Journal of Multidisciplinary Healthcare Dovepress

ORIGINAL RESEARCH

Mapping Knowledge Landscapes and Emerging Trends for the Spread of Health-Related Misinformation During the COVID-19 on Chinese and English Social Media: A Comparative Bibliometric and Visualization Analysis

Yunfan He[1,2], Jun Liang[3–5], Wenguang Fu[6], Yongcheng Liu[7], Fangyu Yang[8], Shunjing Ding[9], Jianbo Lei[10–12]

[1]School of International Relations and Public Affairs, Fudan University, Shanghai, People's Republic of China; [2]National Institute of Intelligent Evaluation and Governance, Fudan University, Shanghai, People's Republic of China; [3]Department of AI and IT, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; [4]Center for Health Policy Studies, School of Public Health, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; [5]Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; [6]Department of Hepatobiliary Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou, People's Republic of China; [7]Netease Group, Hangzhou, Zhejiang, People's Republic of China; [8]School of Nursing, Capital Medical University, Beijing, People's Republic of China; [9]Department of Cardiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, People's Republic of China; [10]Clinical Research Center, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, People's Republic of China; [11]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, Sichuan, People's Republic of China; [12]Center for Medical Informatics, Health Science Center, Peking University, Beijing, People's Republic of China

Correspondence: Jianbo Lei, Center for Medical Informatics, Peking University, Beijing, 100191, People's Republic of China, Tel +86 (10) 8280-5901, Fax +86 (10) 8280-5900, Email jblei@hsc.pku.edu.cn; Shunjing Ding, Department of Cardiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, People's Republic of China, Email dsj6890@sina.com

Background: Online health-related misinformation poses a serious threat to public health. As the coronavirus disease 2019 (COVID-19) pandemic aggravated the spread of misinformation regarding COVID-19, relevant research has surged.

Objective: To systematically summarize Chinese and English articles regarding health-related misinformation about COVID-19 on social media and quantitatively describe research progress.

Methods: Using bibliometrics, we systematically analyzed and compared the characteristics of scientific articles in English and Chinese, examining article numbers, journals, authors, countries, institutions, funding, and research topics, and compared changes in popular research topics.

Results: This study analyzed 1,294 articles, revealing a significant increase in article numbers and citations during the COVID-19 pandemic (1.94 times and 2.95 times, respectively, compared to pre-pandemic data). However, high-impact articles were scarce and the field lacked a core group of authors and collaborative networks. China had the largest number of papers (n=266) and funds (n=292), but articles in English exceeded by far those in Chinese (1,131 vs 163, respectively). Regarding article topics, the transformation from qualitative small-data analyses to quantitative empirical big-data research has been realized.

Conclusion: With the maturity of natural language processing technology, in-depth mining of massive user-generated content has become a hot spot. The outbreak of the COVID-19 pandemic has prompted the research focus to shift from misinformation-related health problems to social problems involving the sources, content, channels, audiences, and effects of communication networks. Using artificial intelligence technology like machine learning to deeply mine large amounts of user-generated content on social media will be a future research hot spot.

Keywords: misinformation, online health information, social media, COVID-19, Chinese, English

Received: 7 November 2024
Accepted: 9 December 2024
Published: 19 December 2024

Introduction

The emergence of social media, especially the rise of user-generated content, has reduced the cost of publishing information[1] but has intensified the phenomenon of online health-related misinformation. Misinformation encompasses inaccurate content, including both deliberately fabricated disinformation and unintentional falsehoods.[2] Social media has become the main channel for obtaining health information,[3] and more than 70% of people in China and worldwide inquire about health- and disease-related information through online platforms.[4,5] During the coronavirus disease 2019 (COVID-19) pandemic, Twitter, Facebook, Weibo, and other social media platforms have become channels to retrieve health information, with the highest usage rates among Chinese- and English-speaking citizens.[6,7] The rise of user-generated content models[8–12] and the absence of regulatory mechanisms have led to widespread health-related misinformation on social media, and the outbreak of the COVID-19 pandemic has further aggravated this trend.

During the COVID-19 pandemic, mass dissemination of health-related misinformation on the internet caused serious health and social problems and may even threaten national security, which has become an important object of public health research.[13] The global dissemination of health-related misinformation exceeds even the pace of the pandemic itself,[14] forming infodemics, which refers to the abundance of accurate and inaccurate information that makes it difficult for people to find trustworthy sources of information and reliable guidance, while the discipline of infodemiology behind infodemics was proposed by Eysenbach, the chief editor of JMIR, early in 2002.[15,16] This poses a significant threat to people's health, seriously harming personal health management[17,18] and public health governance.[19] For example, mass dissemination of COVID-19-related misinformation caused people to misrepresent scientific evidence of epidemic prevention, causing adverse psychological reactions like anxiety, depression, and panic, increasing hate speech, leading to improper allocation of health resources, and increasing vaccination hesitancy.[20,21] As an important means of information warfare, online health information has gradually become one of the most important guarantors of national defense security.[22] Therefore, a growing number of scholars worldwide are paying attention to the spread of COVID-19 misinformation on social media, and related research is rapidly increasing.

Many studies have explored the dissemination of health-related misinformation on social media from different perspectives, and several reviews and bibliometric analyses have summarized the current research status. Regarding topic distribution, Lu[23] applied the Crisis and Emergency Risk and Communication model to investigate the topics and trends of COVID-19 misinformation in China, identifying five major misinformation topics. For a systematic review, Wang et al[24] conducted a systematic review of the characteristics and potential drivers of health-related misinformation by summarizing methodological and empirical articles on the dissemination mechanisms of health-related misinformation. For bibliometric analysis, Pool et al[25] revealed the main nodes and future directions of infodemic-related research through literature concept maps. Wang et al[26] quantitatively analyzed 5666 disinformation articles using Derwent Data Analyzer to explore the forward directions in the field of disinformation. Yeung et al[27] used bibliometrics to systematically summarize the research progress of related articles in the United States and Europe, revealing potential differences in the dissemination of health-related misinformation on social media across different countries. However, there is currently a lack of comparative analysis of Chinese social media platforms.

These studies on health-related misinformation on the one hand have limitations in terms of few retrieval databases, irregular data screening processes, and incomplete analysis areas and angles, on the other hand neglect to address the health-related misinformation during the COVID-19 pandemic and China's research. Regarding data sources, the databases selected in previous studies are insufficient, relying solely on PubMed, Embase, and Web of Science,[28] the recall rate cannot be guaranteed, and the retrieval time is mostly before the COVID-19 pandemic, which cannot accurately reflect the latest research progress. Regarding data screening, some studies did not provide detailed inclusion and exclusion criteria, lacked a description of the screening process, and could not guarantee the accuracy of included articles. Previous research was also limited to the analysis of English articles and lacked data from Chinese publications, where COVID-19 first broke out. Moreover, there is a lack of comparative analysis of research progress before and during COVID-19, and in different countries.

Therefore, this study aimed to summarize Chinese and English articles related to the dissemination of COVID-19 misinformation as of December 27, 2022, use bibliometrics to describe the latest research progress and topic trends in this field, and compare the differences in research between China and other countries, addressing the lack of quantitative analyses in traditional systematic reviews and the lack of research on Chinese social media. This will help the academic community identify the core journals, best collaborators, and research flaws in this field and discover research differences before and after the COVID-19 pandemic. It also provides information for responses to public health crises during an epidemic and for the formulation of policies for normal health education and health promotion. Meanwhile, given that some Chinese authors may publish research in English, the comparison between China and other countries may not necessarily accurately identify the differences, but at least it can provide us with some insights into the potential differences in research on the dissemination of health-related misinformation on social media in different countries, not to mention that we have merged Chinese and English papers on related topics in China.

Materials and Methods

Ethical Statement

All data in this study are sourced from publicly available literature databases and this study does not involve any human experiments. There are no ethical or moral issues involved.

Data Sources and Retrieval Strategies

To cover comprehensive and complete articles, we searched seven international literature databases (two in the field of comprehensive science: Web of Science Core Collection and Scopus; three in the field of medical health: PubMed, Embase, and Cochrane; one in the field of computer science: IEEE; and one in the field of psychology: PsycINFO) and three Chinese literature databases (CNKI, Wanfang, and VIP) to ensure the recall rate. We used keywords like "COVID-19 pandemic", "health-related misinformation", and "social media"[24] to construct a search formula (see Appendix Table 1 for this formula and the search results from different databases) and searched the above 10 databases to obtain articles about the spread of health-related misinformation on social media during the COVID-19 pandemic until December 27, 2022. During the search process, we have excluded non Chinese or English literature through the secondary search function of the database. Of the 5,971 retrieved articles, 3234 remained after excluding duplicates.

Data Filtering

This study strictly followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework[29] to complete the screening and analysis of articles (As this study is a bibliometric analysis of literature, the review was not registered). Based on the protocol of previous research,[24] we formulated inclusion and exclusion criteria (please refer to the end of this paragraph for Inclusion and Exclusion Criteria). To ensure the consistency of screening results, two experts in the field of medical informatics independently screened the same 30 randomly selected articles according to these inclusion and exclusion criteria. Their evaluation showed strong consistency (kappa=0.93). Then, the same two experts screened the remaining articles to ensure that the included articles were closely related to the spread of COVID-19 misinformation on social media. We excluded 47 articles that were classified as reviews, letters, editorial, and comments. Additionally we excluded 42 articles not related to the COVID-19 pandemic, 4 articles with unavailable full text, 1,690 articles where the research subject is not misinformation, 100 articles lacking analytical aspect on information dissemination, 48 articles where the information transmission channel is not social media, and 9 articles not related to health, medical, or public. Finally, 1,294 relevant articles were included (Figure 1).

Inclusion Criteria:

1. Include the articles exploring the period of the COVID-19 pandemic;
2. Include the articles related to misinformation, disinformation, fake news, rumor or misinformation of any kind;
3. The misinformation involved in the included articles must be disseminated through online social media;
4. Incorporate articles must relate to health, medicine, disease, treatment or public health;
5. Include model and empirical analysis of the distribution, spread, or impact of misinformation.

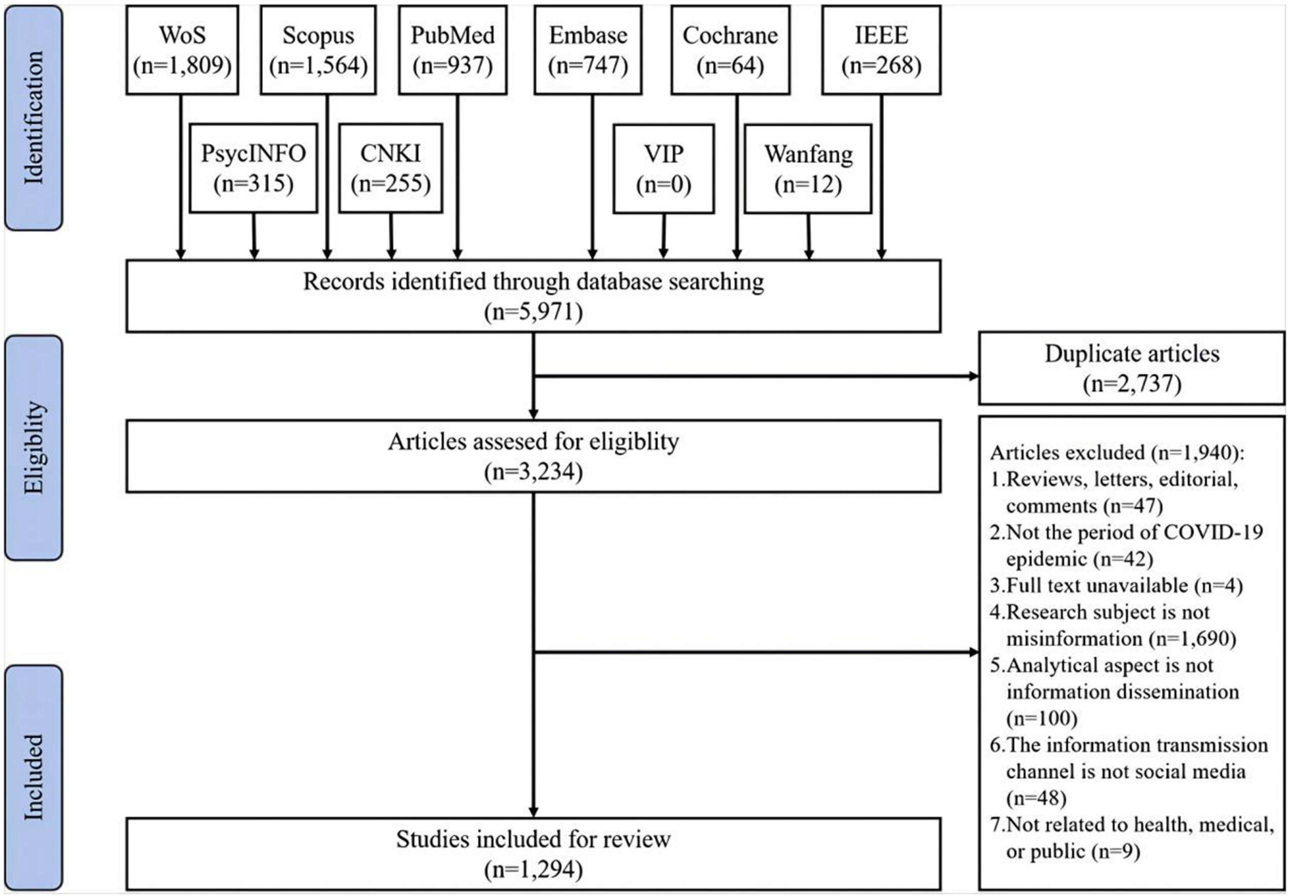

Figure 1 Flowchart of the article screening process.

Exclusion Criteria:

1. Exclude summaries, reviews, letters, errata and comments;
2. Exclude the articles whose period was not during the COVID-19 pandemic;
3. Exclude the articles with incomplete information and full text is not available;
4. Exclude the articles in which misinformation was not the main object of study;
5. Exclude the articles whose main research content is not information dissemination;
6. Exclude the articles whose information dissemination channel is not social media;
7. Articles not related to health, medicine or public health were excluded.

Data Extraction and Analysis

First, we extracted the complete bibliographic information of the included articles, including title, abstract, keywords, authors, author institutions (the authors' name was disambiguated through the author institutions), publication time, journal, and funds, and manually translated the bibliographic information of the Chinese articles into English. Then, we used Python code to unify the exported bibliographic information from different databases into the standard format of bibliometrix. Next, we used bibliometrics to quantitatively analyze the characteristics of the included articles regarding the number of articles, core journals, author influence, institution distribution, as well as funding and topic distribution, to understand the research status and trends in this field and explore differences between Chinese and English articles. Python was used to process data, the bibliometrix package[30] of R was used to analyze the articles, and the ggplot2 package[31] was used to draw graphs. Finally, VOSviewer[32] was used for network visualization.

For topic distribution, we used Python to process keywords, which involved unifying English capitalization, removing stop words, and lemmatization. The high-frequency keyword threshold was calculated as 7 according to the 80/20 rule,[33] and the co-occurrence network of high-frequency keywords was generated using VOSviewer software.

Results

Trend Analysis of the Number of Articles

The number of health-related misinformation articles published in only 3 years during the COVID-19 pandemic (n=1,294) far exceeds that in the 10 years before the pandemic (n=530).[27] The first relevant articles were published in February 2020, and the number of articles continued to increase every quarter afterward (average quarterly growth rate, 41.30%, and average annual growth rate, 44.99%), reaching its peak in the first quarter of 2022 (n=155; Figure 2). Their citation numbers rapidly increased starting in the first quarter of 2020 and reaching their peak in the fourth quarter of 2020 (n=3,864), with an average of 1,474 citations per quarter and 5,895 per year (Figure 3). This field is dominated by articles in English, and the numbers of publications (1,131 vs 163) and citations (16,865 vs 821) exceed by far those of articles in Chinese.

Core Journal Analysis

The core journals of analyzed articles are mostly English journals indexed by the Science Citation Index (SCI)/Social Sciences Citation Index (SSCI), mainly from the fields of Public, Environmental & Occupational Health and Medical Informatics, with high average quality but a lack of high-impact journals. According to Bradford's law and related inferences in bibliometrics,[34] the number of core journals accounts for about 1/31 of the total number of journals. Considering the 22 core journals in this field (Appendix Table 2), the relevant articles in these journals account for 26.66% of the total article number. Among them, only three are Chinese journals, and their article and citation numbers are far lower than those of English journals. Among all core journals, the Journal of Medical Internet Research had the most articles (n=64) and citations (n=2,106). When comparing the numbers of publications (Figure 4) and citations (Figure 5) per quarter of the top 5 journals with the most articles, the publication and citation growth rates of the Journal of Medical Internet Research and the International Journal of Environmental Research and Public Health outpaced those of other journals, reaching their peaks in a short time in 2020.

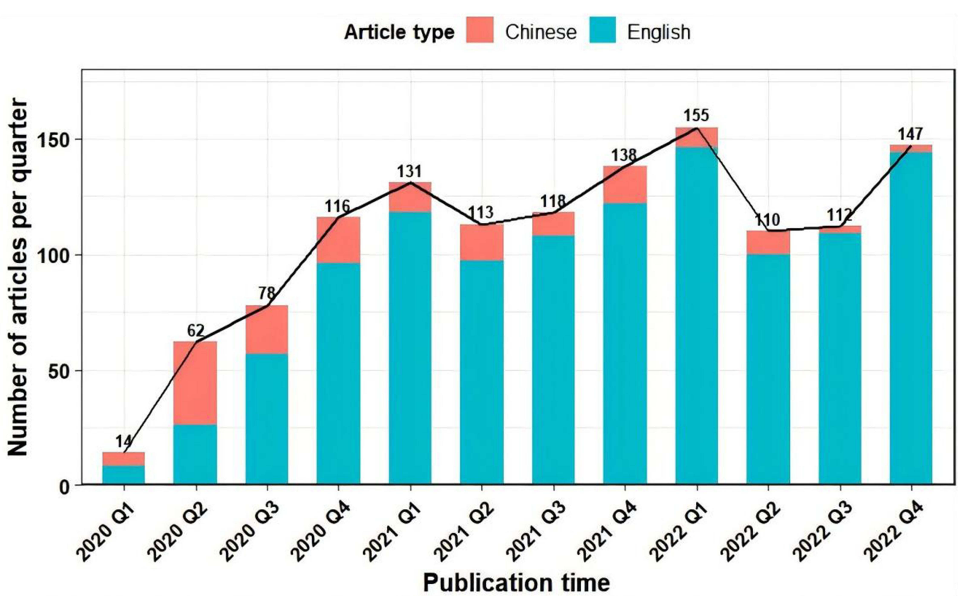

Figure 2 Number of scientific publications per quarter.
Note: Quarter 1 (Q1) refers to January–March, quarter 2 (Q2) to April–June, quarter 3 (Q3) to July–September, and quarter 4 (Q4) to October–December of a year.

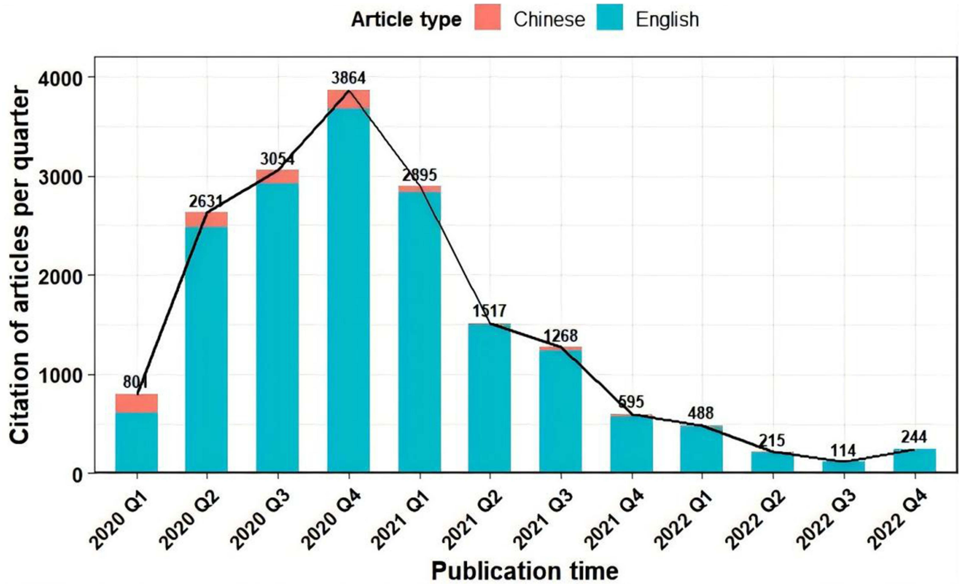

Figure 3 Citation of the scientific production per quarter.

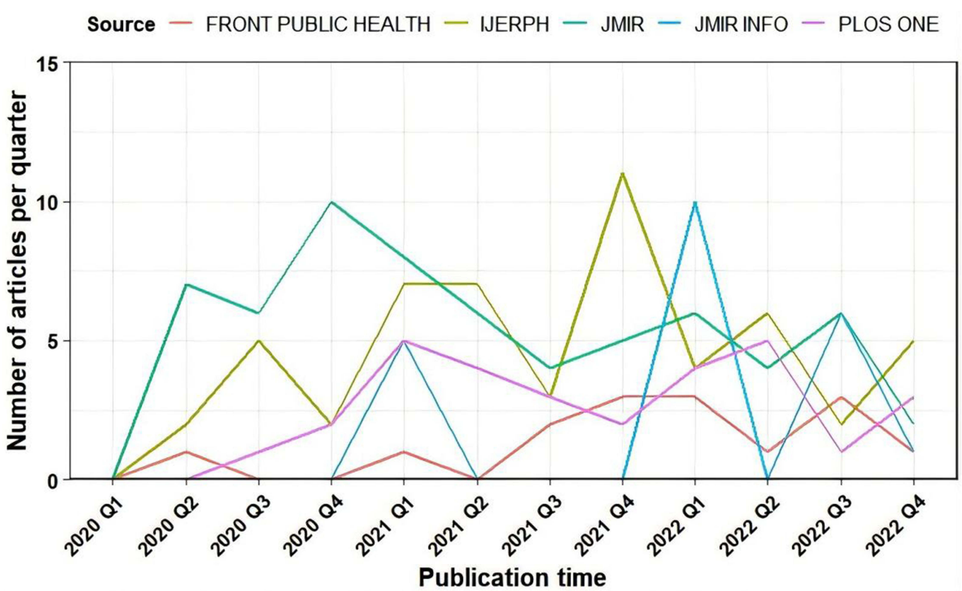

Figure 4 Number of articles per quarter published by the top 5 core journals.

Core Authors Analysis

Of the 4,682 authors in the field, 2,111 were first or corresponding authors. The publication numbers of highly productive authors were generally low, their influence was relatively insufficient, and core author groups and cooperation networks were lacking. Based on Lotka's Law,[35] authors whose publication number exceeds 0.749 times the square root of that of the most productive scientist are considered highly productive authors. Of the 323 highly productive authors, 69 were Chinese authors, and the publications by highly productive authors account for 55.87% of all articles. Based on the numbers of articles and citations, we created a trend chart of the annual publications of the top 10 authors (Figure 6). Most (70%) of these authors have been engaged in related research since 2020, the first year of the COVID-19 pandemic. However, even the most productive author accumulated only 5 articles, and the 3 authors with the most citations are from the same research team; their jointly published four articles were cited 548 times. No obvious cooperative network was

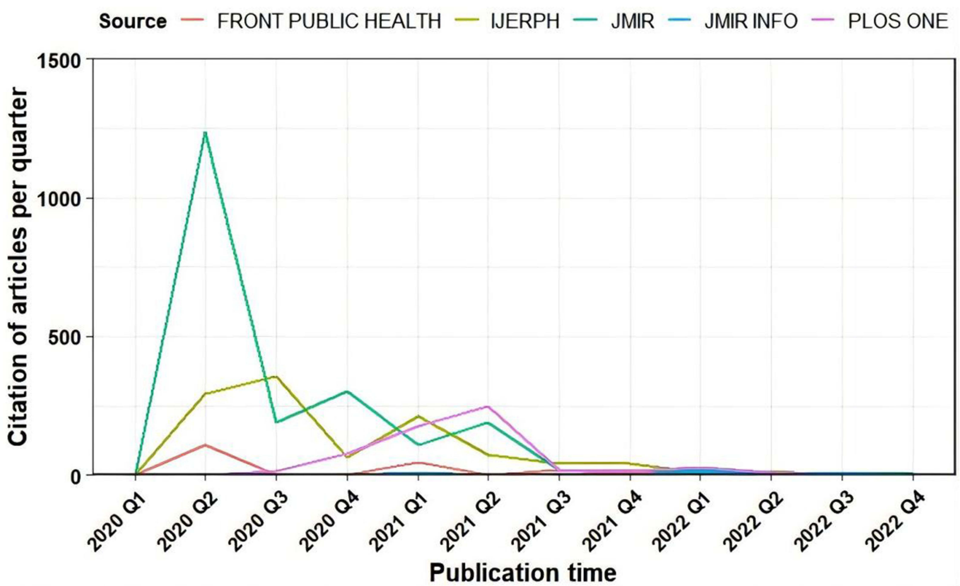

Figure 5 Comparison of the citation frequency of the top 5 core journals per quarter.

Figure 6 Publishing trends of the top 10 authors. Marker size and color shade indicate the annual numbers of publications and citations, respectively.

built among high-yield authors, which shows that the field lacks high-volume, highly cited experts and core author groups, as well as cooperation among experts in different fields.

Analysis of Cooperation Between Countries and Institutions

China and the United States were the countries with the largest numbers of articles in this field (266 and 226, respectively), with relatively close cooperation between these countries. The University of Jordan and King Saud University in the Middle East are the institutions that have published the most articles (29 and 20, respectively). Academic cooperation between international universities is the focus of institutional cooperation. Regarding the country of origin, 4,682 authors came from 76 countries. The distributions of the number of individuals and the cooperative publications of the top 10 countries are shown in Figure 7 China has the most related research (n=266, 20.56%), followed by the United States, the United Kingdom, and Canada. The top 5 institutions are the University of Jordan (n=29, 2.24%), Wuhan University (n=21, 1.62%), King Saud

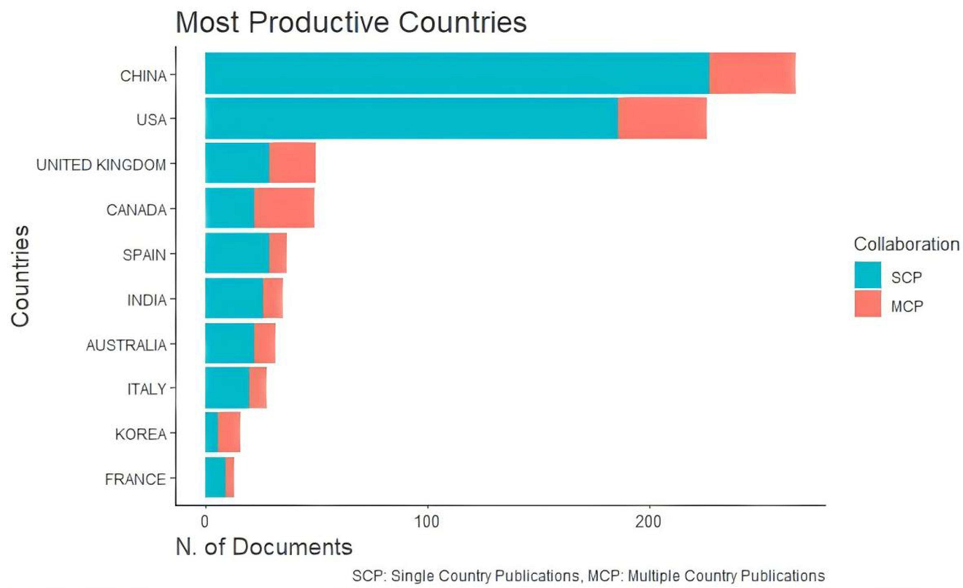

Figure 7 Distribution of the number of individual and cooperative publication of the top 10 high-yield countries.

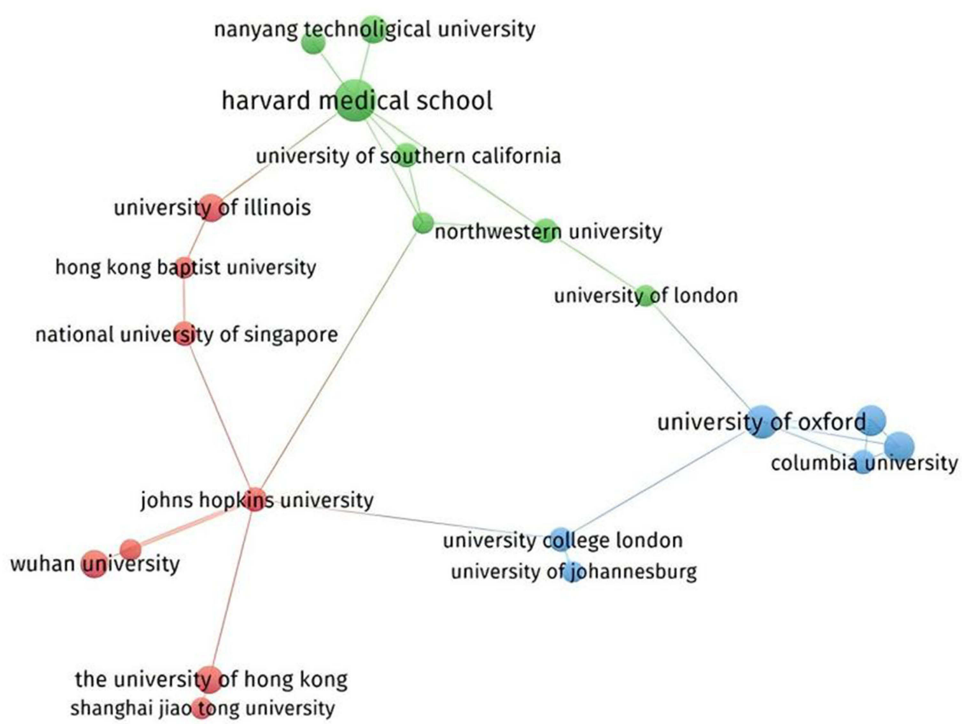

Figure 8 Cooperation networks of research institutions.

University (n=20, 1.55%), University of Pennsylvania (n=19, 1.47%), and University of Oxford (n=16, 1.24%). Based on the institutional data of co-authors, the network map (Figure 8) demonstrated that academic cooperation among international universities is relatively close. The American cluster is dominated by Harvard Medical School, the British cluster by the University of Oxford, and the Chinese cluster by Wuhan University.

Funding Analysis

The funding rate of the included articles is 36.50%. China and the United States are the main funding countries, and most of the funds come from the National Social/Natural Science Foundation of the People's Republic of China and the USA National Institutes of Health. A total of 462 included articles received 683 grants, and the funding rate of related research was 35.70% (of which the funding rate of Chinese articles was 42.94%), with an average of 1.48 grants per paper. For the source countries of funds, a total of 46 countries/country groups have provided funding for related research, among which China (n=292) and the United States (n=147) funded the most project funds, far exceeding other countries. For source institutions of funds, a total of 278 institutions have provided fund support, and 9 institutions have no less than 10 funds (see Appendix Table 3). Among them, National Natural Science Foundation of the People's Republic of China have the most funds (n=50), followed by National Social Science Foundation of the People's Republic of China (n=48), and USA National Institutes of Health (n=39). In addition, among the 292 funding projects in China, 28 (9.59%) are national key research and development projects, covering medical, humanities, military and other fields, with a total funding amount of more than 20 million US dollars.

Hot Topic Analysis

Topics were analyzed in three dimensions: high-frequency keyword clustering, thematic map,[36] and thematic evolution.[37] First, topic clustering was carried out for high-frequency keywords, and distributions and relationships among topics were analyzed using co-occurrence networks. Second, the topics were clustered and mapped according to density and centrality to explore current research hot spots. Finally, we analyzed the trends in research topics over time.

High Frequency Keywords Cluster Analysis

According to the co-occurrence network, the research topics were divided into four parts: social media source, data mining technology, research object, and research field (Figure 9). Among them, the red category represents social media sources (with 15 keywords), such as Twitter, YouTube, Facebook, and Sina Microblog, and English social media

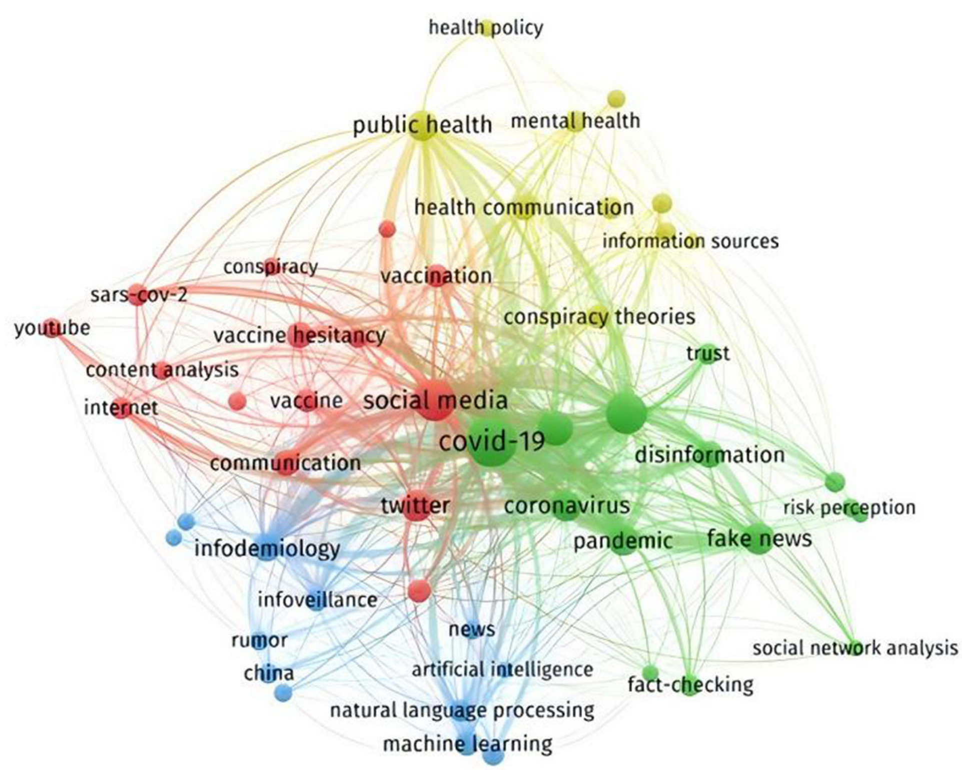

Figure 9 High-frequency keywords co-occurrence network.

https://doi.org/10.2147/JMDH.S501067

platforms are currently the focus of researchers' attention. The blue category (12 keywords) represents data mining technology, including artificial intelligence, natural language processing, and machine learning; the green category (13 keywords) represents the research object, such as coronavirus, fake news, disinformation, and social networks, and text data is currently the main data type that research focuses on; the yellow category (10 keywords) represents the research field, including public health, mental health, and health policy. In terms of the concept of connections between categories. The main link between social media sources and data mining technology are research entities, such as vaccines and respirators. The main connections between data mining technology and research objects are common analysis methods like fact checking and social network analysis. The main relationship between the research object and the research field is the research theory, such as conspiracy theories. The main link between the research field and social media sources are outcome effects, like echo chamber effects.

In order to further explore the thematic characteristics of Chinese research, we attempted to cluster high-frequency keywords in Chinese research. However, the keywords of the 266 articles written by Chinese authors were insufficient to construct an effective co-occurrence network, indicating that Chinese research lack concentration.

Thematic Map Analysis

We calculated the density and centrality of the high-frequency keyword co-occurrence matrix, visualized two-dimensionally the main categories, and drew a theme map that reflects the research's popularity and importance (Figure 10). The first quadrant includes basic disciplines like public health and infodemiology, indicating the relevance of these fields. The second quadrant includes affective states (anxiety, fear) and research objects (social networks, blogs), indicating that research on negative psychological impacts through data mining and analysis of COVID-19 misinformation on blogs is relatively mature but of low importance. The third quadrant only contains qualitative research, indicating that this field is less important and does not concern the public. The fourth quadrant includes not only basic vocabulary (COVID-19, social media) but also analysis methods (machine learning, sentiment analysis), indicating that emerging artificial intelligence (AI) technologies for social media data mining remain a potentially popular research topic.

Topic Evolution Trend Analysis

Utilizing Sangi diagram commonly used in engineering to express process changes, we incorporated the dimension of time into the high-frequency keyword co-occurrence matrix and topic clustering to analyze the topic evolution from 2020 to 2022 (Figure 11). Based on the high-frequency keywords conversion of the sangi diagram, we found that the research type has shifted from qualitative analysis to quantitative evidence, the analysis method has shifted from artificial content

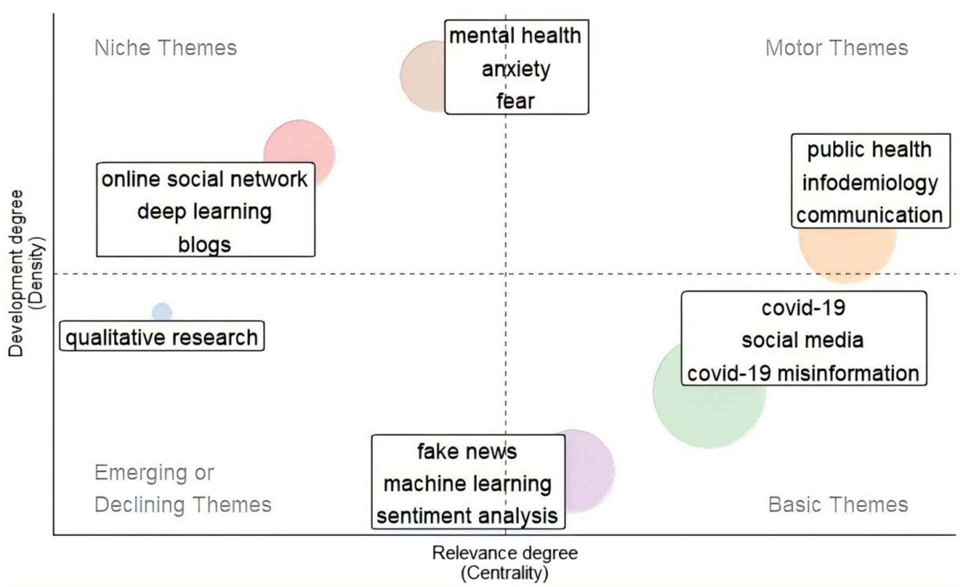

Figure 10 Thematic map per domain.

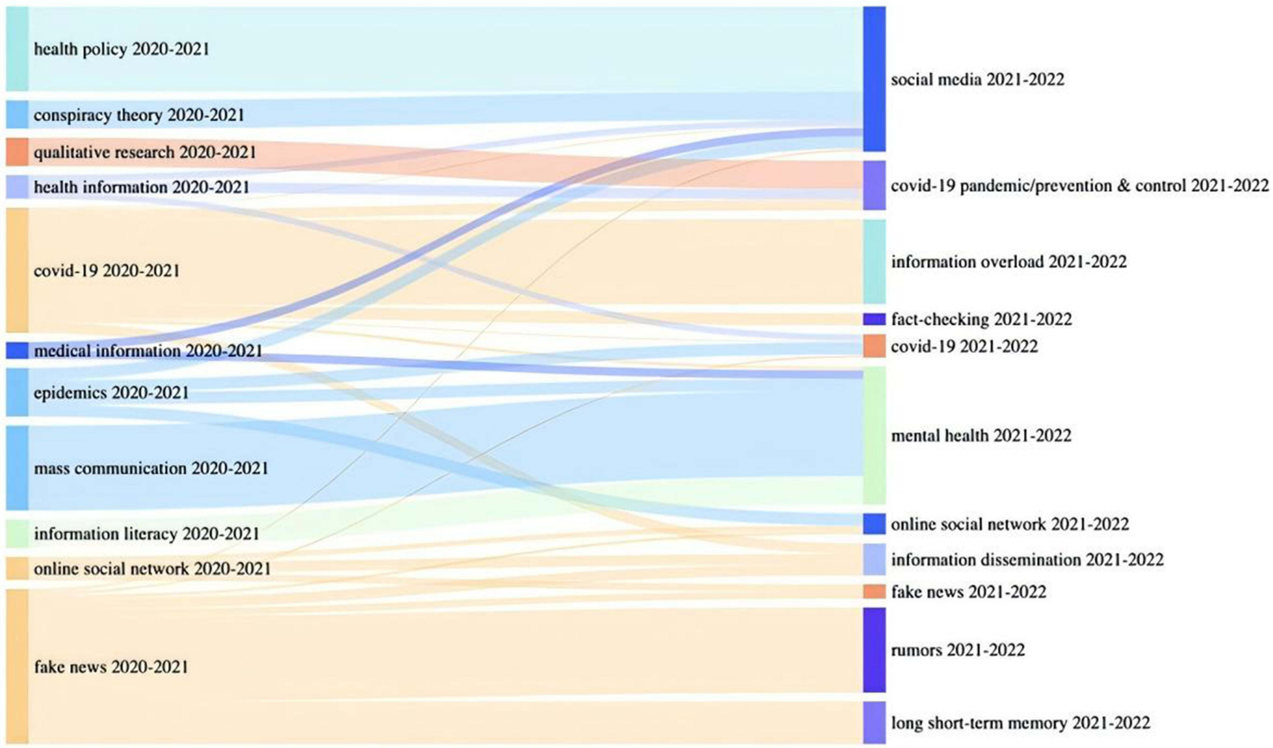

Figure 11 Sankey diagram showing changes in thematic trends from 2020 to 2022.

analysis to AI-based communication network models, the perspective has shifted from single communicator to comprehensive communication source, content, channel, audience, and effect, and the research content has shifted from coarse-grained analysis of macro policies to fine-grained topics like detection of misinformation sources, and analysis of dissemination networks.

Discussion

The Article Number Rapidly Increased Due to the COVID-19 Pandemic

During the COVID-19 pandemic, the number of articles examining the dissemination of health-related misinformation on social media rapidly increased,[28,38] but high-impact articles are lacking. Relevant research existed before this pandemic but received less attention. In September 2019, Wang et al[24] systematically summarized the research status of health-related misinformation spread on social media, comprehensively analyzing 57 articles. They found that relevant articles were first published in 2012 and that the number of articles increased every year, from 7 in 2012 to 41 in November 2018. In January 2021, Yeung et al[27] reported that related research had received more widespread attention. The article number increased from 80 in 2019 to 180 in 2020, and the cumulative number of citations reached nearly 6,000. The present study found that as of December 2022, the number of relevant articles had rapidly increased by 1.94 times from 2020 to 2022 (270 vs 524) and the number of citations by 2.95 times (6,000 vs 17,686). The outbreak of the COVID-19 pandemic may explain the observed surge during this period. The unknowns of COVID-19, the generally low health literacy of the public, and the turbulent international environment exacerbated the spread of COVID-19 misinformation.[15] The speed of COVID-19 misinformation is even faster than that of the pandemic, forming a new infodemic that has caused great harm to personal health management and public health governance.[16] The academic community responded quickly, and many researchers have focused on dissemination mechanisms for health-related misinformation during the COVID-19 pandemic. Through research on misinformation distribution, topics, communication network models, and effects, they hope to alleviate the infodemic. In addition, the number of Chinese articles is significantly lower compared to English articles which may be related to China's unique rumor deletion mechanism.

China is a Major Leader in This Field, but Its Authors are Relatively Fragmented

China and the United States are the leading countries in this field, providing the most funding and publishing the most articles. However, author groups in China are not concentrated, and research institutions are scattered, whereas universities in the United States and Europe are more concentrated. The University of Jordan and Wuhan University are institutions with the highest numbers of publications. As prestigious schools in Asia, their discipline of information management is very impressive, and they intended to use empirical analysis to address the serious infodemic in Asia during the COVID-19 pandemic. Due to the huge harm caused by the infodemic, China has continuously increased its attention and investment in this field since 2020 and provided the most funding. American, British, and Chinese university clusters are the core institutions in this field. Although the cooperation within each cluster is strong, the cooperation between different clusters is relatively weak, which may be related to semantic and grammatical differences in information on mainstream social media like Twitter and Weibo in different countries. Researchers from different countries cooperate to remove barriers caused by language differences in health-related misinformation. The number of high-yield authors is small, with only 6.90% among the 4,682 authors; 93.10% of the authors published only 1 article, and the author with the highest number published only 5 thematically relevant articles. The distribution of Chinese author groups and research institutions is relatively scattered. Most Chinese researchers published only one article, and the core institutions in Chinese university clusters are only Wuhan University and Shanghai Jiaotong University, which may be related to China's unique COVID-19 pandemic prevention policies.

Quantitative Empirical Research Has Become a Research Hot Spot, and the Research Perspective Has Expanded to Multiple Dimensions of Lasswell's Communication Model

During the COVID-19 pandemic, research hot spots in this field have gradually shifted from qualitative communication models to quantitative communication network characteristics. The focus of researchers has gradually shifted from the single communicator to Lasswell's communication model, which refers to describing information dissemination behavior through five dimensions: communicator, information, medium, receiver, and effect,[39] including multiple dimensions related to the communication network like communication sources, content, channels, audiences, and effects. We systematically summarized the relevant research on the spread of health-related misinformation on social media before the outbreak of COVID-19, and found that this conclusion is also applicable to the theme change of relevant research before and after the outbreak of COVID-19. We analyzed the changes in research topics related to misinformation spread on social media before and after the outbreak of COVID-19 from four aspects: research type, analysis method, research subject, and content. The research type gradually shifted from theoretical analysis to data-driven, AI-based empirical analysis. Research on model derivation based on the theoretical framework of psychology, and network science was a hot topic before COVID-19 according to previous studies.[40,41] However, through the high-frequency keywords conversion of the sangi diagram this study showed that researchers pay more attention to the quantitative analysis of big data on social media platforms, mining the characteristics of dissemination networks for COVID-19 misinformation from user-generated content and metadata on social media during the COVID-19 pandemic, such as Nian et al's study.[42] The analysis method gradually shifted to data-mining-oriented AI technology. Before the outbreak of COVID-19, most studies used content analysis alone or as an integral part of the analysis. For example, some researchers observed the number of health-related misinformation, and the health issues involved.[41] Nowadays, researchers focus on the use of new AI technologies like machine learning, and use algorithms like long- and short-term memory networks to effectively detect health-related misinformation.[43] The research subjects gradually shifted from infectious diseases, and chronic non-communicable diseases to coronavirus disease. Before the outbreak of COVID-19, most researchers focused on misinformation related to the Chaka virus, Ebola virus, other infectious diseases,[44] and cancer.[45] Afterward, misinformation related to epidemic prevention and the prevention, diagnosis, and treatment of COVID-19 has become a new research focus, such as Bastani et al's study.[46] Finally, research content gradually shifted from health-related misinformation, and the health problems involved by information disseminators to communication sources,[43] content,[47] channels,[48] audiences,[49] and intervention effects.[50] In addition, the continuous improvement of data training models[51–53] highlights the importance of scholars' attention to health-related misinformation generated by Artificial Intelligence

Generated Content (AIGC). However, the research topics of Chinese authors are relatively scattered, making effective clustering of keywords in their literature challenging. Thus, it is recommended that Chinese authors further focus on research in this field.

Limitations

This study has some limitations. As a bibliometrics study, it lacks a more in-depth content analysis and cannot conduct complete statistical evaluations on the quality, deviation degree, and outcome indicators of the included articles. As a cross-sectional study, it only explores research on the dissemination of health-related misinformation on social media during the COVID-19 pandemic. The publication time range of included studies is relatively short, precluding long-term changes in article numbers and topic evolution. Furthermore, Chinese authors' published literature exhibits a fewer number of high-frequency keywords, making it difficult to effectively cluster them for detailed comparison of thematic differences with other countries' authors. Thus, future research can be further explored.

Conclusions

With the rise of social media and user-generated data, research on health-related misinformation becomes increasingly important. The COVID-19 pandemic facilitated this process. Using bibliometrics, this study quantitatively analyzed 1,294 articles of health-related misinformation disseminated on social media during the COVID-19 pandemic. The number of articles and citations of relevant research during the pandemic increased rapidly, exceeding by far those before COVID-19. The numbers of Chinese articles and citations were far less than those of English articles. China and the United States are the countries with the most papers and the most funding. International cooperation among countries is close, mainly involving academic cooperation between universities. High-impact articles and academic groups are still lacking in this field; Chinese author groups and research institutions are relatively scattered, and the research topics of Chinese authors cannot be effectively clustered. Researchers should combine concepts and methods from different disciplines to further improve misinformation research, and further focus is needed on the research of Chinese authors. Furthermore, research topics have undergone significant changes. Research hot spots have gradually shifted from qualitative research on communication models to quantitative research on communication network characteristics, and the research focus has shifted from misinformation-related health problems to social issues involved in communication sources, content, channels, audiences, and effects. Although studies have been conducted on different social media platforms (Twitter, YouTube, and Facebook in the US, and Weibo and TikTok in China), broader cross-platform research to explore differences in the spread of COVID-19 misinformation on different social media platforms is needed. Meanwhile, the current research landscape lacks further analysis of various types of COVID-19 misinformation, such as images, videos, and even AIGC.

Abbreviations

COVID-19, coronavirus disease 2019, PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Data Sharing Statement

All original data in this study are from literature databases and published articles. Other analysis results are publicly available as attachments.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work. Among them, Yunfan He made significant contributions to the writing of the article, while Jun Liang and Jianbo Lei made significant contributions to the conceptual design of the article.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 81871455), Zhejiang Provincial Natural Science Foundation of China (grant number LY22H180001), Municipal Natural Science Foundation of Beijing of China (grant number 7222306), National TCM innovation team and talent support projects (grant number ZYYCXTD-C-202210), and the National Social Science Major Fund of China (grant number 23&ZD224).

Disclosure

The authors declare no conflicts of interest. This paper has been uploaded to JMIR as a preprint: https://preprints.jmir.org/preprint/49268.

References

1. Shan S, Ju X, Wei Y, Wen X. Concerned or apathetic? Using social media platform (Twitter) to gauge the public awareness about wildlife conservation: a case study of the illegal rhino trade. *Int J Environ Res Public Health*. 2022;19:6869. PMID:35682453. doi:10.3390/ijerph19116869
2. Wardle C, Derakhshan H. Information disorder: an interdisciplinary framework. Available from: https://firstdraftnews.org:443/coe-report/. Accessed May 8, 2023.
3. Tang XJ, Huang CX, Wu XX. New media bluebook: China new media development report no. 11 (2020). Social Science Academic Press; 2020.
4. Fox S, Jones S. The social life of health information. Available from: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx. Accessed May 8, 2023.
5. Qi S, Huiling R, Yifan C, Wanjun X, Jinyin L. Survey on the status of health information use on the internet of community residents in Hefei. *Med Soc*. 2014;27:62–64.
6. Fondazione Bruno Kessler. COVID-19 and fake news in the social media. Available from: https://www.fbk.eu/en/press-releases/covid-19-and-fake-news-in-The-social-media/. Accessed May 8, 2023.
7. Yang Y, Wu S. Health information search behaviors of the public during an epidemic. *Chin J Med Libr Inf Sci*. 2020;29:35–41.
8. Wen D, Zhang X, Liu X, Lei J. Evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. *J Med Internet Res*. 2017;19:e68. PMID:28270382. doi:10.2196/jmir.6874
9. Wang T, Wang W, Liang J, et al. Identifying major impact factors affecting the continuance intention of mHealth: a systematic review and multi-subgroup meta-analysis. *NPJ Digit Med*. 2022;5:145. PMID:36109594. doi:10.1038/s41746-022-00692-9
10. Wang T, Zheng X, Liang J, et al. Use of machine learning to mine user-generated content from mobile health apps for weight loss to assess factors correlated with user satisfaction. *JAMA Netw Open*. 2022;5:e2215014. PMID:35639374. doi:10.1001/jamanetworkopen.2022.15014
11. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. *JMIR MHealth UHealth*. 2018;6:e94. PMID:29650506.
12. Liang J, He Y, Fan L, et al. A preliminary study on the abnormal deaths and work burden of Chinese physicians: a mixed method analysis and implications for smart hospital management. *Front Public Health*. 2021;9:803089. PMID:35059382. doi:10.3389/fpubh.2021.803089
13. 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*. 2021;38:143–149. PMID:32657000. doi:10.1111/hir.12320
14. Zhao S, Hu S, Zhou X, et al. The prevalence, features, influencing factors, and solutions for COVID-19 vaccine misinformation: systematic review. *JMIR Public Health Surveill*. 2023;9:e40201. PMID:36469911. doi:10.2196/40201
15. World Health Organization. Novel coronavirus. nCoV situation report. 2019. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf. Accessed May 8, 2023.
16. Eysenbach G. Infodemiology: the epidemiology of (mis)information. *Am J Med*. 2002;113(9):763–765. doi:10.1016/s0002-9343(02)01473-0
17. Lavorgna L, De Stefano M, Sparaco M, et al. Fake news, influencers and health-related professional participation on the web: a pilot study on a social-network of people with multiple sclerosis. *Mult Scler Relat Disord*. 2018;25:175–178. PMID:30096683. doi:10.1016/j.msard.2018.07.046
18. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. *J Med Internet Res*. 2021;23:e17187. PMID:33470931. doi:10.2196/17187
19. Okuhara TS, Ishikawa H, Okada M, Kato M, Kiuchi T. Assertions of Japanese websites for and against cancer screening: a text mining analysis. *Asian Pac J Cancer Prev*. 2017;18:1069–1075. PMID:28547943. doi:10.22034/APJCP.2017.18.4.1069
20. Borges Do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. *Bull World Health Organ*. 2022;100:544–561. PMID:36062247. doi:10.2471/BLT.21.287654
21. Jabbour D, Masri JE, Nawfal R, Malaeb D, Salameh P. Social media medical misinformation: impact on mental health and vaccination decision among university students. *Ir J Med Sci*. 2023;192:291–301. PMID:35119644. doi:10.1007/s11845-022-02936-9
22. Gill R, Goolsby R, editors. *COVID-19 Disinformation: A Multinational, Whole of Society Perspective*. New York: Springer International Publishing; 2022.
23. Lu JH. Themes and evolution of misinformation during the early phases of the COVID-19 outbreak in China—an application of the crisis and emergency risk communication model. *Front Commun*. 2020;5:57. doi:10.3389/fcomm.2020.00057
24. 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;240:112552. PMID:31561111. doi:10.1016/j.socscimed.2019.112552
25. Pool J, Fatehi F, Akhlaghpour S. Infodemic, misinformation and disinformation in pandemics: scientific landscape and the road ahead for public health informatics research. *Stud Health Technol Inform*. 2021;281:764–768. PMID:34042681. doi:10.3233/SHTI210278
26. Wang S, Su F, Ye L, Jing Y. Disinformation: a bibliometric review. *Int J Environ Res Public Health*. 2022;19:16849. PMID:36554727. doi:10.3390/ijerph192416849
27. Yeung AWK, Tosevska A, Klager E, et al. Medical and health-related misinformation on social media: bibliometric study of the scientific literature. *J Med Internet Res*. 2022;24:e28152. PMID:34951864. doi:10.2196/28152

28. Joseph AM, Fernandez V, Kritzman S, et al. COVID-19 misinformation on social media: a scoping review. *Cureus*. 2022;14:e24601. PMID:35664409.
29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372: n71. PMID:33782057. doi:10.1136/bmj.n71
30. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr*. 2017;11:959–975. doi:10.1016/j.joi.2017.08.007
31. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2016.
32. Van Eck NJ, Waltman L. *Visualizing Bibliometric Networks. Measuring Scholarly Impact: Methods and Practice*. Springer; 2014:285–320.
33. Truewell RL. *Some Behavioral Patterns of Library Users: The 80/20 Rule*. New York: Wilson Libr Bull; 1969.
34. Vickery BC. Bradford's law of scattering. *J Doc*. 1948;4:198–203. doi:10.1108/eb026133
35. de Solla Price DJ. *Little Science, Big Science*. Columbia University Press; 1965.
36. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. *J Informetr*. 2011;5:146–166. doi:10.1016/j.joi.2010.10.002
37. Wang KL, Jin-Hua LI. Thematic evolution analysis of information behavior in China in recent ten years. *Inf Sci*. 2018. doi:10.28945/4028
38. Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD. The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. *Z Gesundh Wiss*. 2021:1–10. PMID:34660175. doi:10.1007/s10389-021-01658-z
39. Lasswell H. *The Structure and Function of Communication in Society*. Communication University of China Press; 1948.
40. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. *Health Commun*. 2018;33:1131–1140. PMID:28622038. doi:10.1080/10410236.2017.1331312
41. 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:e1. PMID:27227144. doi:10.2196/publichealth.5059
42. Nian FZ, Guo X, Li JZ. A new spreading model in the environment of epidemic-related online rumors. *Mod Phys Lett B*. 2022;36:2150569. doi:10.1142/S0217984921505692
43. Rani P, Jain V, Shokeen J, Balyan A. Blockchain-based rumor detection approach for COVID-19. *J Ambient Intell Humaniz Comput*. 2022:1–15. PMID:35611303. doi:10.1007/s12652-022-03900-2
44. Pathak R, Poudel DR, Karmacharya P, et al. YouTube as a source of information on Ebola virus disease. *N Am J Med Sci*. 2015;7:306–309. PMID:26258077. doi:10.4103/1947-2714.161244
45. Chen B, Shao J, Liu K, et al. Does eating chicken feet with pickled peppers cause avian influenza? Observational case study on Chinese social media during the Avian influenza A (H7N9) outbreak. *JMIR Public Health Surveill*. 2018;4:e32. PMID:29599109. doi:10.2196/publichealth.8198
46. Bastani P, Bahrami MA. COVID-19 related misinformation on social media: a qualitative study from Iran. *J Med Internet Res*. 2020. PMID:32250961.
47. 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:e24585. PMID:33480853. doi:10.2196/24585
48. Brian houston JB, Thorson E, Kim E, Mantrala MK. COVID-19 communication ecology: visualizing communication resource connections during a public health emergency using network analysis. *Am Behav Sci*. 2021;65:893–913. doi:10.1177/0002764221992811
49. Tran HTT, Nguyen MH, Pham TTM, et al. Predictors of eHealth literacy and its associations with preventive behaviors, fear of COVID-19, anxiety, and depression among undergraduate nursing students: a cross-sectional survey. *Int J Environ Res Public Health*. 2022;19:3766. PMID:35409448. doi:10.3390/ijerph19073766
50. Bak-Coleman JB, Kennedy I, Wack M, et al. Combining interventions to reduce the spread of viral misinformation. *Nat Hum Behav*. 2022;6:1372–1380. PMID:35739250. doi:10.1038/s41562-022-01388-6
51. Liu C, Wang DL, Zhang H, et al. Using simulated training data of voxel-level generative models to improve 3D neuron reconstruction. *IEEE Trans Med Imaging*. 2022;12:3624–3635. PMID:35834465. doi:10.1109/TMI.2022.3191011
52. Huang Z, Wen J, Chen S, Zhu L, Zheng N. Discriminative radial domain adaptation. *IEEE Trans Image Process*. 2023;32:1419–1431. PMID:37018670. doi:10.1109/TIP.2023.3235583
53. Wang PF, Wang HY, Zheng NG. Stochastic cubic-regularized policy gradient method. *Knowledge-Based Syst*. 2022;255:109687. doi:10.1016/j.knosys.2022.109687