

Impact of Public Health Event on COVID-19 Vaccine Hesitancy on Social Media: National Infoveillance Study

Zizheng Zhang, Guanrui Feng, Jiahong Xu, Yimin Zhang, Jinhui Li, Jian Huang, Babatunde Akinwunmi, Casper J.P. Zhang, Wai-Kit Ming

Submitted to: JMIR Public Health and Surveillance
on: August 16, 2021

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5
Supplementary Files..... 19
 Figures 20

Ahead Of Print
JMIR Publications

Impact of Public Health Event on COVID-19 Vaccine Hesitancy on Social Media: National Infection Surveillance Study

Zizheng Zhang^{1,2,3} BSc; Guanrui Feng¹ MPH; Jiahong Xu¹ MPH; Yimin Zhang⁴ MBBS; Jinhui Li⁵ PhD; Jian Huang^{6,7} PhD, MPH; Babatunde Akinwunmi^{8,9} MD, MMSc; Casper J.P. Zhang¹⁰ PhD, MPH; Wai-Kit Ming¹ MD, PhD, MPH, MMSc

¹Department of Public Health and Preventive Medicine, School of Medicine Jinan University Guangzhou CN

²Jinan University-University of Birmingham Joint Institute Jinan University Guangzhou CN

³School of Mathematics, College of Engineering and Physical Sciences University of Birmingham Birmingham GB

⁴School of Medicine Jinan University Guangzhou CN

⁵School of Journalism and Communication Jinan University Guangzhou CN

⁶Singapore Institute for Clinical Sciences (SICS) Agency for Science, Technology and Research (A*STAR) Singapore SG

⁷Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine Imperial College London London GB

⁸Department of Obstetrics and Gynecology Brigham and Women's Hospital Boston US

⁹Center for Genomic Medicine (CGM), Massachusetts General Hospital, Harvard Medical School Harvard University Boston US

¹⁰School of Public Health The University of Hong Kong Hong Kong HK

Corresponding Author:

Wai-Kit Ming MD, PhD, MPH, MMSc

Department of Public Health and Preventive Medicine, School of Medicine

Jinan University

Room 324, 3/F, Second Social Science Building, Jinan University, Guangzhou, China

Guangzhou

CN

Abstract

Background: The ongoing Coronavirus 2019 (COVID-19) pandemic has brought unprecedented challenges to every country in the world. A call for global vaccination of COVID-19 plays a pivotal role in the fight against this virus. With the development of COVID-19 vaccines, public willingness to accept a vaccination has become an important concern of public health given the vaccine hesitancy observed in the world. Social media is powerful in monitoring public attitudes and assess the dissemination, which would provide valuable information for policy makers.

Objective: This study aimed to investigate the responses of vaccine positivity on social media when public health events were reported.

Methods: A total of 340,783 vaccine-related posts was captured with the poster's information on Weibo, the biggest social platform in China. After data cleaning, 156,223 posts were included in the subsequent analysis. By using pandas and SnowNLP Python libraries, posts were classified into 2 categories, positive and negative. After model training and sentiment analysis, the proportion of positive posts was computed to measure the public positivity toward the COVID-19 vaccine.

Results: The positivity of COVID-19 vaccine in China tends to fluctuate over time in the range of 45.7% to 77.0% and is intuitively correlated with public health events. In terms of gender, males were more positive (70.0% of the time) than females. In terms of region, when regional epidemics arose, not only the region with the epidemic and surrounding regions but even the whole country showed more positive attitudes to varying degrees. When the epidemic subsided temporarily, positivity decreased with varying degrees in each region.

Conclusions: In China, public attitudes of COVID-19 vaccination vary from gender and region. A regional epidemic or news on social media may cause variations in willingness to accept a vaccination. It is crucial for policy makers to adjust their policies through the use of positive incentives with prompt responses to pandemic-related news to promote vaccination acceptance.

(JMIR Preprints 16/08/2021:32936)

DOI: <https://doi.org/10.2196/preprints.32936>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org/](#)

Original Manuscript



Original Paper

Zizheng Zhang^{1,2,3}, Bsc; Guanrui Feng¹, MPH; Jiahong Xu¹, MPH; Yimin Zhang⁴, MBBS; Jian Huang^{5,6}, PhD; Jinhui Li⁷, PhD; Babatunde Akinwunmi^{8,9}, MD, MMSc; Casper J.P. Zhang¹⁰, PhD and Wai-Kit Ming^{1,*}, MD, PhD, MPH, MMSc.

¹ Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou 510632, Guangdong, China.

² Jinan University-University of Birmingham Joint Institute, Jinan University, Guangzhou 510632, Guangdong, China.

³ School of Mathematics, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, Birmingham B12 2TT, UK.

⁴ School of Medicine, Jinan University, Guangzhou 510632, Guangdong, China.

⁵ Singapore Institute for Clinical Sciences (SICS), Agency for Science, Technology and Research (A*STAR), Singapore 138632.

⁶ Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, W2 1PG, UK.

⁷ School of Journalism and Communication, Jinan University, Guangzhou 510632, Guangdong, China.

⁸ Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Boston, MA 02115, USA

⁹ Center for Genomic Medicine (CGM), Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, MA 02115, USA.

¹⁰ School of Public Health, The University of Hong Kong, Pokfulam, Hong Kong, China.

* Corresponding Author:

Wai-Kit Ming, MD, MPH, PhD, MMSc
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
West Huangpu Road 601
Guangzhou
CN
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Impact of Public Health Event on COVID-19 Vaccine Hesitancy on Social Media: National Infoveillance Study

Abstract

Background: The ongoing Coronavirus 2019 (COVID-19) pandemic has brought unprecedented challenges to every country in the world. A call for global vaccination of COVID-19 plays a pivotal role in the fight against this virus. With the development of COVID-19 vaccines, public willingness to accept a vaccination has become an important concern of public health given the vaccine hesitancy observed in the world. Social media is powerful in monitoring public attitudes and assess

the dissemination, which would provide valuable information for policy makers.

Objective: This study aimed to investigate the responses of vaccine positivity on social media when major public events (major outbreaks) or major adverse event of vaccination (COVID-19 or other similar vaccines) were reported.

Methods: A total of 340,783 vaccine-related posts was captured with the poster's information on Weibo, the biggest social platform in China. After data cleaning, 156,223 posts were included in the subsequent analysis. By using pandas and SnowNLP Python libraries, posts were classified into 2 categories, positive and negative. After model training and sentiment analysis, the proportion of positive posts was computed to measure the public positivity toward the COVID-19 vaccine.

Results: The positivity of COVID-19 vaccine in China tends to fluctuate over time in the range of 45.7% to 77.0% and is intuitively correlated with public health events. In terms of gender, males were more positive (70.0% of the time) than females. In terms of region, when regional epidemics arose, not only the region with the epidemic and surrounding regions but even the whole country showed more positive attitudes to varying degrees. When the epidemic subsided temporarily, positivity decreased with varying degrees in each region.

Conclusions: In China, public positivity of COVID-19 vaccines fluctuates over time and a regional epidemic or news on social media may cause significant variations in willingness to accept a vaccination. Furthermore, public attitudes of COVID-19 vaccination vary from gender and region. It is crucial for policy makers to adjust their policies through the use of positive incentives with prompt responses to pandemic-related news to promote vaccination acceptance.

Keywords: COVID-19; Vaccine Hesitancy; Social Media; China; Sentiment Analysis; Infoveillance.

Introduction

At the end of 2019, the first case of the Coronavirus 2019 (COVID-19) infection was reported in Wuhan, China. The disease spread rapidly throughout China after which it soon evolved into a global pandemic. By the end of May, 2021, the total number of confirmed cases globally exceeded 100 million and the cumulative number of deaths was reported a > 3 million with a mortality rate of about 2.07% [1]. Although the rate was lower than severe acute respiratory syndrome corona virus 1 and Middle Eastern respiratory coronavirus (SARS-CoV-1, 9.5% and MERS, 34.4% respectively), it cannot be ruled out that COVID-19 has stronger transmissibility than either one of those viruses [2, 3]. The rapid spread of COVID-19 has brought unprecedented challenges to each country in the world in terms of social, economic, cultural, and/or political aspects.

A vaccination is considered to be the most effective and safest way to give the human body immunity to face new infectious diseases. According to statistics, the world's current kinds of vaccines can save more than 3 million lives related to > 20 diseases every year [4]. In order to control the worldwide spread of the COVID-19, a call for the global vaccination of COVID-19 is required [5]. In Mid-March, 2020, China's recombinant COVID-19 vaccine was approved, and clinical trials were initiated [6]. Up to now, at least 13 different COVID-19 vaccines have been put into use throughout the world, including the Sinopharm COVID-19 vaccine [7]. However, with the continuous development of the global Internet, the anti-vaccine campaign is also spreading rapidly through social media platforms, thus causing a threat to optimal global vaccine delivery [8-10].

Social media has played a key role in information dissemination during the COVID-19 pandemic. Through social media, important epidemic-related information can be easily disseminated, and people across the world can quickly obtain relevant disease-related information, participate in the discussions, and express their own views about COVID-19 [11, 12]. In the meantime,

misinformation, defined as erroneous or incorrect information, has also been widely spreading during the pandemic[13]. Although misinformation about COVID-19 is posted more than evidence-based information in social media [14], science-based information has had more reposts than the false information [14], and the platforms have responded to much misinformation identified by fact-checkers [15].

Weibo is one of the representative social media with most users in China, which has more than 500 million active users and more than 700 billion views [12, 16-19]. It has become one of the primary social platforms for Chinese Internet users to disseminate and acquire health information [20]. Up to June 2020, China had nearly 1,000 million netizens, accounting for 67% of all Chinese citizens [21, 22]. According to the 2020 annual report released by Weibo, users checked the COVID-19 information 16.1 billion times every day during the outbreak [23]. Particular to the COVID-19 vaccine, more than 100,000 Weibo users participated in the discussion with a cumulative reading of more than 500 million times [24]. Given its popularity and the massive information contained within the site, Weibo can be considered as an appropriate data source to investigate the public attitudes toward the COVID-19 vaccine.

For sentiment analysis under public health emergencies, many studies using web crawlers, text mining, and other technologies to collect a variety of public opinion information from Internet can be found [25]. In addition, some studies use the web text data according to different phases and classify these data according to the theme and emotion [26, 27] and adopt various visualization tools to investigate the public sentiment, proving that the social media can be applied to the measure the public attention toward public health emergencies [27]. With the ongoing worldwide COVID-19 and successful entry of its related vaccines onto the market, some studies have begun to focus on the information of social media to analyze the acceptance of vaccines by the public, emphasizing that public's attitude on health issues are strongly influenced by social media [28].

However, many studies have not yet analyzed the sentiments of the Chinese population nationwide through their statements on domestic social media sites, such as Weibo. While a number of studies have been conducted abroad on social media, such as Twitter and Facebook [9, 29-31], we do not yet clearly know the current sentiments and attitudes of the population towards vaccination for the COVID-19 in China. Also, only a few studies have investigated the relationship between social media context and public sentiment toward vaccination [32]. Clear and concise sentiment analysis of textual information on Weibo will not only improve the monitoring of public opinion on the internet but also effectively allow application of the results of emotional psychology studies to provide early warnings of unusual occurrences. The study of such psychological indicators is a very important guide for government policies at this particular stage [33-37] and would enable national governmental departments to better understand the attitude of the public toward vaccination, thus advancing collaboration with multiple parties more effectively to increase the vaccination rate of COVID-19.

The study aimed to investigate the public sentiment of COVID-19 vaccine and to evaluate gender and regional variations in this sentiment. The feasibility of a social sentiment analysis based on online data of hot-spot events and whether the same approach can be used in the future to keep track of online public opinion during the vaccination period was assessed. The aim of this process was to provide a realistic grasp of the dynamic psychology of the public and highlight the leading role of the national government departments. The study also highlights the essential role of national governmental departments in moderating public sentiment through social media.

Methods

Based on the public nature of the Weibo platform, this study used Python for data mining and sentiment analysis of the resulting text to crawl and analyze public comments published by Weibo users on the issue of the COVID-19 vaccination, thus allowing identification of the sentiment tendencies of the resulting text to be made.

Data Collection

Processing

Python 3.9.2 (Python Software Foundation, Beaverton, Oregon) [38] and related libraries were utilized to simulate logging and then capture the required data. The data obtained containing the identifier (ID) of the post, the context of the post, the post time, the re-post times, the number of “likes”, the gender of the person posting, the location of this person, and the posting person’s birthday were saved as multiple csv files. Due to the anti-crawler mechanism of Weibo, outliers beyond the setting date or keyword ranges were excluded.

Inclusion

Data were captured from the search results of Weibo with keyword ‘COVID-19 Vaccine (新冠肺炎疫苗)’ between October 18, 2020 and May 15, 2021 (inclusive of both dates). As the general search criteria, the captured posts could refer to any approved COVID-19 vaccines globally. However, the search result may tend to the vaccine that available in China. Since the availability of the vaccine to the Chinese public can be dated back to mid-October, 2020 [39], the chosen time period is believed to cover the process from the vaccine development to the mass vaccination scheme.

Exclusion Criteria

Given that the study focused on the public opinion in China, any texts written in languages other than Chinese and those posts from users whose locations were outside China were excluded. Any posts consisting of only symbols or numbers were also excluded.

Data Cleaning

Text Cleaning

First, the text contains no Chinese characters, namely posts written in other languages, or posts consists of only symbols or numbers were removed. Then, the posts with missing information, such as location or date, duplicated post and posts from public accounts were also removed[40].

Relevance-based Cleaning

Because of the specific writing style of social media, the relevance between the context and study topic is a vital issue to be considered[41]. In this part, a “base text” is set to describe the proven determinants of COVID-19 vaccine acceptance in China [42] and compared with each crawled post to obtain their similarity. Cosine similarity, which is conceived to be a powerful approach in natural language processing (NLP), was performed to measure the similarity between the crawled post and base text as formulated in the following model:

$$\text{similarity} = \cos(\theta) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

in which A_i and B_i is the i -th item of word frequency vector of the extracted keyword list via term

frequency–inverse document frequency (TF–IDF) from the base text and crawled post, respectively. After a trail contained 1000 randomly chosen posts, a threshold of 0.025 was set to distinguish the relevant post from the irrelevant ones, which attained an accuracy rate of 94.1% (941/1000). The model was then applied to the dataset, and irrelevant posts were removed.

Sentiment Analysis

Sentiment analysis is a natural language processing to detect subjectivity in text, extracting and classifying opinions and sentiments [43]. SnowNLP [44], which is a specialized Python library for Chinese language processing and has been used in social media text mining for medical researches especially COVID-19 related ones given its feasibility and accuracy [21, 45-47], was used to perform sentiment analysis.

Fifteen thousand randomly chosen posts were annotated manually, each of them was coded by 2 researchers, one of whom annotated independently and the other double-checked, of which 12,000 and 3,000 posts were randomly split into the training and test sets. The training set included 9,084 positive and 2,916 negative posts (“neutral” was not used as a category due to its limited research significance [45]).

The process of SnowNLP includes word segmentation, stop words removal, and naïve Bayes classification. The key model is shown below:

$$P(c_i|T) = \frac{P(T \vee c_i) \cdot P(c_i)}{P(T)}$$

where:

$$P(T) = P(T|c_1) \cdot P(c_1) + P(T|c_2) \cdot P(c_2)$$

in which T is the eigenvector of the text, and c_i is the i -th emotion category, in this case, positive and negative. The posts are divided into those with probabilities higher than the threshold (positive category) or negative otherwise. Given the unbalanced distribution of label in training set, a receiver operating characteristics (ROC) curve was introduced [48] to evaluate the model. The area under the ROC curve (AUC) was then computed to measure the outcome of the classifier. After training, the threshold was set as 0.5889 for which the AUC yielded 0.81.

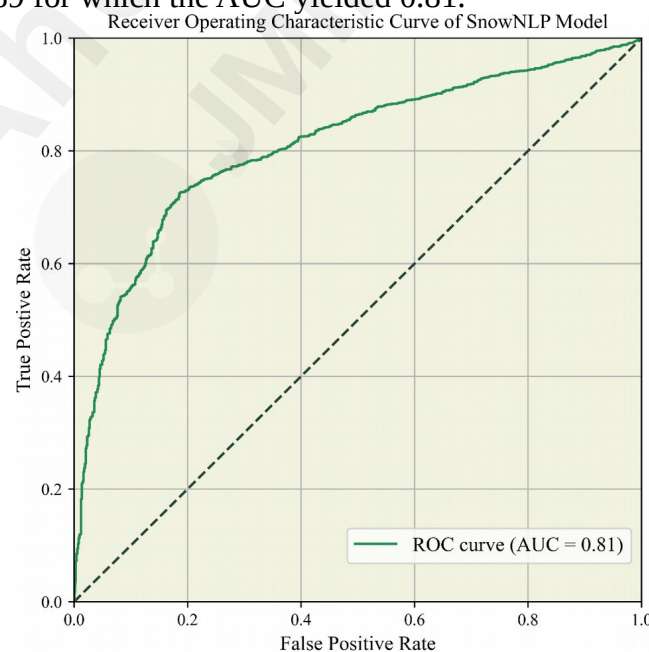


Figure 1. Receiver operating characteristic (ROC) curve of fitted SnowNLP model

After training the model, the emotions in the posts were computed. The sentiment score, S , was calculated via the formula:

$$S = \frac{\text{Pos}}{\text{Pos} + \text{Neg}}$$

in which Pos and Neg are the number of positive and negative posts, respectively. A sentiment score ranged from 0 to 1, indicating the most negative to the most positive.

Results

A total of 340,783 posts, including both original and re-posted posts, and user-related information were retrieved. After cleaning, 156,223 posts were included in analysis. According to the statistics, more female than male posters were noted, and more positive than negative posts were found. Table 1 shows the number of posts and users.

Table 1. Descriptive Statistics of Total Posts

Topics	Number of Posts / Users (%)
Users	98,600
Male Users	45,812 (46.5%)
Female Users	52,788 (53.5%)
Total Posts	156,223
Posts from Male Users	84,353 (54.0%)
Posts from Female Users	71,870 (46.0%)
Positive Posts	93,660 (60.0%)
Negative Posts	62,563 (40.0%)

From Figure 2, the overall public positivity tends to fluctuate over time. The decline in positivity was consistent with the reported information about the side effects of the COVID-19 vaccine or other vaccines in general. In the week after October 18, 2020, 59 people in South Korea were reported to have died after the influenza vaccine injection, and at the same time a rapid decline in the overall user's positivity for the COVID-19 vaccine occurred a short time after that report. In terms of gender, both men and women presented positive attitudes about the COVID-19 vaccine across the most of study periods, and the fluctuation patterns of emotional score between two genders is generally similar. Interestingly, although the trend of male and female emotional fluctuations was generally consistent, the overall positivity of female users was weaker than that of male users most of the time. During the period from February 21, 2021 to March 21, 2021, China Central Television (CCTV) announced that pregnant and lactating women should postpone the vaccination. Unlike male users, female user positivity for the COVID-19 vaccine decreased rapidly during this week (February 21–27, 2021) but began to rise again in the week beginning on March 21, 2021.

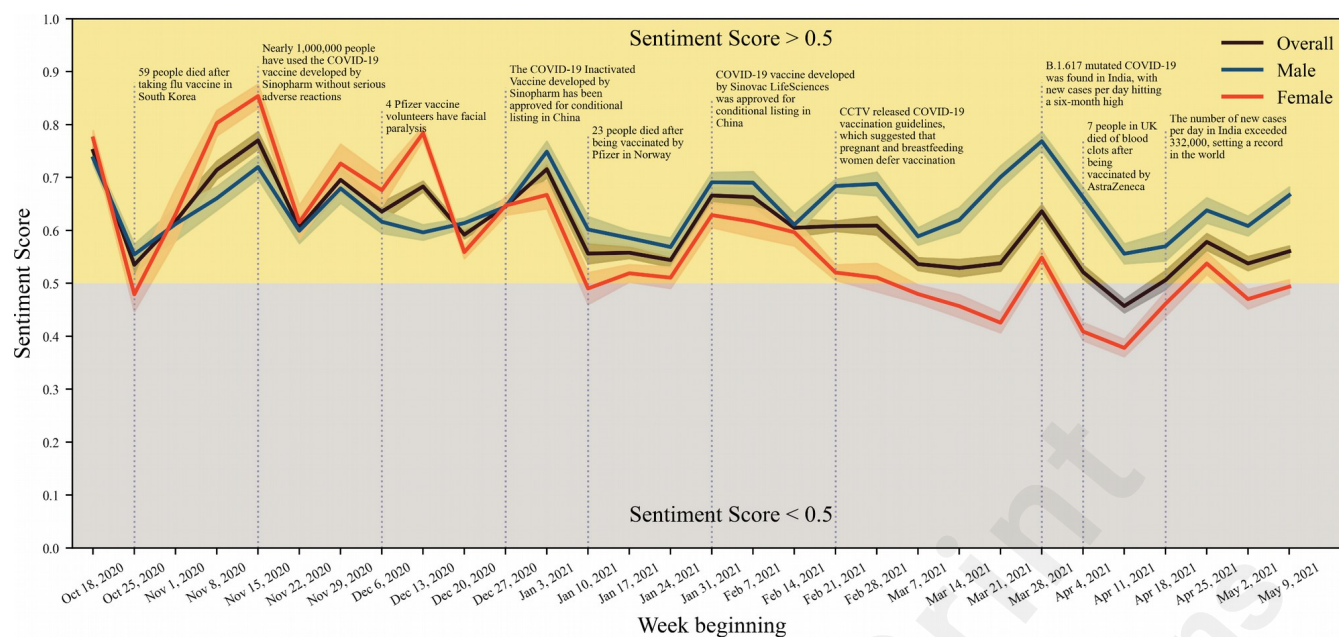
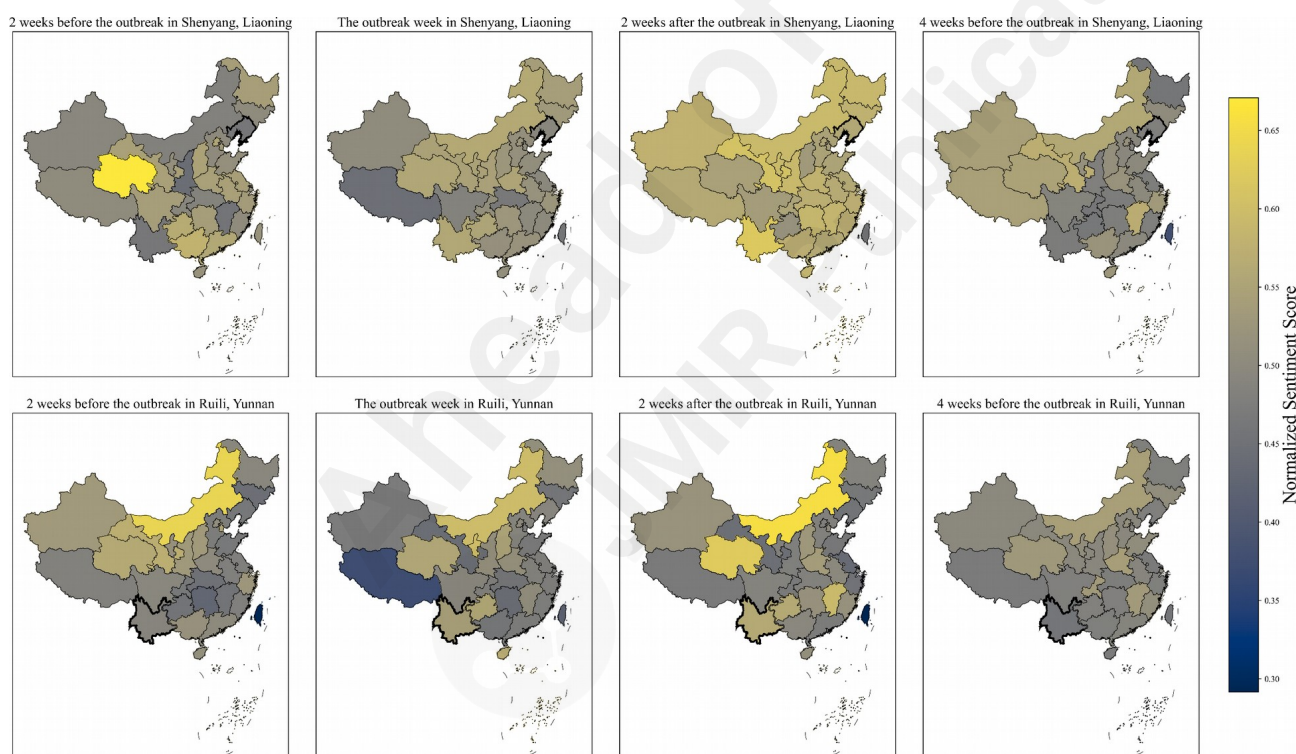


Figure 2. The Variation in Public Sentiment of Coronavirus 2019 (COVID-19) Vaccine over Time between Gender (a higher sentiment score indicates higher positivity of COVID-19 Vaccine)

Figure 3. The Variation of Public Positivity of COVID-19 Vaccines before and after the Outbreak



Period in Ruili and Shenyang (the boundaries of provinces with outbreak is highlighted in bold blacked lines)

In consideration of the two outbreaks in Shenyang, Liaoning Province in January 2021 and Ruili, Yunnan Province in March 2021, we present the heat map of the normalized sentiment score across all regions and provinces of China in heat maps by focusing on the period from the two weeks before the outbreak to four weeks after the outbreak (Figure 3). Since the outbreak in these two regions, the sentiment about vaccination in this region and its surrounding regions had increased significantly whereby the vaccination positivity gradually declined two weeks after the outbreak was reported. In

terms of the outbreak in Shenyang (upper panels in Figure 3), the positivity of vaccination increased not only in its own province but also in Northeast China and even throughout whole country. Similarly, after the outbreak was reported in Ruili, Yunnan Province, the sentiment toward vaccination in Guizhou Province, a neighbor of Yunnan, also increased significantly.

Discussion

COVID-19 vaccine hesitancy is a worldwide phenomenon and is a crucial issue to be solved in the fight against the pandemic. In China, we found that gender-specific emotional responses to vaccines could be influenced by various major public events over time, and the degree of influence varied by gender with women as the more vaccine-hesitant group. Also, public positivity changed significantly in the weeks before and after the COVID-19 outbreak. A fairly substantial body of previous research demonstrating that women experience and express more intense emotion than men with regard to both positive and negative emotions can be found [49-51]. Recent media studies further indicate that female users are more likely to seek emotional support from online communities, while male users tend to provide information-related help [52, 53]. Moreover, the difference in vaccine acceptance between genders has been reported in some previous studies focusing on flu vaccines for which vaccine acceptance is greater in men than women [54-56], findings which also support our results. The current study confirmed the emotional tendency of male and female users towards COVID-19 vaccines on social media, thus extended the literature of gender difference to the specific context of public health events. The findings also inspire the policy makers for refined strategies in negative sentiment management. Interestingly, although previous studies show that women are more vaccine-hesitant [54-56], some studies have reported that vaccine coverage may be higher in women than men [54]. The reason for this difference could result from women visiting preventive healthcare services and physicians more frequently [57]. Therefore, the policy maker should pay more attention to female communities' sentiments concerning major public health events and at the same time monitor the vaccination coverage to provide the in-time responses.

We found that the emotional tendencies of the public are dynamic, and positive and negative tendencies exist with respect to expressed emotional tendencies by the public. Every time a vaccine-related adverse event occurs, it may cause a decline in positive sentiment among Chinese Internet users on Weibo, which is highly contagious in social media. Meanwhile, spreading false and appalling information on Weibo, which may bring about feelings of depression and anxiety for certain groups of people [58, 59]. Therefore, when a COVID-19 emergency occurred in their particular region, people's sense of fear and self-protection led to increased positive emotions toward vaccinations. Interestingly, during the outbreak in Shenyang, Liaoning, the positivity increased in this epidemic region and throughout the whole country. Even though the scenario could have resulted from the outbreak in other provinces in that period, we cannot rule out the possibility that regional outbreaks may affect national positivity. Also, after the outbreak of the epidemic in Shenyang, Liaoning Province, Yunnan Province was the province with one of the most significant increases in vaccination positivity. Therefore, the public positivity of COVID-19 vaccine in previous epidemic regions may have a retrospective effect on vaccination positivity. As the situation of the outbreak improved, people began to gradually decrease the release of positive emotions again due to decreased vigilance. Future studies should consider the use of social media to guide the public sentiment after the epidemic outbreak is over.

Although other emerging research studies have investigated COVID-19 vaccination intention using methods, such as questionnaires, limitations, such as the existence of some bias by inferring the perceptions and attitudes of the group with only a small sample, still exist [60-64]. Sentiment

analysis through the use of big data offers a more direct way to monitor the emotion of the citizens. Further studies should focus on the relationship between the positivity and the case rate growth or death rate.

With the popularity of the Internet and economic development, social media has become a medium for people to express their emotions and opinions. For government officials and public health professionals, understanding public sentiment is critical to developing policies for infectious disease prevention and control and health care resource allocation. In the context of a global COVID-19 outbreak, the vaccine is an important measure to establish herd immunity against COVID-19 in an open border setting [65]. Therefore, understanding public sentiment about vaccines is an effective way for the government to promote COVID-19 vaccination in a rational and orderly manner. Exploring the factors and behaviors that influence positivity between different genders and different regions through the use of Internet data can provide relevant information for government departments that are trying to assist in decision-making and providing health services. It also reminds relevant departments to establish public opinion and sentiment monitoring networks to understand the dynamics of public attitudes toward vaccines, predict changes in sentiment, and plan vaccine production and resource allocation rationally. This process is crucial for the government to better understand public sentiment through social media and convey information accurately and timely, which will also answer vaccine-related queries and increase vaccination motivation.

Sentiment analysis can reveal differences between cities and regions, and when combined with current COVID-19 vaccine postings on social media and dynamic microblogs postings based on geolocation data, can be used as a decision support point for government agencies. This type of analysis can also provide effective and real-time recommendations to government agencies that are based on the average number of microblogs per city and region and emotional tendencies; if this number is well above a significant peak, the information can be quickly reported to official agencies. Our text sentiment analysis tool can be an extension of this research, capturing the relevant information needed in real-time.

This study has some limitations. First, our data collection was conducted only on one social media platform, Weibo. The opinions of those who did not use Weibo were not included. Second, the coding process was not completely independent, which may cause bias in the training process. Third, the gender and location information were self-reported by the users, which is a common issue in research based on social media such as twitter [66]. Fourth, due to the anti-crawler mechanism of Weibo, a small proportion posts randomly lost during crawling. Fifth, sentiment may not be the only factor affecting vaccination acceptance. Local governments may take advantage of social media in promoting vaccination, but other challenges, such as misinformation and the allocation of vaccines, still exist.

Conclusions

The public opinion is closely related to public health events in China. When positive news about COVID-19 vaccine occurs, the public will be more positively sentimental about the vaccine and vice versa. This sentimental reaction appears to be gender-specific, by which men tend to be more open-minded compared to women. In terms of regional differences, the positivity of province and its surrounding (and even the whole country) in which a pandemic occurs, was shown to increase and then decrease back to normal after 2 to 4 weeks. It is crucial for the government to adjust vaccination policies promptly in response to the public health events in order to promote massive vaccination via dynamic monitoring public sentiments.

Author's Contributions

ZZ and WKM conceptualized and design the study. ZZ contributed to methodology, data collection, visualization and project administration. ZZ, GF, JX and YZ analyzed the data and drafted the manuscript. JH, JL, BA, CJPZ and WKM reviewed and edited the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Abbreviations

AUC: Area under ROC Curve

CCTV: China Central Television

COVID-19: Coronavirus 2019

ID: Identifier

NLP: Natural Language Processing

ROC Curve: Receiver Operating Characteristics Curve

TF-IDF: Term Frequency–Inverse Document Frequency

References

1. Press conferences on COVID-19. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
2. Petrosillo, N.; Viceconte, G.; Ergonul, O.; Ippolito, G.; Petersen, E., COVID-19, SARS and MERS: are they closely related? *Clinical Microbiology and Infection* **2020**, 26 (6), 729-734.
3. Munster, V. J.; Koopmans, M.; van Doremalen, N.; van Riel, D.; de Wit, E., A Novel Coronavirus Emerging in China — Key Questions for Impact Assessment. *New England Journal of Medicine* **2020**, 382 (8), 692-694.
4. Vaccines and immunization: What is vaccination? <https://www.who.int/news-room/q-a-detail/vaccines-and-immunization-what-is-vaccination>.
5. Li, Y.-D.; Chi, W.-Y.; Su, J.-H.; Ferrall, L.; Hung, C.-F.; Wu, T. C., Coronavirus vaccine development: from SARS and MERS to COVID-19. *Journal of Biomedical Science* **2020**, 27 (1), 104.
6. Poland, G. A.; Ovsyannikova, I. G.; Kennedy, R. B., SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *The Lancet* **2020**, 396 (10262), 1595-1606.
7. Is there a vaccine for COVID-19? [https://www.who.int/news-room/q-a-detail/coronavirus-disease-\(covid-19\)-vaccines](https://www.who.int/news-room/q-a-detail/coronavirus-disease-(covid-19)-vaccines)
8. Kata, A., A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine* **2010**, 28 (7), 1709-1716.
9. Puri, N.; Coomes, E. A.; Haghbayan, H.; Gunaratne, K., Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics* **2020**, 16 (11), 2586-2593.
10. Depoux, A.; Martin, S.; Karafillakis, E.; Preet, R.; Wilder-Smith, A.; Larson, H., The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine* **2020**, 27 (3).
11. Ahmed, W.; Vidal-Alaball, J.; Downing, J.; López Seguí, F., COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* **2020**, 22 (5), e19458.
12. Bardus, M.; El Rassi, R.; Chahrour, M.; Akl, E. W.; Raslan, A. S.; Meho, L. I.; Akl, E. A., The Use of Social Media to Increase the Impact of Health Research: Systematic Review. *J Med*

Internet Res **2020**, 22 (7), e15607.

13. In: The Oxford English Dictionary, 2nd ed. 1989. misinformation. Oxford University Press, p.871.

14. Pulido, C. M.; Villarejo-Carballido, B.; Redondo-Sama, G.; Gómez, A., COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International Sociology* **2020**, 35 (4), 377-392.

15. Brennen, J. S.; Simon, F.; Howard, P. N.; Nielsen, R. K., Types, sources, and claims of COVID-19 misinformation. *Reuters Institute* **2020**, 7 (3), 1.

16. Xinhua. 2019. Aug 20, [2020-04-13]. Weibo reports robust Q2 user growth. http://www.xinhuanet.com/english/2019-08/20/c_138323288.htm.

17. Xu, Q.; Shen, Z.; Shah, N.; Cuomo, R.; Cai, M.; Brown, M.; Li, J.; Mackey, T., Characterizing Weibo Social Media Posts From Wuhan, China During the Early Stages of the COVID-19 Pandemic: Qualitative Content Analysis. *JMIR Public Health Surveill* **2020**, 6 (4), e24125.

18. Yuan, Q.; Gao, Q. In *The Analysis of Online News Information Credibility Assessment on Weibo Based on Analyzing Content*, Engineering Psychology and Cognitive Ergonomics, Cham, 2016//; Harris, D., Ed. Springer International Publishing: Cham, 2016; pp 125-135.

19. Feng, Z.; Bo, W.; Yingxue, C., Research on China's city network based on users' friend relationships in online social networks: a case study of Sina Weibo. *GeoJournal* **2016**, 81 (6), 937-946.

20. Rao, Q.; Zhang, Z.; Lv, Y.; Zhao, Y.; Bai, L.; Hou, X., Factors Associated With Influential Health-Promoting Messages on Social Media: Content Analysis of Sina Weibo. *JMIR Med Inform* **2020**, 8 (10), e20558.

21. Yu, S.; Eisenman, D.; Han, Z., Temporal Dynamics of Public Emotions During the COVID-19 Pandemic at the Epicenter of the Outbreak: Sentiment Analysis of Weibo Posts From Wuhan. *J Med Internet Res* **2021**, 23 (3), e27078.

22. The 46th China Statistical Report on Internet Development in China. China Internet Network Information Center. Beijing; 2020 Sep. <https://cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf>.

23. Weibo 2020 User Development Report. <https://weibo.com/ttarticle/p/show?id=2309404613871951282183&sudaref=www.google.com.hk>.

24. COVID-19 vaccine search results on Weibo. https://s.weibo.com/topic?q= &pagetype=topic&topic=1&Refer=Stopic_box.

25. Li, J.; Xu, Q.; Cuomo, R.; Purushothaman, V.; Mackey, T., Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. *JMIR Public Health and Surveillance* **2020**, 6 (2), e18700.

26. Kang, G. J.; Ewing-Nelson, S. R.; Mackey, L.; Schlitt, J. T.; Marathe, A.; Abbas, K. M.; Swarup, S., Semantic network analysis of vaccine sentiment in online social media. *Vaccine* **2017**, 35 (29), 3621-3638.

27. Zhao, Y.; Cheng, S.; Yu, X.; Xu, H., Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *J Med Internet Res* **2020**, 22 (5), e18825.

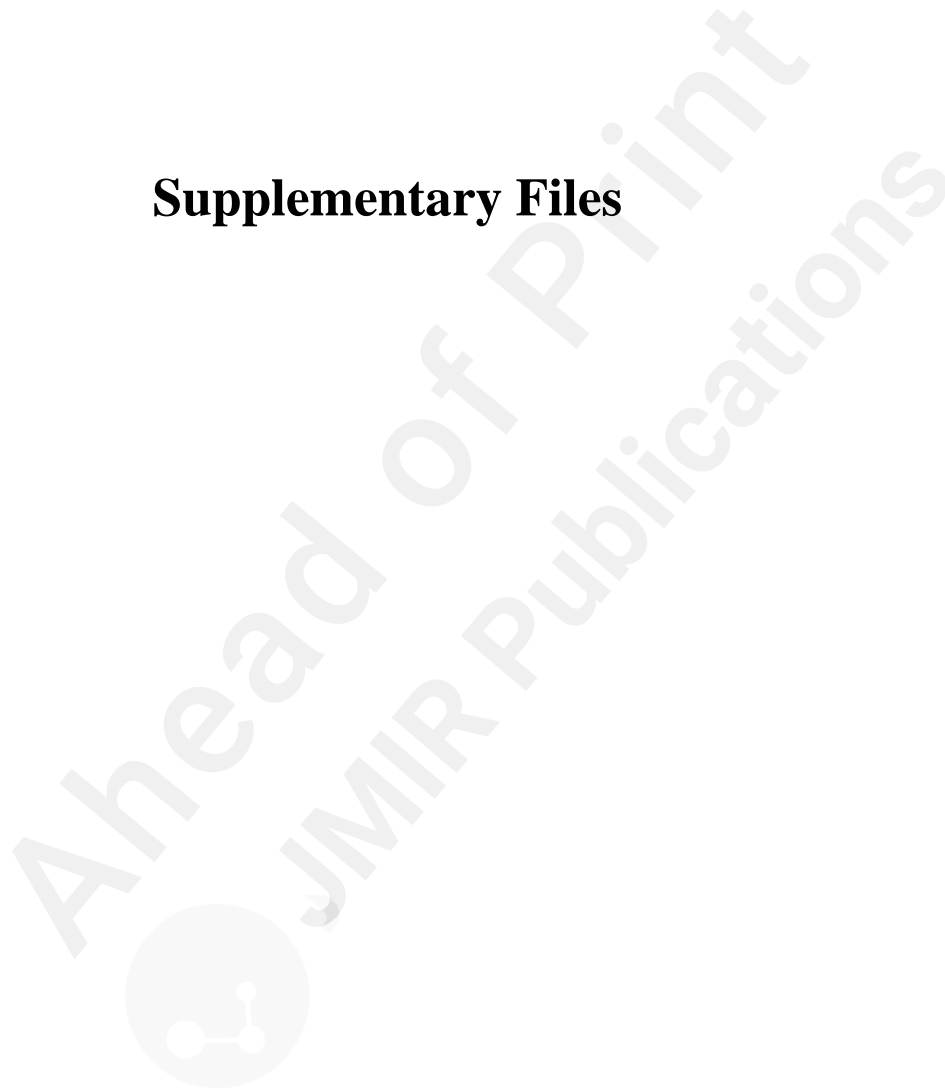
28. Eibensteiner, F.; Ritschl, V.; Nawaz, F. A.; Fazel, S. S.; Tsagkaris, C.; Kulnik, S. T.; Crutzen, R.; Klager, E.; Völkl-Kernstock, S.; Schaden, E.; Kletecka-Pulker, M.; Willschke, H.; Atanasov, A. G., People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. *J Med Internet Res* **2021**, 23 (4), e28973.

29. Jeon, J.; Baruah, G.; Sarabadani, S.; Palanica, A., Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data. *J Med Internet Res* **2020**, 22 (10), e20509.

30. Ali, K. F.; Whitebridge, S.; Jamal, M. H.; Alsafy, M.; Atkin, S. L., Perceptions, Knowledge, and Behaviors Related to COVID-19 Among Social Media Users: Cross-Sectional Study. *J Med Internet Res* **2020**, 22 (9), e19913.
31. Hussain, A.; Tahir, A.; Hussain, Z.; Sheikh, Z.; Gogate, M.; Dashtipour, K.; Ali, A.; Sheikh, A., Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. *J Med Internet Res* **2021**, 23 (4), e26627.
32. Wilson, S. L.; Wiysonge, C., Social media and vaccine hesitancy. *BMJ Global Health* **2020**, 5 (10), e004206.
33. Xie, Y.; Qiao, R.; Shao, G.; Chen, H., Research on Chinese social media users' communication behaviors during public emergency events. *Telematics and Informatics* **2017**, 34 (3), 740-754.
34. Westerman, D.; Spence, P. R.; Van Der Heide, B., Social media as information source: Recency of updates and credibility of information. *Journal of computer-mediated communication* **2014**, 19 (2), 171-183.
35. Chou, W.-y. S.; Hunt, Y. M.; Beckjord, E. B.; Moser, R. P.; Hesse, B. W., Social Media Use in the United States: Implications for Health Communication. *J Med Internet Res* **2009**, 11 (4), e48.
36. Bizid, I.; Nayef, N.; Boursier, P.; Doucet, A., Detecting prominent microblog users over crisis events phases. *Information Systems* **2018**, 78, 173-188.
37. Jin, P., Health and health care in Chinese government manifestos: a content analysis of the Report on the Work of the Government from 1954 to 2016. *The Lancet* **2016**, 388, S43.
38. Python 3.9.2. <https://www.python.org/downloads/release/python-392/>.
39. The Municipal Center for Disease Control and Prevention issued instructions on COVID-19 vaccination. http://www.jiaxing.gov.cn/art/2020/10/19/art_1578778_59000979.html.
40. HUANG Lei, L. S., WANG Jingjing, User-Type Classification in Micro-Blog Based on Information of Authenticated User. *Journal of Frontiers of Computer Science and Technology(in chinese)* **2015**, 9 (6), 719-725.
41. Hartmann, J.; Huppertz, J.; Schamp, C.; Heitmann, M., Comparing automated text classification methods. *International Journal of Research in Marketing* **2019**, 36 (1), 20-38.
42. Yin, F.; Wu, Z.; Xia, X.; Ji, M.; Wang, Y.; Hu, Z., Unfolding the Determinants of COVID-19 Vaccine Acceptance in China. *J Med Internet Res* **2021**, 23 (1), e26089.
43. Alessia, D.; Ferri, F.; Grifoni, P.; Guzzo, T., Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications* **2015**, 125 (3).
44. SnowNLP: Simplified Chinese Text Processing. <https://github.com/isnowfy/snownlp>.
45. Dai, W.-q.; Huang, G.; Pan, W.; Hu, C.; Pan, W.; Liao, S., China Public Psychology Analysis About COVID-19 Under Considering Sina Weibo Data. *Available at SSRN* 3618369 **2020**.
46. Liu, J.; Gao, L., Analysis of topics and characteristics of user reviews on different online psychological counseling methods. *International Journal of Medical Informatics* **2021**, 147, 104367.
47. Liu, J.; Zhou, Y.; Jiang, X.; Zhang, W., Consumers' satisfaction factors mining and sentiment analysis of B2C online pharmacy reviews. *BMC Medical Informatics and Decision Making* **2020**, 20 (1), 194.
48. Jin, H.; Ling, C. X., Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* **2005**, 17 (3), 299-310.
49. Allen, J. G.; Haccoun, D. M., Sex differences in emotionality: A multidimensional approach. *Human Relations* **1976**, 29 (8), 711-722.
50. Brody, L. R.; Hall, J. A., Gender and emotion in context. *Handbook of emotions* **2008**, 3, 395-408.

51. Chaplin, T. M., Gender and emotion expression: A developmental contextual perspective. *Emotion Review* **2015**, 7 (1), 14-21.
52. Liu, X.; Sun, M.; Li, J., Research on gender differences in online health communities. *International journal of medical informatics* **2018**, 111, 172-181.
53. Sun, B.; Mao, H.; Yin, C., Male and female users' differences in online technology community based on text mining. *Frontiers in Psychology* **2020**, 11, 806.
54. Kini, A.; Morgan, R.; Kuo, H.; Shea, P.; Shapiro, J.; Leng, S. X.; Pekosz, A.; Klein, S. L., Differences and disparities in seasonal influenza vaccine, acceptance, adverse reactions, and coverage by age, sex, gender, and race. *Vaccine* **2021**.
55. Quinn, S. C.; Jamison, A.; An, J.; Freimuth, V. S.; Hancock, G. R.; Musa, D., Breaking down the monolith: understanding flu vaccine uptake among African Americans. *SSM-population health* **2018**, 4, 25-36.
56. Banach, D. B.; Ornstein, K.; Factor, S. H.; Soriano, T. A., Seasonal influenza vaccination among homebound elderly receiving home-based primary care in New York City. *Journal of community health* **2012**, 37 (1), 10-14.
57. Roy, M.; Sherrard, L.; Dubé, È.; Gilbert, N. L., Determinants of non-vaccination against seasonal influenza. *Health reports* **2018**, 29 (10), 12-22.
58. Limaye, R. J.; Sauer, M.; Ali, J.; Bernstein, J.; Wahl, B.; Barnhill, A.; Labrique, A., Building trust while influencing online COVID-19 content in the social media world. *The Lancet Digital Health* **2020**, 2 (6), e277-e278.
59. Zhang, Y.-t.; Li, R.-t.; Sun, X.-j.; Peng, M.; Li, X., Social Media Exposure, Psychological Distress, Emotion Regulation, and Depression During the COVID-19 Outbreak in Community Samples in China. *Frontiers in Psychiatry* **2021**, 12 (528).
60. Lin, Y.; Hu, Z.; Zhao, Q.; Alias, H.; Danaee, M.; Wong, L. P., Understanding COVID-19 vaccine demand and hesitancy: A nationwide online survey in China. *PLoS neglected tropical diseases* **2020**, 14 (12), e0008961.
61. Wang, K.; Wong, E. L.-Y.; Ho, K.-F.; Cheung, A. W.-L.; Yau, P. S.-Y.; Dong, D.; Wong, S. Y.-S.; Yeoh, E.-K., Change of willingness to accept COVID-19 vaccine and reasons of vaccine hesitancy of working people at different waves of local epidemic in Hong Kong, China: Repeated cross-sectional surveys. *Vaccines* **2021**, 9 (1), 62.
62. Qin, W.; Wang, E.; Ni, Z., Chinese consumers' willingness to get a COVID-19 vaccine and willingness to pay for it. *Plos one* **2021**, 16 (5), e0250112.
63. Han, K.; Francis, M. R.; Zhang, R.; Wang, Q.; Xia, A.; Lu, L.; Yang, B.; Hou, Z., Confidence, Acceptance and Willingness to Pay for the COVID-19 Vaccine among Migrants in Shanghai, China: A Cross-Sectional Study. *Vaccines* **2021**, 9 (5).
64. Kreps, S.; Dasgupta, N.; Brownstein, J. S.; Hswen, Y.; Kriner, D. L., Public attitudes toward COVID-19 vaccination: The role of vaccine attributes, incentives, and misinformation. *npj Vaccines* **2021**, 6 (1), 73.
65. Xiaomei, Z.; Jing, Y.; Jianpei, Z.; Hongyu, H., Microblog sentiment analysis with weak dependency connections. *Knowledge-Based Systems* **2018**, 142, 170-180.
66. Bamman, D.; Eisenstein, J.; Schnoebelen, T., Gender in twitter: Styles, stances. *andsocial networks. Technical Report 1210.4567, arXiv* **2012**.

Supplementary Files



Figures

