Consumer Informatics and COVID-19 Pandemics: Challenges and Opportunities for Research
Findings from the Yearbook 2020 Section on Education and Consumer Health Informatics

Pascal Staccini[1], Annie Y. S. Lau[2], Section Editors for the IMIA Yearbook Section on Consumer Health Informatics
[1] Département IRIS, URE RETINES, Faculté de Médecine, Université Côte d'Azur, France
[2] Center for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Australia

Summary

Objective: To summarise the state of the art during the year 2020 in consumer health informatics and education, with a special emphasis on "Managing Pandemics with Health Informatics - Successes and Challenges".

Methods: We conducted a systematic search of articles published in PubMed using a predefined set of queries, which identified 147 potential articles for review. These articles were screened according to topic relevance and 15 were selected for consideration of best paper candidates, which were then presented to a panel of international experts for full paper review and scoring. The top five papers were discussed in a consensus meeting. Three papers received the highest score from the expert panel, and these papers were selected to be representative papers on consumer informatics for managing pandemics in the year 2020.

Results: Bibliometrics analysis conducted on words found in abstracts of the candidate papers revealed 4 clusters of articles, where the clustering outcomes explained 77.04% of the dispersion. The first cluster composed of articles related to the use of mobile apps for video consultation and telehealth during the pandemic. The second revealed studies reporting the lived experience of healthcare workers and patients during COVID-19. The third focused on ways people used the internet to seek for health information during the pandemic and the dissemination of fake news. The last cluster composed of articles reporting the use of social listening methods (e.g., via tweet hashtags) to explore the spread of the virus around the world.

Conclusions: The pandemic outbreak of the novel coronavirus disease (COVID-19) constitutes a grave risk to the global community and sparks a significant increase in public interest and media coverage, especially on social media. Consumers are facing a new set of challenges that were not considered before COVID-19, often finding themselves in a world that is constantly changing—blended with facts and fake information—and unable to decide what to do next. Despite most people understanding the good will behind public health policies, one must not forget it is individuals we are supporting and that their personal circumstances may affect how they perceive and comply with these policies. Consumers more than ever need help to make sense of the uncertainty and their situation and we need to help them navigate the best option in a world that is constantly evolving.

Introduction

For this 30th edition of the Yearbook of Medical Informatics, the topic of "Managing Pandemics with Health Informatics - Successes and Challenges" comes at dark times when massive populations face the spread of a mortal infectious disease (the coronavirus disease 2019 (COVID-19)) around the world. For clinicians, researchers, public health officials and political representatives, time is pressing to find solution to identify and care for vulnerable patients, to establish policies to slow down virus diffusion, and to monitor the impact of lock down policy, sanitary rules compliance, as well as vaccination campaigns.

Regarding the topic "management pandemics" recall back in 2009, Google Flu Trends (GFT), a digital disease detection system that uses the volume of selected Google search terms to predict influenza-like illnesses (ILI) activity, was identified by many as one of the first examples of how big data would transform traditional statistical predictive analysis [1]. Despite the development of these more refined and accurate predictive algorithms [2], Google announced in May 2015 that GFT would be discontinued and that their raw data would be made accessible to selected scientific teams. These initiatives have laid the foundation for crunching publicly available online-generated data for

public health purposes. Since the beginning of COVID19 pandemics, numerous open-source datasets and statistical online tools have emerged compiling and modelling by country the number of deaths and the number of hospitalized patients with moderate or severe clinical symptoms requiring intensive care. A recent study showed that the main source of data comes from social media and Internet search engines [3], raising the possibility of whether *social media listening* can be implemented in-real-time to support these predictive analyses on scale.

In this paper, we will explore the literature to explore the role of consumer informatics in the COVID-19 pandemics. Specifically,

we will identify how research in the past year have addressed: 1) innovative uses of social media listening for disease monitoring and public health purposes; 2) the lived experience of COVID as reported online; 3) the spread of 'fake information' and ways to determine quality in online information related to the pandemics; and 4) how people used telehealth and other technologies to maintain contact with their healthcare team.

2 Methodology

2.1 Search Strategy

We used PubMed to conduct our search, capturing papers on consumer-using technologies and emerging disease concerned published in the year 2020. The search strategy was based on the PICO framework ((P-Population/Problem, I–Intervention, C-Comparison, O-Outcome), where 'Population' refers to patient, client, consumer, 'Problem' refers to the COVID-19 disease and clinical symptoms, 'Intervention' refers to social media and information technology. 'Comparison' and 'Outcome were not included as we were looking for innovative usage and we did not want to constrain the query with prespecified comparison or outcome measures. We started from a core query adopted in previous work. Step by step we added 'coronavirus disease' related Medical Subject Headings terms or Supplementary Concepts. The syntax "[All Fields]" was used wherever possible to ensure our search strategy was comprehensive. The final search query is listed below:

((2020/1/1:2020/12/31[DP] NOT "Epub ahead of print") NOT Bibliography[pt] NOT Comment[pt] NOT Editorial[pt] NOT Letter[pt] NOT News[pt] NOT Case Reports[pt] NOT Published Erratum[pt] NOT Historical Article[pt] NOT Legal Case[pt] NOT legislation[pt] NOT ("review"[pt] OR "review literature as topic"[MeSH Terms] OR "literature review"[All Fields]))

AND ("patient"[All Fields] OR "client"[All Fields] OR "consumer"[All Fields] OR "consumer health informatics"[MeSH Terms] OR "consumer health information"[MeSH Terms] OR "consumer behavior"[MeSH Terms] OR "consumer

organizations"[MeSH Terms] OR "Direct-To-Consumer Screening and Testing"[MeSH Terms] OR "Direct-To-Consumer Advertising"[MeSH Terms])

AND ("covid 19"[All Fields] OR "covid 2019"[All Fields] OR "severe acute respiratory syndrome coronavirus 2"[Supplementary Concept] OR "severe acute respiratory syndrome coronavirus 2"[All Fields] OR "2019 ncov"[All Fields] OR "sars cov 2"[All Fields] OR "2019ncov"[All Fields] OR "wuhan"[All Fields]) AND ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields])

AND ("ehealth"[All Fields] OR "e-health"[All Fields] OR "tele-health"[All Fields] OR "mhealth"[All Fields] OR "mobile health"[All Fields] OR "telemedicine"[All Fields] OR "electronic patient-physician communication"[All Fields] OR "electronic medical record"[All Fields] "electronic health record"[All Fields] OR "personal health record"[All Fields] OR "electronic patient record"[All Fields] OR "online"[All Fields] OR "electronic"[All Fields] OR "information technology"[All Fields] OR "communication technology"[All Fields] OR "mobile"[All Fields] OR "online systems"[All Fields] OR "internet"[All Fields] OR "web"[All Fields] OR "website"[All Fields] OR "patient portal"[All Fields] OR "cell phone"[All Fields] OR "tele*"[All Fields] OR "digital*"[All Fields] OR "remote*"[All Fields] OR "video*"[All Fields] OR "virtual"[All Fields] OR "virtuality"[All Fields] OR "virtualization"[All Fields] OR "virtualized"[All Fields] OR "virtualizing"[All Fields] OR "virtuals"[All Fields])

AND ("social media"[All Fields] OR "facebook"[All Fields] OR "twitter"[All Fields] OR "youtube"[All Fields] OR "instagram"[All Fields] OR "social network site"[All Fields] OR "social web"[All Fields] OR "online social network"[All Fields] OR "social environment"[All Fields] OR "social process"[All Fields] OR "social competition"[All Fields] OR "social norm"[All Fields] OR "social feedback"[All Fields] OR "social influence"[All Fields] OR "social comparison"[All Fields] OR "social network"[All Fields] OR "discussion group"[All Fields] OR "support group"[All Fields] OR "social support"[All Fields] OR "community network"[All Fields] OR "online community"[All Fields])

2.2 Bibliometrics Analyses

To understand the state of the literature, we applied various bibliometrics tools onto the original set of articles returned from the search query. The "Bibliometrix" package from R [5] was used on the citation set of retrieved articles. We reported frequency of keywords. We illustrated the analysis of abstracts (measure of word frequency) by a word cloud drawing. We analysed keywords to uncover links between concepts through co-occurrences network. We also plotted a thematic map to analyse these clusters according to the quadrant in which they are placed [6]. Themes in the upper-right quadrant are both well-developed and important for the structuring of a research field. They are known as the motor-themes of the specialty given that they present strong centrality and high density. Themes in the upper-left quadrant have well-developed internal ties but unimportant external ties and so are of only marginal importance to the field. These themes are very specialized and peripheral in character. Themes in the lower-left quadrant are both weakly-developed and marginal. The themes of this quadrant have low density and low centrality, mainly representing either emerging or disappearing themes. Themes in the lower-right quadrant are important for a research field but are not as well-developed. Each theme is represented as a sphere, its volume being proportional to the number of documents associated with the theme.

3 Results

3.1 State of the Literature

A descriptive analysis of 147 articles was conducted, analysing the frequency of keywords, and the frequency of words in titles and abstracts. 703 distinct keywords were used, 640 distinct words in titles and 4302 words in abstracts. Figure 1 lists the 50 most cited keywords ranked according to decreasing frequency. Table 1 shows how the most cited keywords (10 or more occurrences) can be reorganised to describe the structure of the initial query.
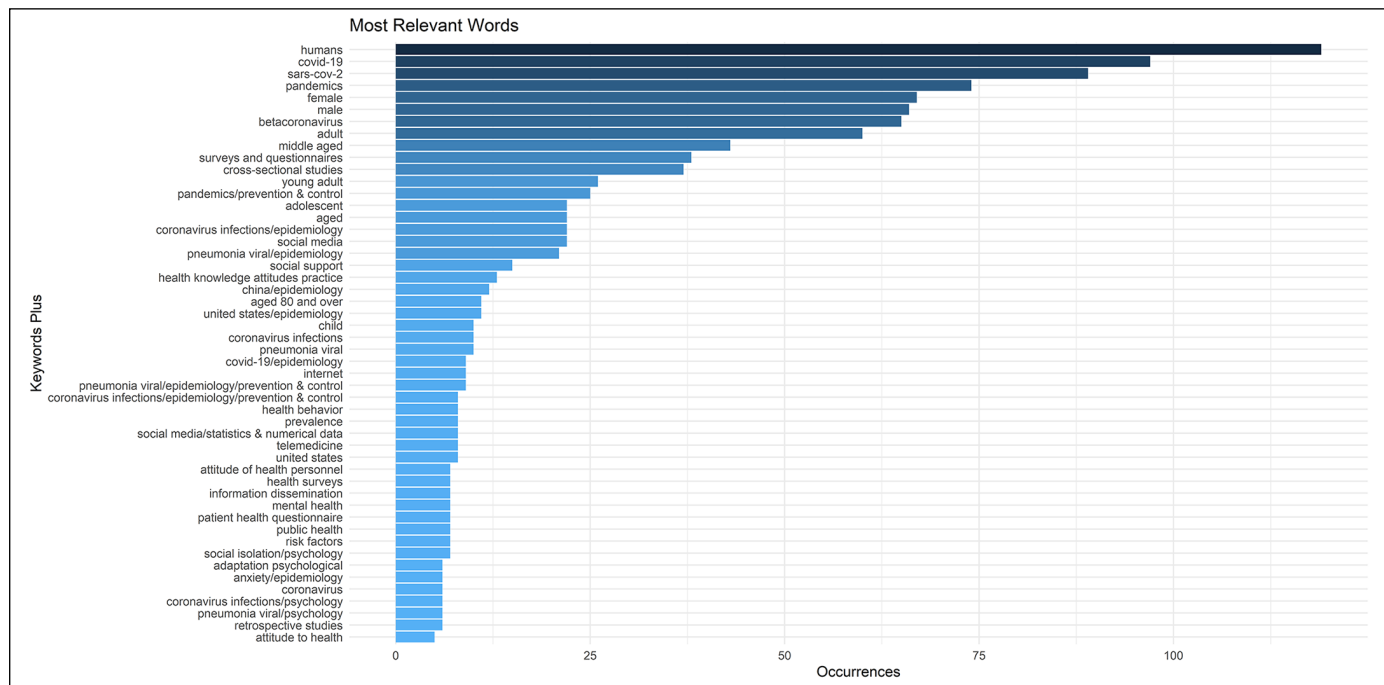
Fig. 1 Occurrences of keywords.

Table 1 Frequency of most cited keywords (10 or more occurrences) classified according to PICO search components

Words	Occurrences	
humans	119	
female	67	
male	66	
adult	60	
middle aged	43	Patient
young adult	26	
adolescent	22	
aged	22	
aged 80 and over	11	
child	10	
covid-19	97	
sars-cov-2	89	
betacoronavirus	65	Problem
pandemics	74	
coronavirus infections	10	
pneumonia viral	10	
surveys and questionnaires	38	
cross-sectional studies	37	Intervention
social media	22	
pandemics/prevention & control	25	
coronavirus infections/epidemiology	22	
pneumonia viral/epidemiology	21	
china/epidemiology	12	Outcome
united states/epidemiology	11	
Social support	15	
health knowledge attitudes practice	13	

Figure 2 shows a triangle-shaped words cloud created from the analysis of abstracts. Along the tip of the triangle, we can see keywords such as "coronavirus, study, pandemic, health, covid"; and along the base of the triangle, we can see keywords such as "social media, patients" and "associated factors and risk reported". Regarding the conceptual structure of the set of 147 articles, Figure 3 shows the co-occurrences of keywords. It identifies the relationships between keywords in accordance to the groups reported in Figure 2. Regarding thematic maps of keywords found in Figure 4, clusters according to centrality (relevance degree) and density (development degree) are reported in each quadrant. Most common themes found across papers are related to covid infection as a human disease. Next set of frequently reported themes focus on "epidemiology, prevention and control". Niche themes revealed by this analysis are represented by two clusters: first, mobile applications and covid epidemiology, second, clinical epidemiology studies to describe the severity of illness. The cluster "attitude to health" is in central and neutral position.

Fig. 2 Triangle-shaped cloud of the 50 most frequent cited words in abstracts.

3.2 Best Paper Selection

To identify the 15 candidate papers, co-editors independently assessed the 147 retrieved papers using the Rayyan web-tool [4], followed by discussion. Elements that were considered in the screening decision include: 1) level of relevance regarding the 2021 Yearbook topic "Managing Pandemics with Health Informatics"; 2) whether the study was focused only on patients and consumers; 3) nature of the issues addressed; and 4) level of innovative approach. Section co-editors' agreement was measured with Cohen's kappa coefficient: κ=0.474 95%CI [0.098 ; 0.7168] (moderate agreement). The 15 articles were then presented to a panel of international experts for full paper review and scoring according to the IMIA Yearbook best paper selection process. The final selection of three best papers is completed after discussions at the annual IMIA Yearbook board meeting.

Figure 5 shows the factorial map of papers, based on words in abstracts (400 terms), revealing four clusters. The result clustering explains 77.04% of the dispersion. This clustering was blinded to the reviewers. Cluster 1 is composed of articles relating to the use of mobile apps for video consultation with healthcare professionals. Sabrir *et al.* [7] investigated the effects of WhatsApp video consultation on patient admission and discharge times in comparison to bedside consultation during the pandemic. Consultation via WhatsApp was reported to have reduced both contact time with patients with COVID-19 and that the number of medical staffs contacting patients. Furthermore, the study reported reducing length of stay in the emergency department. Xu *et al.* [8] built a system based on a popular social media smartphone app called WeChat; the app was used to establish two-way communication between a multidisciplinary team with home-quarantined

individuals including those diagnosed with COVID-19. The system requires few staff to manage a large cohort of patients. They found this telemedicine system reduces the risks of delayed hospitalisation due to better management of disease progression, optimising usage of resources, and preventing cross-infections among medical workers and patients.

Cluster 2 is composed of studies reporting the lived experience of patients and health workers during the pandemic [9-13]. El-Awaisi *et al.* [9] used social listening technique to explore unfiltered public perceptions during the COVID-19 pandemic and to elaborate on the emotional reactions in response to an online social media post about healthcare professionals. The post asked who do you think works in hospitals versus who really works in hospitals? Despite the word 'Heroes' being commonly used, the post brought considerable attention to the role of the interprofessional team and generated
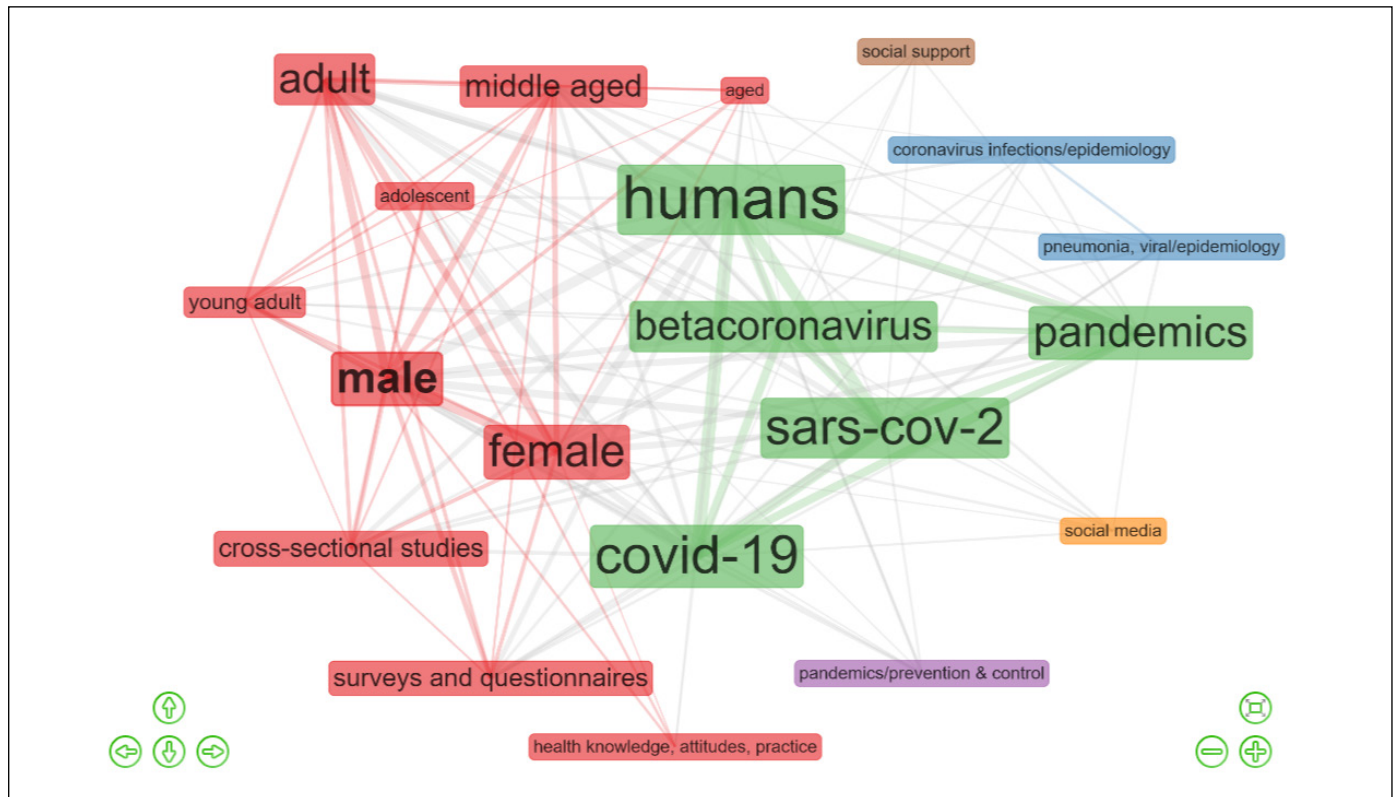
Fig. 3 Co-occurrences network of the 20 most frequent keywords

many feelings of frustration and exclusion. Bennett *et al.* [10] study aimed to gain better understanding of the experiences and concerns of front-line National Health Service (NHS) workers while caring for patients with COVID-19. Key findings highlighted that while healthcare workers shared intensely positive experiences, caring for COVID-19 patients brought a significant emotional toll, and strained relationships between immediate front-line staff, their families, management and even government. There was a sense that at the beginning of the pandemic, staff were driven by adrenalin and optimism; but over time this has dissipated and is replaced by exhaustion, numbness, and dreaded expectation of a 'second wave'. Healthcare workers could reasonably be considered as 'second victims' of COVID-19.

Cluster 3 composed of articles exploring how people used the internet to seek health information during the pandemic [14-18] and the dissemination of fake news [19]. Zhao *et*

al. [15] found that a high proportion of Chinese female users were seeking health information and help for their parents or for older adults at home. The most searched information included accessing medical treatment, managing self-quarantine, as well as offline to online support. Li *et al.* [16] explored the predictive role of social media use on public preventive behaviors in China during the COVID-19 pandemic. They also found that different social media types differed in predicting an individual's preventive behaviors during COVID-19.

Cluster 4 composed of articles reporting the use of social listening methods (e.g., tweet hashtags) to explore the spread of the virus around the world (i.e., infoveillance) [20-21]. See Appendix for more information.

Finally, three papers were selected to be best papers after discussions at a consensus meeting at the April 30th 2021 IMIA Yearbook editorial meeting. They are listed in table 2. See Appendix to read the description of the studies and the main results.

4 Conclusions

The pandemic outbreak of the novel coronavirus (COVID-19) constitutes a grave risk to the global community and sparks a significant increase in public interest and media coverage, especially on social media. Consumers are facing a new set of challenges that were not conceived before COVID-19, often finding themselves in a world that is constantly changing, blended with facts and fake information, and many consumers unable to decide what to do next. Despite most people understanding the good will behind public health policies, one must not forget it is *individuals* we are supporting and that their personal circumstances may affect how they perceive and comply with these policies. Consumers more than ever need help to make sense of the uncertainty and their situation and help them navigate the best option in a world that is constantly evolving.
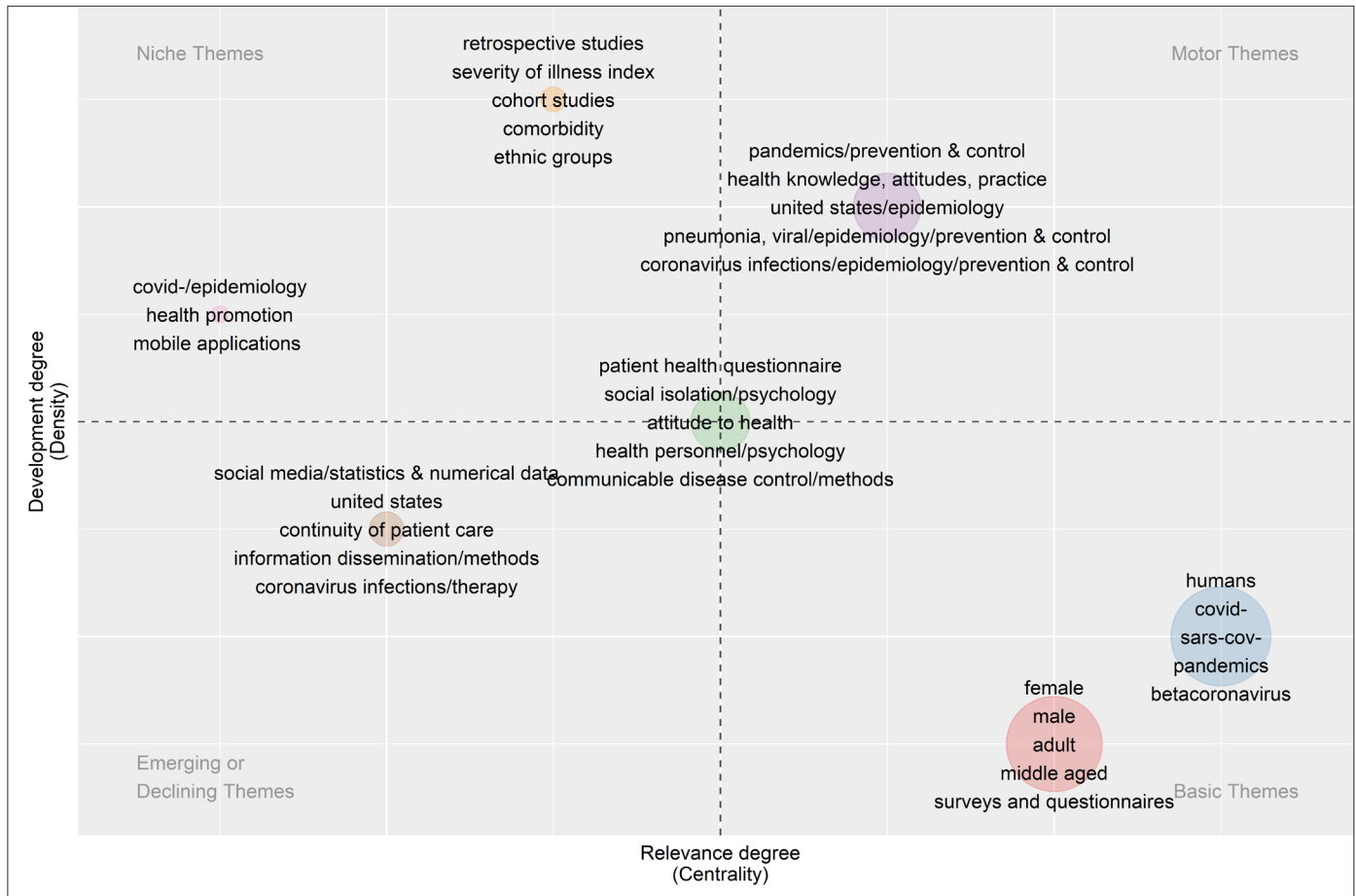
Fig 4 Thematic map of clusters of keywords.

References

1. Helft M (November 11, 2008) Google uses searches to track flu's spread. NY Times. Available at www.nytimes.com/2008/11/12/technology/internet/12flu.html?_r=0#. Accessed May 17, 2021.

2. Yang S, Santillana M, Koua SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473–8.

3. Corsi A, de Souza FF, Pagani RN, Kovaleski JL. Big data analytics as a tool for fighting pandemics: a systematic review of literature. J Ambient Intell Humaniz Comput 2020;1-18.

4. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarid A. Rayyan - a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210.

5. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics 2017;11(4):959-75.

6. Callon M, Courtial JP, Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research—the case of polymer chemistry. Scientometrics 1991;22(1):155–205.

7. Sabırlı R, Karsli E, Canacik O, Ercin D, Çiftçi H, Sahin L, et al. Use of WhatsApp for Polyclinic Consultation of Suspected Patients With COVID-19: Retrospective Case Control Study. JMIR Mhealth Uhealth 2020;8:e22874.

8. Xu H, Huang S, Qiu C, Liu S, Deng J, Jiao B, et al. Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study. J Med Internet Res 2020;22:e19514.

9. El-Awaisi A, O'Carroll V, Koraysh S, Koummich S, Huber M. Perceptions of who is in the healthcare team? A content analysis of social media posts during COVID-19 pandemic. J Interprof Care 2020;34:622-32.

10. Bennett P, Noble S, Johnston S, Jones D, Hunter R. COVID-19 confessions: a qualitative exploration of healthcare workers experiences of working with COVID-19. BMJ Open 2020;10:e043949.

11. Hamlin M, Steingrimsson S, Cohen I, Bero V, Bar-Tl A, Adini B. Attitudes of the Public to Receiving Medical Care during Emergencies through Remote Physician-Patient Communications. Int J Environ Res Public Health 2020;17(14):5236.

12. Xie B, Charness N, Fingerman K, Kaye J, Kim MT, Khurshid A. When Going Digital Becomes a Necessity: Ensuring Older Adults' Needs for Information, Services, and Social Inclusion During COVID-19. J Aging Soc Policy 2020;32:460-70.

13. Picone M, Inoue S, DeFelice C, Naujokas MF, Sinrod J, Cruz VA, et al. Social Listening as a Rapid Approach to Collecting and Analyzing COVID-19 Symptoms and Disease Natural Histories Reported by Large Numbers of Individuals. Popul Health Manag 2020;23:350-60.

14. 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. JMIR Public Health Surveill 2020;6(4):e21071.

15. Zhao X, Fan J, Basnyat I, Hu B. Online Health Information Seeking Using "#COVID-19 Patient Seeking Help" on Weibo in Wuhan, China: Descriptive Study. J Med Internet Res 2020;22(10):e22910.

16. Li X, Liu Q. Social Media Use, eHealth Literacy, Disease Knowledge, and Preventive Behaviors in the COVID-19 Pandemic: Cross-Sectional Study on Chinese Netizens. J Med Internet Res 2020;22(10):e19684.

17. Brar Prayaga R, Prayaga RS. Mobile Fotonovelas Within a Text Message Outreach: An Innovative
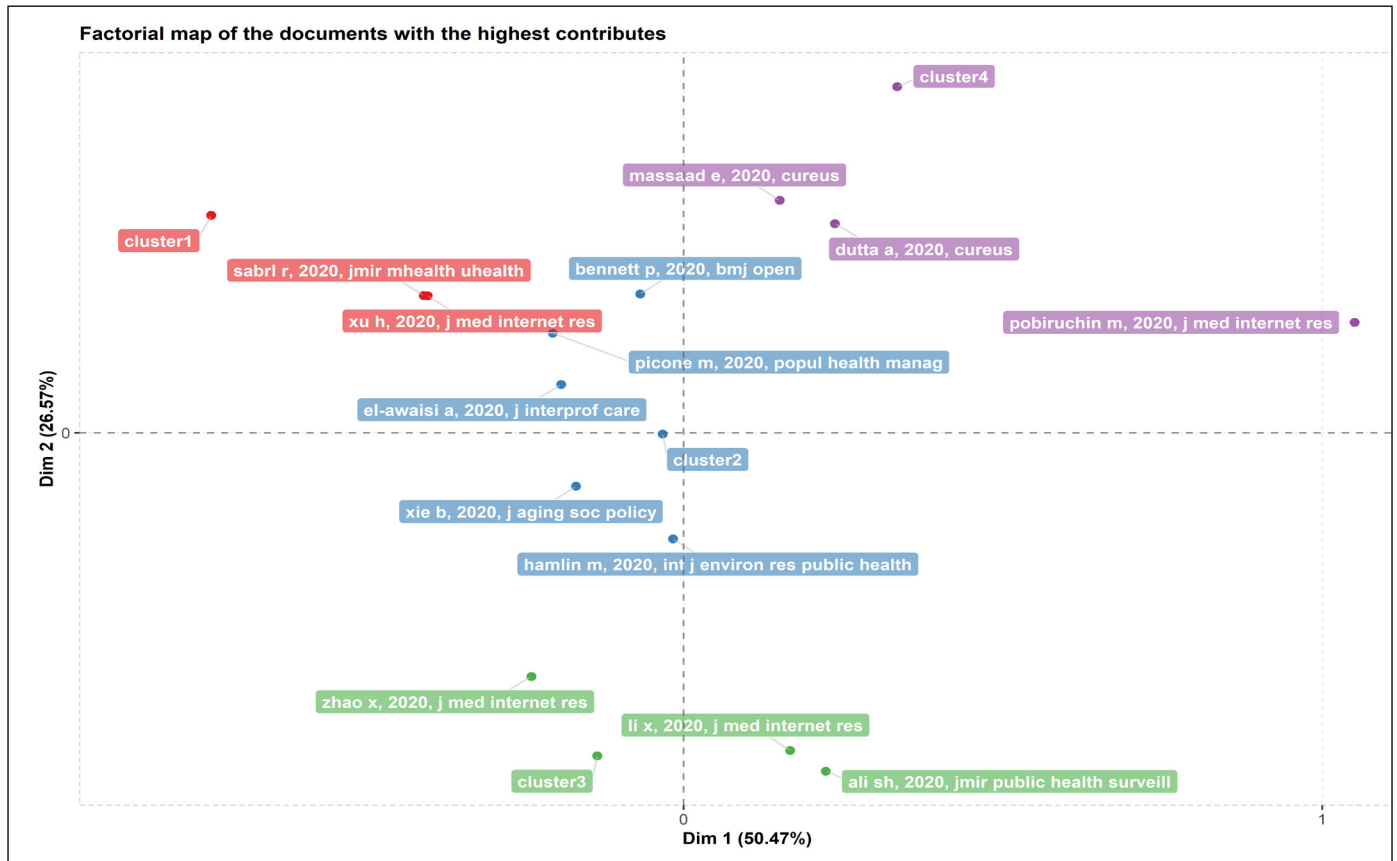
Factorial map of the documents with the highest contributes

Fig. 5 Factorial map of papers, based on words in abstracts (multiple correspondence analysis based on 400 abstract terms).

Table 2 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2021 in the special section 'Consumer Health Informatics'. The articles are listed in alphabetical order of the first author's surname.

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. JMIR Public Health Surveill 2020;6(4):e21071.
▪ 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. Popul Health Manag 2020;23(5):350-60.
▪ 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. J Med Internet Res 2020;22(8):e19629.

Tool to Build Health Literacy and Influence Behaviors in Response to the COVID-19 Pandemic. JMIR Mhealth Uhealth 2020;8(8):e19529.

18. Dutta A, Beriwal N, Van Breugel LM, Sachdeva S, Barman B, Saikia H, et al. YouTube as a Source of Medical and Epidemiological Information During COVID-19 Pandemic: A Cross-Sectional Study of Content Across Six Languages Around the Globe. Cureus 2020;12(6):e8622.

19. Moscadelli A, Albora G, Biamonte MA, Giorgetti D, Innocenzio M, Paoli S, et al. Fake News and Covid-19 in Italy: Results of a Quantitative Observational Study. Int J Environ Res Public Health 2020;17(16):5850.

20. Massaad E, Cherfan P. Social Media Data Analytics on Telehealth During the COVID-19 Pandemic. Cureus 2020;12(4):e7838.

21. 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. J Med Internet Res 2020;22(8):e19629.

Correspondence to:
Pascal Staccini
Département IRIS - URE RETINES
Faculté de Médecine
28 av de Valombrose - Nice, France
E-mail: pascal.staccini@univ-cotedazur.fr

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.