department of Research, Canadian Memorial Chiropractic College, Toronto, ON, Canada
[9]Department of Family & Community Medicine, University of Toronto, Toronto, ON, Canada
[10]Department of Family Medicine and Public Health Medicine, University of Botswana, Gaborone, Botswana
[11]Department of Medicine, Oregon Health & Science University, Portland, OR, United States
[12]Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR, United States
[13]Office of Public Health Studies, University of Hawaii, Manoa, HI, United States
[14]Department of Physiotherapy, Beaumont Hospital, Dublin, Ireland
[15]Department of Neurosurgery, National Neurosurgical Spinal Service, Beaumont Hospital, Dublin, Ireland
[16]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
[17]Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Norway
[18]School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, United States
[19]Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, United States
[20]Department of Physical Medicine and Rehabilitation, Mayo Clinic, Alix School of Medicine, Rochester, MN, United States
[21]Canadian Memorial Chiropractic College, Toronto, ON, Canada
[22]World Spine Care Europe, Holmfirth, United Kingdom
[23]World Federation of Chiropractic, Toronto, ON, Canada
[24]World Spine Care, Ottawa, ON, Canada
[25]School of Public Health, University of Alberta, Edmonton, AB, Canada
[26]World Spine Care, South Africa, Bethlehem, South Africa
[27]Volunteer Programs and Operations, World Spine Care, Ottawa, ON, Canada
[28]Digital Communications, World Spine Care, Oregon City, OR, United States
[29]World Spine Care, Toronto, ON, Canada
[30]World Spine Care Yoga Project, Vancouver, BC, Canada

Corresponding Author:
Scott Haldeman, MD, PhD, DC
World Spine Care
10131 Brier Lane
Santa Ana, CA, 92705
United States

XSL•FO
RenderX

Phone: 1 714 651 2949
Email: shaldema@uci.edu

Abstract

Background: The COVID-19 pandemic has greatly limited patients' access to care for spine-related symptoms and disorders. However, physical distancing between clinicians and patients with spine-related symptoms is not solely limited to restrictions imposed by pandemic-related lockdowns. In most low- and middle-income countries, as well as many underserved marginalized communities in high-income countries, there is little to no access to clinicians trained in evidence-based care for people experiencing spinal pain.

Objective: The aim of this study is to describe the development and present the components of evidence-based patient and clinician guides for the management of spinal disorders where in-person care is not available.

Methods: Ultimately, two sets of guides were developed (one for patients and one for clinicians) by extracting information from the published Global Spine Care Initiative (GSCI) papers. An international, interprofessional team of 29 participants from 10 countries on 4 continents participated. The team included practitioners in family medicine, neurology, physiatry, rheumatology, psychology, chiropractic, physical therapy, and yoga, as well as epidemiologists, research methodologists, and laypeople. The participants were invited to review, edit, and comment on the guides in an open iterative consensus process.

Results: The Patient Guide is a simple 2-step process. The first step describes the nature of the symptoms or concerns. The second step provides information that a patient can use when considering self-care, determining whether to contact a clinician, or considering seeking emergency care. The Clinician Guide is a 5-step process: (1) Obtain and document patient demographics, location of primary clinical symptoms, and psychosocial information. (2) Review the symptoms noted in the patient guide. (3) Determine the GSCI classification of the patient's spine-related complaints. (4) Ask additional questions to determine the GSCI subclassification of the symptom pattern. (5) Consider appropriate treatment interventions.

Conclusions: The Patient and Clinician Guides are designed to be sufficiently clear to be useful to all patients and clinicians, irrespective of their location, education, professional qualifications, and experience. However, they are comprehensive enough to provide guidance on the management of all spine-related symptoms or disorders, including triage for serious and specific diseases. They are consistent with widely accepted evidence-based clinical practice guidelines. They also allow for adequate documentation and medical record keeping. These guides should be of value during periods of government-mandated physical or social distancing due to infectious diseases, such as during the COVID-19 pandemic. They should also be of value in underserved communities in high-, middle-, and low-income countries where there is a dearth of accessible trained spine care clinicians. These guides have the potential to reduce the overutilization of unnecessary and expensive interventions while empowering patients to self-manage uncomplicated spinal pain with the assistance of their clinician, either through direct in-person consultation or via telehealth communication.

(JMIR Public Health Surveill 2021;7(2):e25484) doi:10.2196/25484

KEYWORDS

COVID-19; spinal disorders; evidence-based care; physical distancing care; clinical guides; low- and middle-income communities; telehealth; telemedicine; evidence-based; spine; guide; management

Introduction

This paper has been developed primarily in response to the impact of the COVID-19 pandemic caused by the SARS-CoV-2 virus on the delivery of health care for people with spine-related symptoms. Although the SARS-CoV-2 virus was first identified in late 2019, it was not until January and February 2020 that the seriousness of the situation became evident to health authorities globally [1,2]. By March 2020, the world began to recognize the severity of COVID-19 and a lockdown of health care services was considered for all but emergency care. Internationally, public health professionals and government have either mandated or recommended staying at home and physical or social distancing as key interventions necessary to reduce viral transmission. These restrictions have had a dramatic impact on clinicians' ability to deliver in-person spine care, which traditionally consists of close-contact clinical examination and physical interventions [3,4].

In March 2020, the World Spine Care (WSC) charities [5,6] were forced to close their clinics in all countries where they were operational. This was, in each case, an edict by the health care authorities in the respective countries. Volunteer clinicians had to return home to avoid being stranded in the countries where they were providing services.

During this period, however, people with spine-related symptoms, especially those disabled by these disorders, did not suddenly cease to require care that was simply not immediately available. A series of articles offering advice to clinicians within several specialties that commonly manage people with chronic pain and musculoskeletal and spinal disorders began appearing [7-10]. However, by March and April 2020, no organization had provided guidance that both patients and clinicians with different levels of training and education could follow to triage

XSL•FO
RenderX

spine-related symptoms and disability and provide advice on what the current evidence would suggest as a reasonable management strategy for all potential symptom presentations.

In 2018, the Global Spine Care Initiative (GSCI), an initiative of WSC, published 15 articles in a special issue of the European Spine Journal, summarized in an executive summary [11]. The goal of the GSCI was to bring together leading health care providers, scientists, specialists, government agencies, and other stakeholders to review all current guidelines and to develop a care pathway and model of care that was consistent with current evidence and could be used by all clinicians, irrespective of their training and experience. The final articles were the result of collaboration between 68 participants from 24 countries on 5 continents [12]. The publication of the GSCI provided a unique evidence-based foundation to support the development of guides that could be used to support the delivery of spine care during the COVID-19 pandemic.

Physical distancing between clinicians and patients with spine-related symptoms is not limited to restrictions imposed by pandemic-related lockdowns. For people experiencing these symptoms in most low- and middle-income countries, as well as many underserved communities in high-income countries, extraordinarily little if any evidence-based spine care is available and practiced [13].

The objective of this paper is to describe the process of developing and presenting evidence-based patient and clinician guides that can be used for triage, advice, education via telecommunication, and referral for appropriate care when necessary when direct in-person care is not available.

Methods

The proposed patient and clinician guides were based on 4 principles:

1. Useful during situations where spine care is delivered remotely.
2. Consistent with the recommendations of the most widely accepted evidence-based approach to spinal disorders.
3. Useful for all clinicians who manage spinal disorders, irrespective of their training. This includes family physicians, chiropractors, physical therapists, and medical specialists in such fields as neurology, rheumatology, physiatry, and others.
4. Clear and easily understandable so that patients with different levels of education can reasonably use it, if necessary, with the help of a clinician via telecommunication.

In the end, two sets of guides, one for patients and one for clinicians, were developed by extracting information from the published GSCI papers [14-19]. These papers had reviewed all current, widely accepted guidelines for the management of spinal disorders. The steps recommended in these guides were extracted primarily from the articles on the GSCI Classification of Spinal Disorders [16] and the GSCI Care Pathway [14]. Special attention was paid to the flash cards in the GSCI Care Pathway article. These flash cards provided evidence-based intervention recommendations for each class and subclass of spinal disorder in the GSCI system.

The stimulus for the development of these guides came after a presentation to the Skoll World Forum on April 1, 2020, by World Spine Care, titled “Pain and disability: Making the vulnerable more vulnerable.” This presentation has been viewed by approximately 3000 people. Questions and discussion after this event made it clear that there was a need for information and guidance for the management of spinal disorders during the pandemic. The process of initial extraction and review was completed by two of the lead investigators of the GSCI, a layperson without health care qualifications, and an emeritus dean of an educational institution that graduates clinicians who treat patients with spine-related symptoms.

To ensure a wide consensus and an evidence-based approach that was not limited to members of the GSCI, it was decided that an international and interdisciplinary team would review, provide input on, edit, and modify the guides as necessary. Therefore, 14 individuals who had been coauthors of the GSCI articles and 15 individuals who did not participate in the GSCI were invited to review drafts of the guides. The group included 8 participants who had extensive experience in developing and reviewing evidence-based guides on the management of low back or neck pain; 10 participants who currently held, or had held, senior teaching or administration positions for clinicians who manage spinal disorders in the fields of family practice, chiropractic, and physical therapy; and 15 participants who had extensive clinical experience in family medicine, neurology, physiatry, rheumatology, psychology, chiropractic, physical therapy, and yoga. The participants were from 10 countries on 4 continents.

An open, progressive, iterative approach was used to ensure that every participant would be able to have input and review opinions and recommendations from all other participants. When a participant submitted edits or recommendations, they were incorporated into the guides, with all other participants invited to review and comment on the revisions. A total of 10 versions of the guides were prepared using this process. When there were no further comments or edits from any of the participants, the final draft was converted into a formal educational document for distribution in a professionally produced education format. The process was completed within a 4-week time frame and the final document was posted on the World Spine Care website on April 26, 2020. It has subsequently been posted on the websites of the Society of Indian Physiotherapists and the World Federation of Chiropractic and downloaded over 500 times.

Results

The Patient Guide

It was agreed that the patient guide should be limited to a simple 2-step process. The first step requires the patient to describe the nature of their symptoms or concerns. The second step provides the following:

- Information that a patient could use to determine the level of care required for their symptoms.

XSL•FO
RenderX

- Reassurance and recommendations where the evidence suggests that self-care is likely to be of benefit.
- Advice about when it would be important to seek the care of a spine care clinician or to seek emergency care.

It was stressed that not all spine-related symptoms require immediate (clinical) treatment, imaging, or expensive interventions. The following statement was included in the Patient Guide: "You are empowered to self-assess. Your spine care provider is your partner in health, do not hesitate to reach out to them by phone or email if you have questions."

The first step in the patient guide includes a series of five questions for patients to self-complete that provide adequate information on pain location, severity, and activity impact, and major red flags for serious pathology. Based on feedback from laypeople involved in creating the guides, it was felt that the questions were clear enough that an average patient, even with limited education, could provide answers. Textbox 1 includes the questions that are felt to be essential for this purpose. It was also realized that patients, especially in situations where reading capacity or comprehension might be limited, may require assistance from a clinician, family member, or community helper to answer specific questions. This clarification is included within Textbox 1. The process of using the guides starts with the following statement: "I have back or neck pain or symptoms such as pain, numbness or weakness in my arms or legs or headaches that might be coming from my back or neck. How do I determine the severity of these symptoms?"

XSL•FO
RenderX

Textbox 1. Step 1 of Patient Guide.

Answer the following five questions to determine whether it is necessary to seek professional help. This can be done personally, or with the help of a licensed clinician or first responder via a telehealth consultation by phone or videoconferencing.

1. What are my problems/symptoms?

A. No or minimal discomfort

B. Mild pain

C. Moderate pain

D. Severe pain

E. Numbness or tingling

F. Muscle weakness

G. Loss of balance

H. Recent onset of bladder or bowel problems like loss of control

2. Do I feel pain beyond my spine?

A. No

B. Down both legs

C. Down one leg

D. Down both arms

E. Down one arm

F. New or different headaches

G. Chest pain

3. Are my symptoms stopping me from doing my normal activities?

A. No. I can do everything

B. Yes, a little. I can do most activities

C. Yes, a lot. I have difficulty doing anything

4. Have I had a recent fall or accident?

A. No

B. Yes

5. Do I have, or have I had, any other serious diseases?

A. No

B. Cancer

C. Infection such as tuberculosis or HIV/AIDS

D. Osteoporosis or steroid use

E. Any condition that causes inflammation of my joints or muscles

F. Any condition that affects my nerves or brain

The second step in the patient guide is a listing of concerns for which a patient might seek answers. This step includes information that would apply to specific responses from the questionnaire in step 1. The level and severity of the complaints is divided into subgroups consistent with the GSCI Classification of Spinal Disorders [19]: minimal or no discomfort, mild spine pain, moderate spine pain, severe spine pain, symptoms consistent with nerve involvement, possible spine or bone fracture, and possible complications of serious systemic diseases that might cause spine-related symptoms (Textbox 2).

The first question likely to be asked by a patient is whether they need to see a clinician in the office. The answers, as noted in Textbox 2, vary from “not necessary” through “not usually necessary,” “may need,” “may be necessary,” “emergency care is needed,” and “seek care from a medical specialist.” For patients with spine pain and no red flags, self-care recommendations are made that are consistent with current evidence-based guidelines, with emphasis on reassurance and education. This differentiation of symptoms into patterns also ensures that those patients with red-flag symptoms are referred appropriately. It is necessary to ensure that patients who have ongoing concerns can get further information and advice via telehealth communications with a knowledgeable clinician, without having a face-to-face office or hospital appointment.

XSL•FO
RenderX

Textbox 2. Step 2 of the Patient Guide.

Now that I have described my symptoms, what should I do?

Minimal or no discomfort and no other symptoms, "Yes" on 1A and "No" on all other questions.

It is not necessary for you to see a licensed clinician in the office.

If concerned about activities that can cause spine pain, seek information from a reliable source and stay active. Consider a telehealth communication with your licensed clinician if you have questions.

Mild spine pain, "Yes" on 1B and "No" to all other questions.

Office-based treatment by a licensed clinician is not usually necessary. Consider self-care recommendations and telehealth communication with a licensed clinician.

Self-care suggestions:

- Keep moving and try to maintain as much of your normal activity as you can.
- Try applying heat or cold over the area of discomfort for 20 minutes maximum. Take care to avoid skin burns if too hot or cold.
- Avoid prolonged sitting or stationary positions.
- Consider home exercises and relaxation techniques such as yoga and tai chi.
- Consider over-the-counter medication such as paracetamol, ibuprofen, and naproxen. These medications should not be taken without first contacting a medical physician if they have caused any prior adverse symptoms or if you have symptoms consistent with the flu.
- Mild spine pain is very common and usually does not become disabling. Usually the pain will improve or resolve within a few days or weeks.
- If the symptoms persist for prolonged period (greater than 6 weeks), consider contacting a licensed clinician who is knowledgeable about spinal disorders.
- Tests such as X-ray and magnetic resonance imaging are not very helpful in the decision of which treatment to consider.

Moderate spine pain, "Yes" on 1C and 3B.

Office-based treatment by a licensed clinician is not always necessary. Initially try self-care.

Self-care suggestions:

- Keep moving and try to maintain as much of your normal activity as you can.
- Try applying heat or cold over the area of discomfort for 20 minutes maximum. Take care to avoid skin burns if too hot or cold.
- Avoid prolonged sitting or stationary positions.
- Consider home exercises and relaxation techniques such as yoga and tai chi.
- Consider over-the-counter medication such as paracetamol, ibuprofen, and naproxen. These medications should not be taken without first contacting a medical physician if they have caused any prior adverse symptoms or if you have symptoms consistent with the flu.
- Moderate spine pain is not uncommon. Usually the pain will improve or resolve over time.
- If the pain does not resolve over a period of 2 weeks, or the pain is intolerable, it may be necessary to seek the advice of a licensed clinician for recommendations on how to reduce symptoms.
- X-ray or magnetic resonance imaging or other testing is generally not required unless the pain does not improve over a period of 6 weeks.

Severe spine pain, "Yes" on 1D and 3C and "No" on all other questions.

Office-based treatment by a licensed clinician is not always necessary. Initially try self-care.

Self-care suggestions:

- Keep moving and try to maintain as much of your normal activity as you can.
- Try applying heat or cold over the area of discomfort for 20 minutes maximum. Take care to avoid skin burns if too hot or cold.
- Avoid prolonged sitting.
- Consider over-the-counter medication such as paracetamol, ibuprofen, and naproxen. These medications should not be taken without first contacting a medical physician if they have caused any prior adverse symptoms or if you have symptoms consistent with the flu.
- Severe spine pain is less common. Even severe spine pain, in most cases, in the absence of major injury, nerve symptoms, or serious disease, tends to improve or resolve with time.
- If the pain is intolerable, it may be necessary to seek the advice of a licensed clinician for recommendations on whether you should be seen in an outpatient setting, require testing, and how to reduce symptoms.
- X-ray or magnetic resonance imaging or other testing may be necessary.

XSL•FO
RenderX

Symptoms consistent with nerve involvement: pain, numbness, or tingling in arms or legs ("Yes" to 1D and/or 1E). Muscle weakness ("Yes" to 1F). Loss of balance ("Yes" to 1G). New onset of bowel or bladder problems ("Yes" to 1H). Severe new onset of headaches or chest pain ("Yes" to questions 2B, 2C, 2D, and/or 2E).

Office-based treatment by a licensed clinician may be necessary.

- If symptoms are of recent onset (less than one week), consider contacting a health care practitioner or emergency room for a detailed examination.
- If you have experienced recent onset of incontinence, loss of bowel or bladder function, or marked loss of balance, muscle weakness, or difficulty walking, go to the emergency department. The licensed clinician or emergency physician will determine if you require X-ray or magnetic resonance imaging, or other testing and treatment.

Possible spine/bone fracture, severe fall, or accident with severe spine pain ("Yes" on 1C and 4B).

Emergency treatment is necessary.

- Have someone call for an ambulance or nearest help.
- Keep still and do not move.

Possible complication of a serious problem that is affecting the spine, "Yes" on any of the conditions noted in question 5.

Seek care from a licensed clinician or medical specialist to determine whether your serious disease is causing your spine-related symptoms. Diseases such as cancer, certain infections, and inflammatory rheumatologic diseases can impact the spine and cause pain.

The Clinician Guide

The consensus of the panel was that a clinician guide should be clear and useful to all clinicians, irrespective of their professional qualifications and experience, as well as those who may not be familiar with spine care guidelines. It should be consistent with accepted evidence-based clinical practice. It should be sufficiently comprehensive to allow for documentation of the clinician thought process and ensure adequate medical record keeping, while at the same time not be overly time consuming. A 5-step process was recommended:

1. Obtain and document patient demographics, location of primary clinical symptoms, and psychosocial information. The guide recommends a minimum of the following: patient name, date, age, sex or gender, normal occupation, current work situation, worker's compensation status (if applicable), location of primary complaint, comorbidities, level of anxiety and concern about the condition, and psychiatric history. A downloadable page with these questions is made available but it is recognized that clinicians may wish to develop their own charting method and include additional information on patient demographics or take a more comprehensive psychosocial history.

2. Ensure that the questionnaire recommended in the Patient Guide (Textbox 1) has been completed. The clinician should complete this questionnaire with the patient if it is incomplete and review the answers with the patient to ensure that the questions are understood and answered correctly. Clinicians may also want to quantify the patient's level of pain and disability to a greater extent using visual analog scale (VAS) scores and the NIH Pain Consortium Impact Classification (NIH PCIC) score [20], as recommended in the GSCI publication [19]. The values for each of the questions on pain and disability are consistent with the terms minimal, mild, moderate, and severe symptoms as noted in the Patient Guide. Rating scales are not included in the Patient Guide. It is also noted that any valid disability scale that provided similar differentiation of levels of disability could be used instead of the NIH PCIC score.

3. Determine the classification of the patient's complaints according to the GSCI Classification of Spinal Disorders (Textbox 3, Major Classes). The nomenclature of the GSCI classification system is consistent with the "minimal" through "severe" terminology noted in the patient guide. In the Patient Guide, the patient's definition and judgment of the terms "minimal" through "severe" is used. The class of spinal disorder in the Clinician Guide requires consideration of the VAS score, NIH PCIC score, or a similar disability score, together with questioning of the patient to understand what these terms mean to them.

XSL•FO
RenderX

Textbox 3. Steps 3 and 4 of Clinician Guide.

Step 3 of Clinician Guide: What GSCI class of back or neck pain best represents the patient's symptoms?

Step 4 of Clinician Guide: Determination of GSCI spinal disorder subclassification that best describes the patient presentation [16].

Class 0: minimal or no discomfort but no other symptoms. "Yes" on 1A and "No" on all other questions.

- Subclass 0a: no history of risk factors
- Subclass 0b: history of risk factors

It is not necessary to see a clinician in his or her office. Telehealth may be considered.

Class I: mild spine pain. "Yes" on 1B and "No" to all other questions.

- Subclass Ia: acute (duration <3 months)
- Subclass Ib: chronic (duration >3 months)

Office-based treatment by a clinician is not usually necessary. Telehealth may be necessary. Advice and reassurance may be helpful.

Class II: moderate or severe spine pain. "Yes" on 1D and 3C or 3D and "No" on all other questions.

- Subclass IIa: acute (duration <3 months) moderate pain and disability
- Subclass IIb: chronic (duration >3 months) moderate pain and disability
- Subclass IIc: acute, severe pain
- Subclass IId: chronic, severe pain and disability

IIa and IIb: Office-based treatment by a clinician is not always necessary. Telehealth is important. Advise and reassure. Regular telehealth follow-up may be necessary.

IIc: Office-based treatment by a clinician may be necessary. In many cases, acute symptoms can be managed via telehealth in the absence of red flags.

IId: Office-based treatment by a clinician may not be necessary unless there is a flare-up of incapacitating symptoms. Consider telehealth consultation first.

Class III: symptoms consistent with nerve problem. Pain, numbness, or tingling in arms or legs, new onset marked muscle weakness, and/or new onset of bowel or bladder problems ("Yes" to 1D and/or 1E and/or 1F and/or 1G and/or 1H). Severe new onset of headaches or chest pain ("Yes" to 2B and/or 2C and/or 2D and/or 2E). Consider additional questions on gait difficulty, loss of balance, and/or loss of hand function (including clumsiness and change in dexterity) that may represent symptoms of myelopathy.

- Subclass IIIa: minor or nonprogressive
- Subclass IIIb: acute, major or progressive
- Subclass IIIc: chronic and stable

Office or emergency room treatment by appropriate clinician is necessary if the symptoms are acute or progressive.

Class IV: possible spine/bone fracture. Severe fall or accident with severe spine pain ("Yes" on 1C and 4B).

- Subclass IVa: stable spine structural pathology with no serious symptoms or red flags. Office-based treatment is not necessary. Telehealth may be considered for advice and reassurance.
- Subclass IVb: acute (eg, fracture) or chronic (eg, instability) spine structural pathology that correlates with symptoms. Emergency treatment is necessary.

Subclass V: possible complication of a serious problem that is affecting the spine. "Yes" on any of the conditions noted in question 5.

- Subclass Va: severe acute spine pathology. Requires immediate attention (emergency).
- Subclass Vb: slowly progressive spinal pathology. Requires intervention (nonemergency).
- Subclass Vc: symptoms originating from nonspinal pathology. Requires immediate attention (emergency). Referral to patient's medical family/general physician or specialist to determine whether serious disease is causing the patient's spine-related symptoms. Advise patient if emergency attention is required.

4. Determine the GSCI subclassification of the symptom pattern. This requires three additional questions.

Question 1 is the determination of the duration of the symptoms. Most spine guidelines recommend different interventions for acute versus chronic spine-related symptoms [21,22]. The differentiation between these terms varies somewhat in review articles and guidelines, with acute being defined as a few days to a few weeks and sometimes subdivided into acute (1-4 weeks) and subacute (4-12 weeks) [21,23]. It was elected to use the recommendations of the NIH Task Force on Research Standards, which defines chronic low back pain as present for >3 months

XSL•FO
RenderX

and present for at least half the days in the past 6 months [20]. In specific patients, especially in the red flag classifications, acute may involve a very recent onset of symptoms and may require immediate or emergency care (1 day to 1 week).

Question 2 is the determination as to whether symptoms are progressive or stable. Again, these subclasses are most important in the red-flag and neurologic symptom classifications but also are useful in determining how to manage pain symptoms.

Question 3 includes enquiries on risk factors for spine pain and comorbidities. The information gained from this question allows the clinician to determine the presence of psychosocial factors known to impact disability, but also can be important for the understanding of a patient's general health and activity level and whether comorbidities are controlled. The subclassifications are noted in Textbox 3 under each of the major classes.

5. Treatment interventions. It was not possible to list the extensive number of treatment recommendations for each class of spinal disorder. The clinician is advised to review the GSCI flash cards for each class and subclass of spinal disorder [19]. Clinicians are also advised to refer to websites presenting leading government guidelines [21,24] and those developed by major professional spine care societies, including guidelines specifically adopted by family physicians [22,23], chiropractors [25], and physical therapists [26] and which are felt to be reasonably consistent with each other. Lead statements on the importance of self-care, advice and reassurance, and exercise, as well as potential adverse events from pharmaceuticals, are included. A further statement stresses that imaging is rarely helpful in spinal pain complaints in the absence of other symptoms or red flags. Finally, it was felt that patients should be empowered to embrace self-care but should be made to feel that they are being cared for through follow-up contact with their clinician when they have questions or concerns (Textbox 4).

The Clinician Guide includes a textbox that provides guidance on some of the legal and ethical issues related to telehealth that may not be immediately clear to a practicing clinician (Textbox 5). These include licensing, malpractice, patient informed consent and patient record keeping and confidentiality (Textbox 5).

XSL•FO
RenderX

Textbox 4. Step 5 of the Clinician Guide: treatment considerations.

On completion of the questionnaire and determination of the class and subclass of spinal disorder, consider the following:

1. Follow World Health Organization, US Central Disease Control, UK National Health Service, and other national government agency guidelines for the current status and recommendations regarding the prevention and management of COVID-19.

2. Consult national evidence-based treatment guidelines for evidence-based treatment options for each subclass of spinal disorder. Consider the Global Spine Care Initiative flash cards for a review of these guidelines and the interventions recommended for each class and subclass of spinal disorder.

Other resources:

- UK National Health Service Back Pain Guidelines
- American College of Physicians Clinical Practice Guideline
- Canadian Chiropractic Guideline Initiatives on Neck Pain, Low Back Pain, Self-Management Resources, Exercise Videos
- American Physical Therapy Association Clinical Practice Guidelines
- Task Force on Neck Pain Executive Summary

3. Advise patients over the phone, provide video consultations (in some regions partially or fully reimbursed by health insurance), and use social media including Facebook to educate patients. Considerations include the following:

- Research existing online educational media that satisfies the requirements for each individual patient.
- Consider leading online educational classes with patients.
- Consider leading or referring to online yoga, including the World Spine Care Yoga Project, tai chi, Pilates, or rehabilitative exercise group classes with patients.

4. For Class 0, Class I, and moderate Class II, reinforce that the current research suggests that self-care is usually enough to control symptoms. Provide advice and reassurance that may relieve the pain and aid recovery.

5. The taking of over-the-counter medication including paracetamol and nonsteroidal anti-inflammatory drugs (NSAIDs), including ibuprofen and naproxen, is recommended for short-term relief of back and neck pain by most evidence-based guidelines. Patients should be aware of potential adverse events, including gastrointestinal bleeding and ulcers, and cardiovascular and renal disorders. The possibility that NSAIDs may negatively impact COVID-19 pulmonary symptoms has been raised but not confirmed at the time of publication [27]. Clinicians should be able to advise patients or refer patients to a reliable source of information when asked by patients about taking these medications.

6. Reinforce that testing such as X-ray and magnetic resonance imaging rarely help inform the decision of which treatment to consider. Referral to a surgeon is unlikely to be necessary unless the patient is incapacitated, has had a significant injury, or has red flags that result in Class III, IV, or V assessment.

7. Be available for regular follow-up contact with the patient (virtually).

8. Recognize that patient symptoms on rare occasions can get worse, so that they could eventually fall into a different class or subclass and require different or more immediate care than originally recommended.

9. Patients should be empowered to embrace self-management but still feel they are being cared for.

Textbox 5. Clinician Guide for telehealth considerations.

1. Find out if telehealth communication is allowed in your jurisdiction.

2. In many countries and regional jurisdictions, including the United States, be careful about providing telehealth sessions to patients who live outside of your jurisdiction. You may be sanctioned for practicing without a license. Follow all legal recommendations for telehealth.

3. Check with your malpractice carrier to make sure there are no restrictions for coverage while conducting telehealth sessions.

4. In addition, make sure to obtain a verbal consent from the patient in which they acknowledge there are limitations to a telehealth session, which is not the same as a physical examination (eg, it is not possible to do a detailed neurological examination on the phone).

5. Be careful not to violate any patient privacy rules. Use a video program or platform that provides an appropriate level of patient confidentiality to satisfy regional or national laws.

6. Find out if compensation is allowed, including billing requirements that may necessitate mandatory documentation regarding telehealth sessions with your patients.

7. Keep records of your consultation.

8. Make sure to get the address of the physical location of the patient. This may be necessary in case an emergency arises during the telehealth session.

The Clinician Guide concludes with a section that has the following heading: “During this period of social and physical distancing what precautions must I consider when seeing a patient while protecting both the patient, me, and my staff from infection?” This section provides guidance that is widely recommended by the World Health Organization (WHO),

XSL•FO
RenderX

Centers for Disease Control and Prevention (CDC), the UK National Health Service (NHS), and other government agencies based on the geographical area of practice. It is noted that these guidelines are fluid and clinicians should stay informed of best practices and requirements in their jurisdictions.

The Patient Guide and Clinician Guide can be downloaded in their professionally produced educational formats [28,29].

Discussion

Principal Results

The guides recommended in this article have been developed by an international team from 10 countries on 4 continents. Care was taken to have a large panel of 29 participants that included members from most of the nonsurgical spine care professions and specialties as well as patients. The process of developing these guides was an open iterative approach that allowed all participants to edit, comment on, or modify the recommendations. The guides are therefore not dependent on the recommendations from a single published spine care guideline, the opinions of a small number of clinicians or scientists, a single culture or part of the world, or a single spine care profession or discipline. These guides were extrapolated from a recent review of the most current international spine care guidelines, which were published in September 2018 [11]. The Patient and Clinician Guides minimize any potential conflicting information given to patients and their clinicians. Finally, these guides cover the entire spectrum of spinal disorders from minimal pain or concerns, through mild, moderate, and severe spine pain, neurological symptoms, trauma, and systemic diseases that can result in spine-related symptoms.

Comparison With Prior Work

The Patient and Clinician Guides described in this article are not the first or only guides that have been developed to help clinicians and patients during the current COVID-19 pandemic. The current COVID-19 pandemic has generated a great deal of interest and a number of articles that have focused on the difficulties of managing patients with chronic pain when in-person care is not feasible [8,9,30]. Most of these articles have focused on the challenges and terminology, or different telehealth or eHealth tools. There are also a number of articles that address specific subgroups of patients with spinal disorders such as low back pain [31], general musculoskeletal disorders [7,10,32,33], and inflammatory joint diseases [34], as well as triage of surgical patients [3]. Many of the articles addressing spine care during COVID-19 make recommendations based on a single or unspecified guideline [35]. We are not aware of any specific international, interdisciplinary comprehensive spine care guide that has been developed for use during a pandemic where lockdown restrictions have limited in-person care. None that we have found have addressed the full spectrum of spinal disorders with recommendations for both patients and clinicians.

Future Considerations

The COVID-19 pandemic and resulting limitation of in-person contact between patients and their clinicians will come to an end at some point. We are confident that the utility of these guides will persist after the pandemic and that the new normal will make greater use of virtual spine care interactions. This is likely to be more important where there are geographically dispersed populations or where spine specialists are not available, accessible, or affordable.

People with spine-related complaints or disabilities are rarely receiving care that is consistent with current evidence-based guidelines. This is true in high-, middle-, and low-income countries [13].

In high-income countries, the problem in the delivery of spine care is widely considered to be overutilization of many costly yet low-value interventions such as indiscriminate use of imaging, opioids, and invasive procedures [36]. The broad usage of the proposed linked guides for clinicians and patients ensures similar assessment and management options. This has the potential to reduce the overutilization of unnecessary and expensive interventions while increasing the empowerment of patients to self-manage uncomplicated spinal pain with the assistance of their clinician, either through direct in-person consultation or via telehealth communication.

In low- and middle-income countries, physical distancing between patients and clinicians trained and capable of providing evidence-based care is not available for most of the public. Foster et al [13] pointed out in their review of the literature that in many countries, the primary management strategy for patients who present with spine pain is either emergency room care, admission to hospital, or direct referral to an orthopedic surgeon. Irving et al [37] noted that for 50% of the world population, primary care physicians spend less than 5 minutes of face-to-face time with patients during a consultation. The most common intervention for back pain is limited to the prescription of NSAIDs or opioids. The likelihood that primary care or family practice clinicians could find the time to offer evidence-based spine care—which requires a clinical assessment and includes education, reassurance, and advice on lifestyle and exercise, as well as symptom relief—is remote. For much of the world, there is a dearth of any effective care by nonsurgical clinicians who have specific training and an interest in managing patients with spinal disorders. The proposed guides have the potential to reduce this deficit.

Telehealth Considerations

This lack of alternatives raises the possibility that patient and clinician guides such as those presented in this paper and those developed by local agencies (eg, the Ontario Inter-professional Spine Assessment and Education Clinics [38]) could be used to connect patients with clinicians knowledgeable about the management of spinal disorders through telehealth. Such tools have the potential to improve care and reduce the load on emergency rooms, hospitals, and advanced specialists and surgeons without burdening family and primary care physicians. Patients would have access to information and the ability to consult virtually with a knowledgeable clinician. This will hopefully result in improved outcomes and well-being. There is early evidence that telehealth [39] or digital care can have a positive impact on low back pain [39-41] and that these modalities are well accepted by patients, at least in conjunction with in-person access to a clinician when needed. The more general use of an app to educate patients with musculoskeletal

XSL•FO
RenderX

pain is also being explored [32]. This is becoming more feasible with the extensive and increasing availability of access to mobile phones in many low- and middle-income countries [42,43].

Telehealth is an expanding field with increasing understanding of what is possible through this medium for both acute and chronic consultations [43]. There are now several publications that describe a virtual neurologic examination [44,45]. There is also a growing number of publications addressing the possibility of virtual spine examination [46,47,48], although none of these virtual examinations have been studied for validity or reliability.

Limitations

The GSCI classification was developed through an iterative, broad consensus process and has not been validated through clinical trials at the time of writing. Although the guides were developed with extensive international and interprofessional representation, there is underrepresentation from Asian, Middle Eastern, and Western European countries, which may reduce their acceptance or compatibility with cultural norms in these and other regions of the world. Future revisions of these guides would benefit from input from wider geographic and cultural representation. Work is currently underway to test the GSCI model of care in three underserved communities around the world. Studies are also being proposed to test these guides to determine their acceptance by patients and clinicians, the mechanism through which these guides can be implemented, and their impact on patients with spinal disorders. It is also necessary to determine whether these guides influence the use of unnecessary, costly, and often ineffective interventions that do not comply with current evidence-based guidelines. The consideration of guides for spine-related disorders is a new and expanding field and the recommendations noted in Textboxes 2 and 3 should be considered fluid and are likely to change as our experience with telehealth guides increases.

Conclusion

This article describes the development of patient and clinician guides for the management of patients with all levels of spine-related symptoms who are, for various reasons, unable to access/conduct in-person office-based care. The guides were developed by an international, interdisciplinary panel of 29 participants from 10 countries, with the goal of providing guidance to both patients and their clinicians in situations involving physical distancing. The initial goal was to fill the gap in care during the COVID-19 pandemic lockdown period. It is believed, however, that these guides may be helpful in other situations where there is limited availability of evidence-based care in underserved communities. They should provide patients with both guidance and alternative options for access to spine care clinicians either digitally or via telehealth.

Acknowledgments

The development of these guidelines was supported by a grant from The Skoll Foundation to World Spine Care, a global charity on a mission to improve lives in underserved communities and the aim of creating a world in which everyone has access to the highest quality spine care possible.

Conflicts of Interest

None of the authors or participants in the development of these guides received any compensation or financial benefit for their participation.

References

1. Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, et al. Coronavirus disease 2019 (COVID-19): A literature review. J Infect Public Health 2020 May;13(5):667-673 [FREE Full text] [doi: 10.1016/j.jiph.2020.03.019] [Medline: 32340833]
2. Valencia D. Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. Cureus 2020 Mar 24;12(3):e7386 [FREE Full text] [doi: 10.7759/cureus.7386] [Medline: 32337113]
3. Ehrlich H, McKenney M, Elkbuli A. Protecting our healthcare workers during the COVID-19 pandemic. Am J Emerg Med 2020 Jul;38(7):1527-1528 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.024] [Medline: 32336585]
4. Jean WC, Ironside NT, Sack KD, Felbaum DR, Syed HR. The impact of COVID-19 on neurosurgeons and the strategy for triaging non-emergent operations: a global neurosurgery study. Acta Neurochir (Wien) 2020 Jun 21;162(6):1229-1240 [FREE Full text] [doi: 10.1007/s00701-020-04342-5] [Medline: 32314059]
5. Brady O, Nordin M, Hondras M, Outerbridge G, Kopansky-Giles D, Côté P, et al. Global Forum: Spine Research and Training in Underserved, Low and Middle-Income, Culturally Unique Communities: The World Spine Care Charity Research Program's Challenges and Facilitators. The Journal of Bone and Joint Surgery 2016;98(24):e110. [doi: 10.2106/jbjs.16.00723]
6. Haldeman S, Nordin M, Outerbridge G, Hurwitz EL, Hondras M, Brady O, et al. Creating a sustainable model of spine care in underserved communities: the World Spine Care (WSC) charity. Spine J 2015 Nov 01;15(11):2303-2311. [doi: 10.1016/j.spinee.2015.06.046] [Medline: 26096472]
7. Chehade MJ, Yadav L, Kopansky-Giles D, Merolli M, Palmer E, Jayatilaka A, et al. Innovations to improve access to musculoskeletal care. Best Pract Res Clin Rheumatol 2020 Oct;34(5):101559. [doi: 10.1016/j.berh.2020.101559] [Medline: 32718885]

XSL•FO
RenderX

8. Deer T, Sayed D, Pope JE, Chakravarthy KV, Petersen E, Moeschler SM, ASPN COVID Workgroup. Emergence From the COVID-19 Pandemic and the Care of Chronic Pain: Guidance for the Interventionalist. Anesth Analg 2020 Aug;131(2):387-394 [FREE Full text] [doi: 10.1213/ANE.0000000000005000] [Medline: 32452905]
9. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May 2;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
10. Yadav SK, Kar BK, Banta A, Kumar A. Creating Backup Management Resources for Spine Care during the Coronavirus Disease 2019 Pandemic. Asian Spine J 2020 Jun;14(3):382-384 [FREE Full text] [doi: 10.31616/asj.2020.0193] [Medline: 32403895]
11. Haldeman S, Nordin M, Chou R, Côté P, Hurwitz EL, Johnson CD, et al. The Global Spine Care Initiative: World Spine Care executive summary on reducing spine-related disability in low- and middle-income communities. Eur Spine J 2018 Sep;27(Suppl 6):776-785. [doi: 10.1007/s00586-018-5722-x] [Medline: 30151809]
12. Johnson CD, Haldeman S, Nordin M, Chou R, Côté P, Hurwitz EL, et al. The Global Spine Care Initiative: methodology, contributors, and disclosures. Eur Spine J 2018 Sep;27(Suppl 6):786-795. [doi: 10.1007/s00586-018-5723-9] [Medline: 30151808]
13. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. The Lancet 2018 Jun;391(10137):2368-2383. [doi: 10.1016/s0140-6736(18)30489-6]
14. Acaroğlu E, Nordin M, Randhawa K, Chou R, Côté P, Mmopelwa T, et al. The Global Spine Care Initiative: a summary of guidelines on invasive interventions for the management of persistent and disabling spinal pain in low- and middle-income communities. Eur Spine J 2018 Sep 10;27(Suppl 6):870-878. [doi: 10.1007/s00586-017-5392-0] [Medline: 29322309]
15. Ameis A, Randhawa K, Yu H, Côté P, Haldeman S, Chou R, et al. The Global Spine Care Initiative: a review of reviews and recommendations for the non-invasive management of acute osteoporotic vertebral compression fracture pain in low- and middle-income communities. Eur Spine J 2018 Sep 16;27(Suppl 6):861-869. [doi: 10.1007/s00586-017-5273-6] [Medline: 29038868]
16. Cedraschi C, Nordin M, Haldeman S, Randhawa K, Kopansky-Giles D, Johnson CD, et al. The Global Spine Care Initiative: a narrative review of psychological and social issues in back pain in low- and middle-income communities. Eur Spine J 2018 Sep 27;27(Suppl 6):828-837. [doi: 10.1007/s00586-017-5434-7] [Medline: 29374779]
17. Chou R, Côté P, Randhawa K, Torres P, Yu H, Nordin M, et al. The Global Spine Care Initiative: applying evidence-based guidelines on the non-invasive management of back and neck pain to low- and middle-income communities. Eur Spine J 2018 Sep 19;27(Suppl 6):851-860. [doi: 10.1007/s00586-017-5433-8] [Medline: 29460009]
18. Haldeman S, Johnson CD, Chou R, Nordin M, Côté P, Hurwitz EL, et al. The Global Spine Care Initiative: care pathway for people with spine-related concerns. Eur Spine J 2018 Sep;27(Suppl 6):901-914. [doi: 10.1007/s00586-018-5721-y] [Medline: 30151811]
19. Haldeman S, Johnson CD, Chou R, Nordin M, Côté P, Hurwitz EL, et al. The Global Spine Care Initiative: classification system for spine-related concerns. Eur Spine J 2018 Sep 27;27(Suppl 6):889-900. [doi: 10.1007/s00586-018-5724-8] [Medline: 30151807]
20. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014 Jun;15(6):569-585 [FREE Full text] [doi: 10.1016/j.jpain.2014.03.005] [Medline: 24787228]
21. National Institute for Health and Care Excellence (UK). Low back pain and sciatica in over 16s: assessment and management. In: National Institute for Health and Care Excellence: Clinical Guidelines. London, UK: National Institute for Health and Care Excellence (UK); Sep 22, 2020.
22. Qaseem A, Wilt TJ, McLean RM, Forciea MA. Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. Ann Intern Med 2017 Feb 14;166(7):514. [doi: 10.7326/m16-2367]
23. Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle P, Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 02;147(7):478-491 [FREE Full text] [doi: 10.7326/0003-4819-147-7-200710020-00006] [Medline: 17909209]
24. National Institute of Neurological Disorders and Stroke - National Institutes of Health. Low back pain fact sheet for patients and the public. J Pain Palliat Care Pharmacother 2004;18(4):95-110. [Medline: 15760814]
25. Canadian Chiropractic Guideline Initiative. 2020. URL: https://www.ccgi-research.com [accessed 2021-02-04]
26. Delitto A, George SZ, Van Dillen L, Whitman JM, Sowa G, Shekelle P, Orthopaedic Section of the American Physical Therapy Association. Low back pain: Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. J Orthop Sports Phys Ther 2012 Apr;42(4):A1-57 [FREE Full text] [doi: 10.2519/jospt.2012.42.4.A1] [Medline: 22466247]
27. Little P. Non-steroidal anti-inflammatory drugs and covid-19. BMJ 2020 Mar 27;368:m1185. [doi: 10.1136/bmj.m1185] [Medline: 32220865]

XSL•FO
RenderX

28. World Spine Care. Managing Back and Neck Pain in the Time of COVID-19. Clinician Guide. URL: https://www.worldspinecare.org/wp-content/uploads/2020/04/World-Spine-Care-COVID-19-Clinician-Guide.pdf [accessed 2021-02-05]
29. World Spine Care. Managing Back and Neck Pain in the Time of COVID-19. Patient Guide. URL: https://www.worldspinecare.org/wp-content/uploads/2020/04/World-Spine-Care-COVID-19-Patient-Guide.pdf [accessed 2021-02-05]
30. Shanthanna H, Strand NH, Provenzano DA, Lobo CA, Eldabe S, Bhatia A, et al. Caring for patients with pain during the COVID-19 pandemic: consensus recommendations from an international expert panel. Anaesthesia 2020 Jul 26;75(7):935-944 [FREE Full text] [doi: 10.1111/anae.15076] [Medline: 32259288]
31. Amorim AB, Pappas E, Simic M, Ferreira ML, Tiedemann A, Jennings M, et al. Integrating Mobile health and Physical Activity to reduce the burden of Chronic low back pain Trial (IMPACT): a pilot trial protocol. BMC Musculoskelet Disord 2016 Jan 19;17(1):36 [FREE Full text] [doi: 10.1186/s12891-015-0852-3] [Medline: 26787469]
32. Bailey J, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
33. Viswanath A, Monga P. Working through the COVID-19 outbreak: Rapid review and recommendations for MSK and allied heath personnel. J Clin Orthop Trauma 2020 May;11(3):500-503 [FREE Full text] [doi: 10.1016/j.jcot.2020.03.014] [Medline: 32292256]
34. Schulze-Koops H, Holle J, Moosig F, Specker C, Aries P, Burmester G, im Auftrag der Kommission Pharmakotherapie und des Vorstandes der Deutschen Gesellschaft für Rheumatologie. [Current guidance of the German Society of Rheumatology for the care of patients with rheumatic diseases during the SARS-CoV-2/Covid 19 pandemic]. Z Rheumatol 2020 May 27;79(4):385-388 [FREE Full text] [doi: 10.1007/s00393-020-00799-y] [Medline: 32342184]
35. Zou J, Yu H, Song D, Niu J, Yang H. Advice on Standardized Diagnosis and Treatment for Spinal Diseases during the Coronavirus Disease 2019 Pandemic. Asian Spine J 2020 Apr;14(2):258-263 [FREE Full text] [doi: 10.31616/asj.2020.0122] [Medline: 32222720]
36. Buchbinder R, van Tulder M, Öberg B, Costa LM, Woolf A, Schoene M, Lancet Low Back Pain Series Working Group. Low back pain: a call for action. Lancet 2018 Jun 09;391(10137):2384-2388. [doi: 10.1016/S0140-6736(18)30488-4] [Medline: 29573871]
37. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 2017 Nov 08;7(10):e017902 [FREE Full text] [doi: 10.1136/bmjopen-2017-017902] [Medline: 29118053]
38. Ontario Inter-professional Spine Assessment and Education Clinics. 2020. URL: https://www.isaec.org [accessed 2020-02-04]
39. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov 15;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
40. Rutledge T, Atkinson JH, Holloway R, Chircop-Rollick T, D'Andrea J, Garfin SR, et al. Randomized Controlled Trial of Nurse-Delivered Cognitive-Behavioral Therapy Versus Supportive Psychotherapy Telehealth Interventions for Chronic Back Pain. J Pain 2018 Sep;19(9):1033-1039. [doi: 10.1016/j.jpain.2018.03.017] [Medline: 29673974]
41. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019 Jan 7;2(1):1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
42. Bastawrous A, Armstrong MJ. Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. J R Soc Med 2013 Apr;106(4):130-142 [FREE Full text] [doi: 10.1177/0141076812472620] [Medline: 23564897]
43. Totten AM, Hansen RN, Wagner J, Stillman L, Ivlev I, Davis-O'Reilly C, et al. Telehealth for Acute and Chronic Care Consultations. Rockville, MD, USA: Agency for Healthcare Research and Quality (US); 2019.
44. Al Hussona M, Maher M, Chan D, Micieli JA, Jain JD, Khosravani H, et al. The Virtual Neurologic Exam: Instructional Videos and Guidance for the COVID-19 Era. Can J Neurol Sci 2020 Sep 21;47(5):598-603 [FREE Full text] [doi: 10.1017/cjn.2020.96] [Medline: 32434626]
45. Hatcher-Martin JM, Adams JL, Anderson ER, Bove R, Burrus TM, Chehrenama M, et al. Telemedicine in neurology: Telemedicine Work Group of the American Academy of Neurology update. Neurology 2020 Jan 07;94(1):30-38. [doi: 10.1212/WNL.0000000000008708] [Medline: 31801829]
46. Satin AM, Lieberman IH. The Virtual Spine Examination: Telemedicine in the Era of COVID-19 and Beyond. Global Spine J 2020 Jul 31:1 [FREE Full text] [doi: 10.1177/2192568220947744] [Medline: 32734784]
47. Yoon JW, Welch RL, Alamin T, Lavelle WF, Cheng I, Perez-Cruet M, et al. Remote Virtual Spinal Evaluation in the Era of COVID-19. Int J Spine Surg 2020 Jun;14(3):433-440 [FREE Full text] [doi: 10.14444/7057] [Medline: 32699768]
48. Viswanath A, Monga P. Working through the COVID-19 outbreak: Rapid review and recommendations for MSK and allied heath personnel. J Clin Orthop Trauma 2020 May;11(3):500-503 [FREE Full text] [doi: 10.1016/j.jcot.2020.03.014] [Medline: 32292256]

Abbreviations

CDC: Centers for Disease Control and Prevention

XSL•FO
RenderX

GSCI: Global Spine Care Initiative
MRI: magnetic resonance imaging
NHS: National Health Service
NIH: National Institutes of Health
NIH PCIC: National Institutes of Health Pain Consortium Impact Classification
NSAID: nonsteroidal anti-inflammatory drug
PCIC: Pain Consortium Impact Classification
VAS: visual analog scale
WHO: World Health Organization
WSC: World Spine Care

Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by T Aslanidis, A Hidki,Dr; comments to author 25.11.20; revised version received 30.11.20; accepted 19.01.21; published 17.02.21.

Please cite as:
Haldeman S, Nordin M, Tavares P, Mullerpatan R, Kopansky-Giles D, Setlhare V, Chou R, Hurwitz E, Treanor C, Hartvigsen J, Schneider M, Gay R, Moss J, Haldeman J, Gryfe D, Wilkey A, Brown R, Outerbridge G, Eberspaecher S, Carroll L, Engelbrecht R, Graham K, Cashion N, Ince S, Moon E
Distance Management of Spinal Disorders During the COVID-19 Pandemic and Beyond: Evidence-Based Patient and Clinician Guides From the Global Spine Care Initiative
JMIR Public Health Surveill 2021;7(2):e25484
URL: http://publichealth.jmir.org/2021/2/e25484/
doi:10.2196/25484
PMID:33471778

©Scott Haldeman, Margareta Nordin, Patricia Tavares, Rajani Mullerpatan, Deborah Kopansky-Giles, Vincent Setlhare, Roger Chou, Eric Hurwitz, Caroline Treanor, Jan Hartvigsen, Michael Schneider, Ralph Gay, Jean Moss, Joan Haldeman, David Gryfe, Adam Wilkey, Richard Brown, Geoff Outerbridge, Stefan Eberspaecher, Linda Carroll, Reginald Engelbrecht, Kait Graham, Nathan Cashion, Stefanie Ince, Erin Moon. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 17.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX