Appendix: Content Summary of Selected Best Papers 2021 for the IMIA Yearbook Section Consumer Health Informatics

Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ

Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: nationwide cross-sectional study

During the rapid escalation of the COVID-19 pandemic in March and April 2020, the authors conducted an online survey on the sources of information used and trusted by US adults for acquiring COVID-19 information and ascertained how these sources varied according to key sociodemographic characteristics. They also assessed how differences in information sources were associated with variation in beliefs and levels of knowledge related to COVID-19. The sample was a self-selected nonprobability sample of social media users on Facebook and its affiliated platforms that was recruited through an on-platform advertisement campaign. Participants were sampled in two rounds about one month apart, from March 20th to 30th, and from April 16th to 21th, 2020. Eligible participants included US residents aged ≥18 years. The survey was based on the Health Belief Model, which has been previously utilized in recent surveys on other viral outbreaks. Participants were asked whether or not they used any of the 11 predetermined sources to find information about COVID-19. A variable indicating the total number of sources used by each participant was created by summing the number of "yes" responses for each of the information sources. They were asked to identify the information source they trusted the most. A total of 13,201 respondents were eligible to participate, of whom 11,242 provided data on their sources of COVID-19 information. Males were significantly less likely than females to use all identified sources, excluding spouses/family/friends and religious leaders. Participants aged 40-59 years and ≥60 years were less likely to use government websites compared to those aged 18-38 years (adjusted odds ratio [AOR] 0.59, 95% CI [0.47-0.71]; AOR 0.47, 95% CI [0.37-0.60]). Participants identifying as races other than non-Hispanic White were more likely to use doctors (AOR 1.39, 95% CI 1.18-1.64) and religious leaders (AOR 1.40, 95% CI 1.03-1.86) as a source of information. Those with a bachelor's degree or higher were more likely to use all the sources, except traditional media. The largest individual information source was government websites (87.6%), but they represented only 45.2% of the most trusted information source. These findings suggest that public health professionals seeking to effectively communicate information on COVID-19 must acknowledge and appropriately adapt to disparities in public trust and information source preferences, particularly to address the differences in knowledge and beliefs regarding the pandemic. These findings also suggest that trust in information sources may differ across time, place, culture, and type of disease outbreak, emphasizing the importance of updated surveillance on trends in information seeking behaviors during pandemics.

Picone M, Inoue S, DeFelice C, Naujokas MF, Sinrod J, Cruz VA, Stapleton J, Sinrod E, Diebel SE, Wassman J, Robert E

Social listening as a rapid approach to collecting and analyzing COVID-19 symptoms and disease natural histories reported by large numbers of individuals

Given the severe and rapid impact of COVID-19, the pace of information sharing has been accelerated. However, traditional methods of disseminating and digesting medical information can be time-consuming and cumbersome. To help people determine whether they might have COVID-19, leading health information providers list symptoms of COVID-19 online. But details about impacts and experiences with COVID-19 in daily life pre and post hospitalization might not be communicated to clinicians or captured in publications. The goal of this pilot study was to determine whether, by analyzing voices of patients from social media posts, new information could be identified that might have been missed by other methods of collecting information about patient experiences (such as CDC, WHO, Mayo Clinic, or WebMD lists of symptoms). Two data sets from social media groups of people with or presumed to have COVID-19 were analyzed: a Facebook group poll, and conversation data from a Reddit group including detailed disease natural history-like posts. Supervised learning and natural language processing techniques were used to identify disease burden and disease management concepts. In total, there were 22 symptoms listed in the poll that were mentioned in at least one online list as a direct or related match. In contrast, there were 59 symptoms from the Reddit full conversation data and 80 symptoms mentioned in the Facebook poll that did not appear on any of the four online resource symptom lists. These comparisons show that people are talking about symptoms that are not listed as key symptoms on publicly posted information sites. Symptoms that stood out as unusual were high heart rate, talking causes exhaustion, emotional distress, anxiety, and neurological symptoms such as dizziness and tingling. Furthermore, Reddit and Facebook groups reported that the disease was lasting longer than 3 weeks, and up to 90 days or more. Persistent disease was reported across a wide range of ages, including people younger than age 40. The combined data sets contained more than 60 symptoms that were not listed in the four online symptom lists: CDC, WHO, Mayo Clinic, or WebMD. This pilot study demonstrates that social media data can offer novel insights into patient experiences as a source of real-world data (RWD). Social listening contributes real-world data that falls outside of conventional or traditional modes of research data collection such as conventional randomized controlled clinical trials.

Pobiruchin M, Zowalla R, Wiesner M

Temporal and location variations, and link categories for the dissemination of COVID-19-related information on Twitter during the SARS-CoV-2 outbreak in Europe: Infoveillance study

By monitoring related activities on different social media platforms, public authorities or research institutions can gather valuable insights into regional trends, country-specific trends, or even the global situation. For this infoveillance study, during the early phase of the COVID-19 pandemic, the authors decided to focus on the social media platform Twitter, as the platform allows monitoring of the dynamic pandemic situation on a global scale in real time for different aspects of a topic, languages, as well as regions and even whole countries. Of particular interest were temporal and geographical variations of COVID-19–related tweets, the situation in Europe, and the categories and origin of shared external resources. In total, 16 hashtags were selected for collecting COVID-19–related tweets for the purpose of temporal, geolocation, and link category analyses. Twitter's Streaming application programming interface was used to filter tweets based on these prevalent hashtags related to the COVID-19 outbreak. Each tweet's text and corresponding metadata as well as the user's profile information were extracted and stored into a database. Between February 9 (midnight CEST) and April 11 (11:59 PM CEST), 2020, a total of 21,755,802 distinct tweets posted under the 16 hashtags were collected and stored in the study database. Those tweets were posted by 4,809,842 distinct Twitter accounts of which 83,560 were verified by the platform itself. Overall, the number of daily tweets rose during the study period. The use of #covid19 increased throughout February and March. The trend was similar to the use of #coronavirus. However, the use of the hashtag #COVID—19 was fluctuating periodically. To address geographical variations, all tweets that contained geographical coordinates were included in the analysis. Each tweet was plotted in a geographical map of Europe for each 7-day interval in the observation period. More tweets could be observed in the vicinity of countries' capitals (eg, Paris, Madrid, Vienna, or Berlin) or in densely populated areas such as the Benelux Union or South England. A higher number of tweets with geolocations was observed in Central and Western European countries than compared to Eastern Europe. Interestingly, tweet volumes in Turkey seemed to be higher than in surrounding countries. This study demonstrates how COVID-19–related tweets can be analyzed for a certain region (Europe). With the continuous progression of the pandemic situation, which is to be expected in the next months worldwide, further regions should be analyzed in-depth.