

Original Paper

Characterization of Vaccine Tweets During the Early Stage of the COVID-19 Outbreak in the United States: Topic Modeling Analysis

Li Crystal Jiang¹, PhD; Tsz Hang Chu¹, MPhil; Mengru Sun², BA

¹Department of Media and Communication, City University of Hong Kong, Hong Kong, Hong Kong

²College of Media and International Culture, Zhejiang University, Hangzhou, China

Corresponding Author:

Li Crystal Jiang, PhD

Department of Media and Communication

City University of Hong Kong

M5082, Run Run Shaw Creative Media Centre

18 Tat Hong Avenue, Kowloon

Hong Kong

China (Hong Kong)

Phone: 852 034429332

Email: crystal.jiang@cityu.edu.hk

Abstract

Background: During the early stages of the COVID-19 pandemic, developing safe and effective coronavirus vaccines was considered critical to arresting the spread of the disease. News and social media discussions have extensively covered the issue of coronavirus vaccines, with a mixture of vaccine advocacies, concerns, and oppositions.

Objective: This study aimed to uncover the emerging themes in Twitter users' perceptions and attitudes toward vaccines during the early stages of the COVID-19 outbreak.

Methods: This study employed topic modeling to analyze tweets related to coronavirus vaccines at the start of the COVID-19 outbreak in the United States (February 21 to March 20, 2020). We created a predefined query (eg, "COVID" AND "vaccine") to extract the tweet text and metadata (number of followers of the Twitter account and engagement metrics based on likes, comments, and retweeting) from the Meltwater database. After preprocessing the data, we tested Latent Dirichlet Allocation models to identify topics associated with these tweets. The model specifying 20 topics provided the best overall coherence, and each topic was interpreted based on its top associated terms.

Results: In total, we analyzed 100,209 tweets containing keywords related to coronavirus and vaccines. The 20 topics were further collapsed based on shared similarities, thereby generating 7 major themes. Our analysis characterized 26.3% (26,234/100,209) of the tweets as *News Related to Coronavirus and Vaccine Development*, 25.4% (25,425/100,209) as *General Discussion and Seeking of Information on Coronavirus*, 12.9% (12,882/100,209) as *Financial Concerns*, 12.7% (12,696/100,209) as *Venting Negative Emotions*, 9.9% (9908/100,209) as *Prayers and Calls for Positivity*, 8.1% (8155/100,209) as *Efficacy of Vaccine and Treatment*, and 4.9% (4909/100,209) as *Conspiracies about Coronavirus and Its Vaccines*. Different themes demonstrated some changes over time, mostly in close association with news or events related to vaccine developments. Twitter users who discussed conspiracy theories, the efficacy of vaccines and treatments, and financial concerns had more followers than those focused on other vaccine themes. The engagement level—the extent to which a tweet being retweeted, quoted, liked, or replied by other users—was similar among different themes, but tweets venting negative emotions yielded the lowest engagement.

Conclusions: This study enriches our understanding of public concerns over new vaccines or vaccine development at early stages of the outbreak, bearing implications for influencing vaccine attitudes and guiding public health efforts to cope with infectious disease outbreaks in the future. This study concluded that public concerns centered on general policy issues related to coronavirus vaccines and that the discussions were considerably mixed with political views when vaccines were not made available. Only a small proportion of tweets focused on conspiracy theories, but these tweets demonstrated high engagement levels and were often contributed by Twitter users with more influence.

(*JMIR Infodemiology* 2021;1(1):e25636) doi: [10.2196/25636](https://doi.org/10.2196/25636)

KEYWORDS

topic modeling; social media; infoveillance; vaccine; coronavirus; COVID-19

Introduction

Background

The COVID-19 pandemic has affected more than 200 countries and territories, killed more than 1.2 million people, devastated the global economy, and disrupted the daily life of billions of people [1]. Owing to the lack of effective containment measures during the early stages of the COVID-19 outbreak, many of those heavily affected placed their hope on the development of coronavirus vaccines. Ever since the early stages of the outbreak, extensive news coverage followed the progress of vaccine developments, while web users engaged in heated discussions about coronavirus vaccines or vaccines in general on various social media platforms such as Facebook, Twitter, and Instagram [2-4]. It is crucial to understand media portrayals and public discussions of coronavirus vaccines during the early stages of the outbreak because they influenced policy-making in public health and public perceptions of and attitudes toward vaccination in the later stage [5-11]. A comprehensive understanding of the public opinion during the initial phase of infectious outbreaks will inform how public health professionals and policymakers make decisions in addressing public concerns in future outbreaks of infectious diseases [12].

Infodemic and Early Stages of Outbreaks

Frequent infectious outbreaks are an ongoing reality for globalized societies, and the early stage of an outbreak is always challenging. The beginning of an outbreak is typically characterized by a lack of accuracy, widespread misinformation, as well as heightened uncertainty and fear among the general public [13,14]. In the first couple of months of the COVID-19 pandemic, policymakers had limited knowledge about coronavirus and largely relied on data modeling for predictions and decisions. Similarly, owing to the lack of knowledge, there was little consensus among media professionals, public health professionals, and politicians over containment measures [15]. Instead, geopolitical discourses, conspiracy theories, and racial bigotry created significant amounts of noise for officials trying to manage the pandemic [16-19]. All of these issues brought intensified fear and anxiety to the public.

Social media platforms shape public experience and opinions, while also serving as platforms for public health. During the initial phase of the pandemic, social media became the hotspot of all sorts of issues for the pandemic. Previous studies have shown that social media content about COVID-19 is mixed with a deluge of stigmas, rumors, and misinformation [16-18] and is highly biased by political and social ideologies [19-21]. On February 15, 2020, the World Health Organization officially coined a phenomenon “infodemic,” which refers to the rapid spread of misinformation through social media platforms and other outlets on a global scale [22-25]. An infodemic is a serious threat to public health as it greatly advocates hostile attitudes toward preventive measures and complicates our fight with the COVID-19 pandemic [26].

COVID-19, Vaccines, and Social Media

Despite the scientific consensus that vaccination is a safe and effective approach to prevent infectious diseases, there is more controversy over the use of vaccines than over other preventive measures (eg, hand hygiene, social distancing). These concerns include fear of side effects, uncertainty about vaccine efficacy, and general mistrust of the sciences and the government. These contentions have resulted in vaccine hesitancy, declines in immunization, and even small outbreaks of vaccine-preventable diseases [27-29]. Controversies over vaccination have often manifested in social media communities, leading to increasing research investigating the spread of information and opinions about vaccines on various social media platforms. This inquiry mainly focuses on the intensified competition between provaccination and antivaccination views on social media in recent years [30]. Both manual coding and computational methods have identified similar proportions of provaccination and antivaccination content on YouTube and Twitter [31-33], but antivaccination content—produced by closely connected communities and employing sophisticated antivaccination advocacy strategies—often outweigh the provaccination content [34,35]. There is some variation across specific types of vaccines. For example, influenza-related videos contain more anti-immunization content compared to videos on measles, presumably because influenza vaccination is normally perceived as new and less efficacious [32].

Scholars propose several strategies for tackling the vaccine controversy and addressing antivaccination information on social media [23,24], such as infoveillance. Infoveillance is an emerging approach that tracks what people do and write on the internet to reflect public opinions, behaviors, knowledge, and attitudes related to health issues [36]. Major applications of infoveillance include but are not limited to monitoring health-relevant messages on the internet (eg, antivaccination sites), outlining web-based health information availability (eg, vaccine advocacies), and analyzing search engine queries to predict disease outbreaks (eg, syndromic inquiry). By analyzing social media posts related to public health issues, previous studies have successfully performed surveillance on public opinions and public sentiments [36-40], predicted prevalence and mortality across time and space [41,42], and explained how intended or unintended behavioral responses are shaped by social networks and other information features [43-45]. In the case of analyzing vaccine-related social media messages, infoveillance can provide key stakeholders (eg, health organizations, governments) the benefits of revealing public concerns over vaccines and monitoring public sentiments in real-time. It also helps identify influencers and advocates, directly engages with the vaccine targets (ie, people who are at high risk of infection), and manage misinformation and hostile messages efficiently. Infoveillance is particularly powered by big data and computational techniques as they offer very useful tools for understanding social media content in an unstructured, bottom-up manner. Previous studies have successfully used computational methods to examine public perceptions on

influenza vaccine [46], human papilloma virus vaccines [47], and childhood vaccinations [48,49].

This study aims to investigate the discussion related to coronavirus vaccines on Twitter during the early stage of the COVID-19 outbreak in the United States (February 20, 2020 to March 31, 2020). This study will contribute to our understanding of coronavirus vaccines and connections to attitudes related to vaccines by tracking back to the initial public concerns. The findings of this study will elucidate the public discussions on new vaccines or vaccines under development, and the concerns and issues revealed in this study can show the implications on public health efforts in coping with infectious disease outbreaks in the future. Provided that the coronavirus vaccines show plenty of uncertainty in efficacy and effectiveness, using unsupervised learning methods, we aim to explore the main themes that emerged from the tweets related to coronavirus vaccines during the initial stage of the pandemic in the United States (RQ1). We also seek to examine how these themes evolved over time (RQ2).

Out of the different types of misinformation, conspiracy theories have merged as a significant concern in the “social media infodemic.” Since the COVID-19 pandemic, several studies have analyzed certain types of conspiracy theories such as the coronavirus as a bioweapon [49], the 5G coronavirus [50], or “Film Your Hospital” [51]. However, few studies have looked at the overall spread of conspiracy theories related to COVID-19. A recent analysis of German tweets indicated that less than 1% of the tweets analyzed were related to conspiracy theories, although partisanship boosted the spread of conspiracy theory tweets [52]. This study examined how conspiracy theories related to coronavirus vaccines were represented in the American tweets at the early stage of the outbreak (RQ3).

Previous studies also indicate that the spread of conspiracy theories and antivaccination messages follow a different pattern compared to that of provaccination messages. On social media, antivaccination content, in general, attracts more likes and engages more discussion because content producers are inclined to use a variety of persuasive strategies (eg, health narratives) and present antivaccination in the form of public criticism aggressively [30,53]. Antivaccination messages are normally produced by a small proportion of powerful influencers, but antivaccination supporters perpetuate echo chambers by actively spreading conspiracy theories and misinformation through a more decentralized network [34]. This study also expected some differences in the influences (ie, number of followers) and engagement levels when comparing different themes in Twitter vaccine discussions. Specifically, compared to the tweets discussing other vaccine-related themes, tweets discussing conspiracy theories were likely contributed by Twitter users with more followers (H1a) and produced more engagement than tweets that discuss other themes (H1b).

Methods

Data Source

The study period was set from February 20 to March 31, 2020. We marked this period as the early stage because it corresponded

to a sharp increase in the coronavirus case count and death toll in the United States (eg, over 181,000 cases and 3606 deaths by March 31, 2020). At the end of March 2020, the United States became the country with the most number of confirmed cases in the world. Moreover, in March 2020, most state and local governments declared COVID-19 as a public health emergency, issued stay-at-home orders, and mandated closures of schools and public meeting places [54]. We purposely chose this time frame to capture the tweets during the first phase of the COVID-19 outbreak in the United States. Meltwater [55], a commercial web-based media monitoring service, was used for data collection. Meltwater has access to the full Twitter pipelining data hosting service, providing customized reporting options with the last 15 months of Twitter history. Meltwater geotagged each tweet using the user’s Twitter bio-related or other geo-related information, thus ensuring that all tweets included in the sample were posted by American Twitter users.

Using the social media monitoring and data collection platform provided by Meltwater, we collected tweets originating from the United States and written in English that were related to the coronavirus vaccine by using the following Boolean query: (covid OR coronavirus) AND (vaccine OR vaccines OR vaccination OR vaccinations OR vaccinate OR vax OR vaxine OR vaxx OR vaccinated). Using this strategy, we identified 117,718 tweets (including original tweets and quote tweets but not replies and retweets). The text of the tweet and relevant metadata, including username, date of the post, and follower count, were stored. We also stored the engagement metric provided by Meltwater, which was a composite score representing how many times a tweet was retweeted, quoted, liked, or prompted a reply by other users. A higher engagement value indicated that the tweet received more attention by other Twitter users.

Topic Modeling

To analyze the obtained data set, we applied topic modeling—an unsupervised machine learning algorithm that allows researchers to uncover hidden thematic structures in a sizable collection of documents [56]. A topic model can “produce a set of interpretable topics (groups of words that are associated under a single theme) and assess the strength with which each document exhibits those topics” [57]. In this study, we used Latent Dirichlet Allocation (LDA), one of the widely used topic models that groups words that frequently co-occur in documents into various topics. By providing the text input and setting the desired number of topics, LDA automatically produces a set of topics, words are allocated to the topics, and the topic proportions are attributed for each document [58]. We decided to use LDA, as findings yielded by prior studies indicate that it performs well with both long and short texts. In addition, it has been previously used to examine COVID-19–related discussions on Twitter [59].

Data Preprocessing

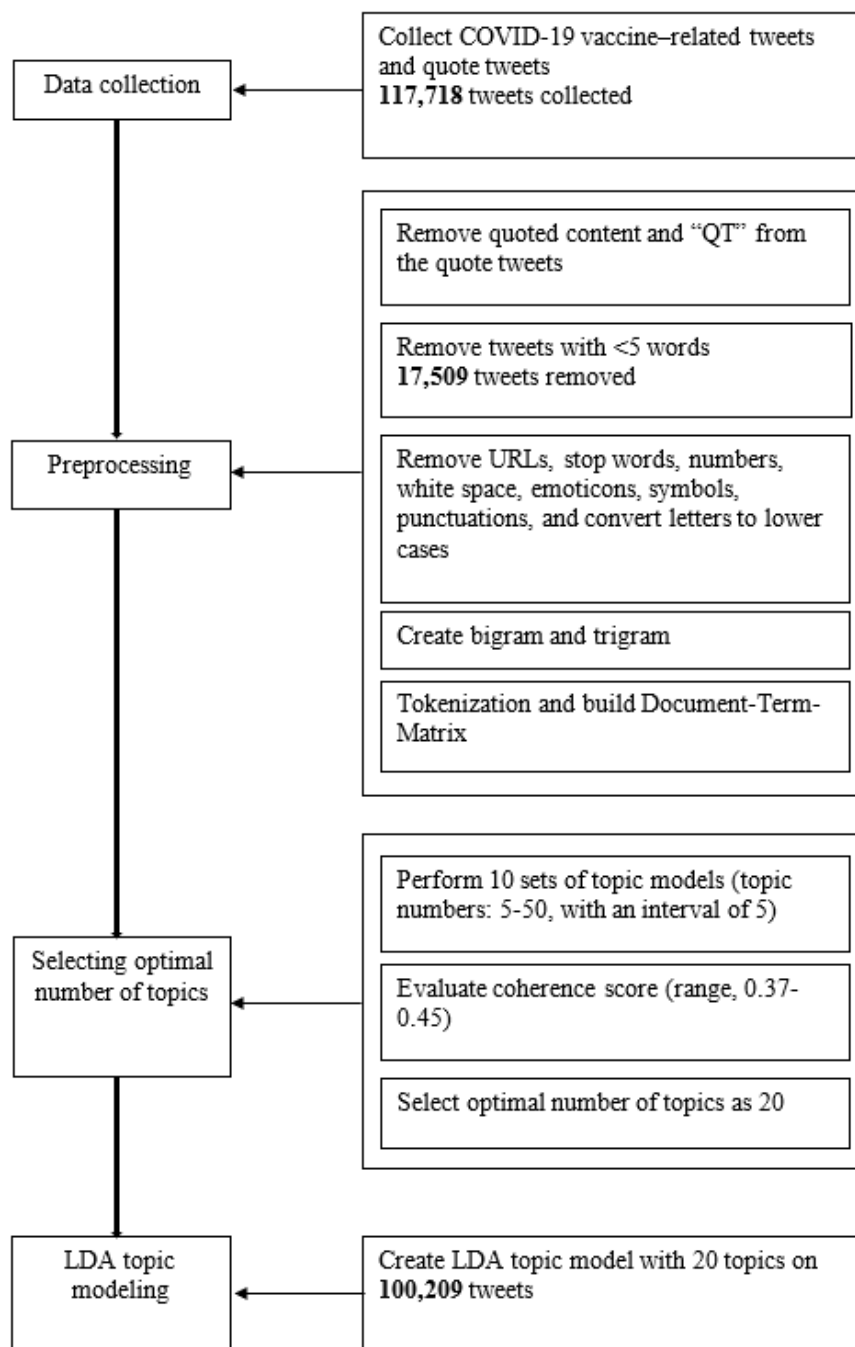
To prepare the corpus for LDA topic modeling, we first removed the quoted content within the quote tweets and the “QT” (meaning a quote tweet) to retain only the original content of the tweet. As the length of document plays a significant role in the topic modeling method [60], tweets with fewer than 5 words

were removed, leaving a total sample of 100,209 tweets. Following this, all the URLs within the tweets were removed. Next, the tweets were preprocessed using standard natural language processing practice [61]. We converted all the letters to lower case, removed all the stop words (eg, the, it, that), lemmatized the words, and removed numbers, white space, emoticons, symbols, and punctuation, with the use of Python packages such as NLTK (Apache) [62] and spaCy (Explosion AI) [63]. Bigram and trigram were also created and added. After tokenization, Document-Term-Matrix was built and used for the LDA topic modeling.

Number of Topics

To determine the optimal number of topics for this tweet set, we performed 10 sets of topic models with topic numbers ranging from 5 to 50 (with intervals of 5) by implementing the LDA model from the Python package MALLET. The topic coherence—a metric focusing on the interpretability—of the 10 topic models were then calculated and evaluated for selecting the appropriate number of topics [64,65]. We decided to use the topic model with 20 topics in this study because it presented the highest topic coherence as compared with the other candidate models. Figure 1 presents the steps of data processing and creating topic models.

Figure 1. Data processing and analysis flowchart. LDA: Latent Dirichlet Allocation; QT: quote tweet.



Topic Interpretation and Further Analyses

The output of the LDA topic model based on 20 topics was reviewed. Although the LDA model presumed that each document contained a mixture of topics and the model produced a probability topic distribution for each topic, we considered only the dominant topic, that is, the topic with the highest probability in that document, and categorized each tweet subject to its dominant topic [66,67]. We then reviewed the 20 top-associated terms, together with the top 5 tweets with the highest topic percentage contribution of each topic, before labeling each topic. These labels were based on the authors' background knowledge regarding vaccine hesitancy as well as the observation of coronavirus vaccine-related news and user-generated opinions on Twitter during the data collection and analysis [68]. The 3 authors involved in this study independently labeled the topics, and the resulting 3 sets of topic labels were compared. The diverse topic labels were discussed and 100% agreement between the authors was reached. The labeled topics were further grouped into distinct themes deductively following discussion. Lastly, differences between the themes on the number of followers and levels of engagement were examined. Nonparametric tests were used, as the outcome variables were not normally distributed within the current data.

Results

Topic Modeling

We analyzed 100,209 tweets in this study. The average number of followers of each tweet was 19,300.62 (SD 431,794.41), and the average engagement value of the tweets was 29.41 (SD 624.91). To examine themes that have emerged in coronavirus vaccine discourses on social media (RQ1), LDA modeling with 20 topics was performed. During the labeling process, it was noticed that 4 of the topics were related to news concerning human trials and testing of coronavirus vaccines. As the 4 topics were similar and closely related, the 3 authors agreed to merge these discussions into 1 overall topic, that is, *News of Vaccine Development*. Next, the remaining 17 topics were organized into 7 themes. The themes, topic labels, and associated words for each topic are presented in Table 1. The majority of the tweets were labeled as *News Related to the Coronavirus and Vaccine Development* (26,234/100,209, 26.2%) and *General Discussion and Seeking of Information on the Coronavirus* (25,425/100,209, 25.4%), followed by *Financial Concerns* (12,882/100,209, 12.9%), *Venting Negative Emotions* (12,696/100,209, 12.7%), *Prayers and Calls for Positivity* (9908/100,209, 9.9%), *Efficacy of Vaccines and Treatments* (8155/100,209, 8.1%), and *Conspiracies about Coronavirus and Its Vaccines* (4909/100,209, 4.9%).

Table 1. Themes and topics from coronavirus vaccine discussions on Twitter.

Themes/topics of discussion	Associated words	Frequency of discussion (N=100,209)	Percentage of discussion	Examples of tweets
News related to coronavirus and vaccine developments				
News of vaccine developments	human, trial, begin, volunteer, receive	17,435	17.4%	As said by the US authorities, the first clinical trial of COVID-19 vaccine on humans has been planned to begin today. The first human subject is going to get the dose today.
News of US government research funding/plans for the pandemic	research, fund, system, medical, government	4012	4.0%	The White House approved the emergency fund to deal with COVID-19 in the United States and abroad. The fund will support the development of the COVID-19 vaccine by providing money for new equipment as well as supplies.
News of research plans for vaccines	pandemic, develop, effort, outbreak, step	4787	4.8%	COVID-19: Mainland China has taken a new step for developing the vaccine. A team from around the world will investigate the initial results on youngsters.
General discussion and seeking of information on coronavirus				
Seeking of information on vaccines	question, understand, information, real, cure	3460	3.5%	I would like to know how the coronavirus vaccine interacts with the flu shot. Although I am not that clever to tell if we should be concerned about it, I want to raise this question out of curiosity.
Discussion about coronavirus trend	case, number, low, current, increase	3268	3.3%	This will never work. For example, there is a rising number of confirmed cases in the Republic of Korea and Taiwan after the ease of restrictions. There is going to be nonstop waves of infection if there is no vaccine. The main purpose of isolation is reducing the load on the health care system.
Discussion about coronavirus and its vaccines	virus, spread, vaccine, fast, mutate	4860	4.8%	Although the mutation of coronavirus is much slower than that of the flu viruses, it is an RNA virus, which normally mutates nearly 100 times faster than viruses based on DNA. It will be much difficult to control or vaccinate in the future if millions of people are infected by it as it will provide more chance for the coronavirus to mutate.
Comparisons with influenza	flu, kill, deadly season, thousand	10,145	10.1%	First, we have the flu vaccine already. Second, compared with the influenza and the Spanish flu that have caused over 50 million deaths, the coronavirus seems more infectious. Third, compared with that with the Spanish flu, the death rate with the coronavirus is higher. Fourth, while the influenza virus has more impact on individuals older than 65 years, the coronavirus does not discriminate individuals according to age.
Preventive measures	protect, hand, safe, force, home	3692	3.7%	The following steps can help in defeating COVID-19: stay calm and keep washing your hands with water and soap or use hand sanitizer. Keep social distancing, open doors with your elbow, and do not rub your nose, face, or shake hands with others.
Financial concerns				

Themes/topics of discussion	Associated words	Frequency of discussion (N=100,209)	Percentage of discussion	Examples of tweets
Disparity in income	American, rich, poor, capitalism, afford	5653	5.6%	We, the taxpayers, are going to pay for the research on COVID-19 vaccines, which we deliver to the select few without any compensation. Rich people can acquire billions from tax cuts and chief executive officers can acquire millions from compensation. The capitalism of the Republican Party is socialism for the rich. We all are the targets.
Price of vaccine	free, affordable, cost, charge, insurance	7229	7.2%	The COVID-19 vaccine should be free of charge for people who do not have enough money for copayment for insurance or those who do not have medical insurance. The fee of my vaccination will be covered by my insurance, and I am able to pay for the difference. We have to ensure that the health insurance companies pay their part first.
Efficacy of vaccines and treatments				
Efficacy of vaccines	prevent, cancer, infect, immunity, antibody	4096	4.1%	A lot of people don't know about the COVID-19 vaccines. They are not injecting your body with the dead virus but harmless spikes. Immunity will be built to the spikes after injecting the vaccine. This could ease the worries of those opposing the vaccine.
Efficacy of treatments/preventions	test, treatment, effective, hospital, prove	4059	4.1%	Lately, many physicians from the United States and France have asserted the effectiveness of antimalarial medication in treating COVID-19. Does it mean that the malaria vaccine would work against COVID-19 also?
Conspiracies about coronavirus and its vaccines				
Conspiracies related to companies/stock/government	profit, market, stock, government, attempt	4909	4.9%	Is it possible that the Republican Party and Trump manipulated the stock market and profited through insider trading of Moderna's stock? This biotechnology company, which invented the new vaccine, had its stock increased by 15%.
Venting negative emotions				
Negative emotions (toward Trump and big pharmacies)	wrong, damn, business, stupid, idiot	7019	7.0%	The vaccine makers could create whatever they want. Even if someone got injured or died, we cannot sue them. If someone dies, that's just bad luck. If some child dies, that's just bad luck. If someone becomes paralyzed, that's just bad luck. The profits of the pharmacies grow because we never fight back. We are just the slaves of the big pharmacies.
Trump-related	trump, lie, truth, blame, reality	5677	5.7%	Agreed. What Trump and his incompetent administration do is to lie about everything: the seriousness of the disease, keeping the disease on a tight rein already, getting a vaccine soon.
Prayers and calls for positivity				
Emotions/prayers	good, hope, happen, pretty, remember	5228	5.2%	That is some good news! All of us need some optimism.
Calls for positivity	great, love, call, idea, good	4680	4.7%	I enjoy seeing the positivity in the current state.

Themes and Topics From Coronavirus Vaccine Discussions on Twitter

News Related to the Coronavirus and Vaccine Development

During the COVID-19 pandemic, social media users frequently shared news related to coronavirus as well as the development of coronavirus vaccines. There were 3 topics under this theme: *News of Vaccine Development*, *News of US government Research Funding/Plans for the Pandemic*, and *News of Research Plans for Vaccines*. Tweets categorized in this theme included general news on the progress of human trials and vaccine development across different countries (eg, Germany), announcements of US government funding for scientists and companies conducting research, and upcoming prevention plans for the pandemic released by official bodies as well as coronavirus vaccine research across the globe (eg, “The White House has permitted an emergency funding of US \$1 billion overall in order to fight the COVID-19 outbreak. The emergency fund will offer resources as well as financial support for COVID-19 vaccine development for the states”).

General Discussion and Seeking of Information on the Coronavirus

A total of 5 topics were grouped under this theme: *Seeking of Information on Vaccines*, *Discussion of the Coronavirus and Its Vaccines*, *Discussion of Coronavirus Spread and Infection Trends*, *Comparisons with Influenza*, and *Preventive Measures*. The coronavirus was often compared with the influenza virus in terms of death rate, speed of transmission, and so on (eg, “Up till now, there is no cure for COVID-19 but only treatment for the symptoms. The long-term plan is to invent a new vaccine; yet, there would be no vaccine available in the next couple of months”). The importance of preventive measures, including handwashing and social distancing, was also stressed because there is currently no vaccine nor effective treatment for the coronavirus infection (eg, “We all need to get rid of bad habits. Stop touching your face when you are in public space. Scratch your nose only after washing your hands or scratch it with your sleeve. And remember to wash your hands once you get home”).

Financial Concerns

There were 2 topics under this theme: *Disparity over Income and Price of the Vaccines*. In the topic *Disparity over Income*, conversations were related to the gap between the rich and the poor during the pandemic as well as the differences in access to future coronavirus vaccines (eg, “All this is turning into a class war now. Only rich people can get the COVID-19 vaccine as none of us can be sure that the vaccine will be affordable for everyone”). Worries of inequality brought about by capitalism in obtaining vaccination were also expressed (eg, “Capitalism should never get closed to health care systems. The operating costs of the traditional Medicare and the administrative costs of the US health spending is extremely high. The new vaccine should be free for everyone”). As for the price of vaccination, “free” instead of “affordable” coronavirus vaccines for all Americans were urged (eg, “Citizens who were not able to pay for the COVID-19 vaccine will just keep spreading the disease.

The COVID-19 vaccine should be affordable or free for everyone!”).

Venting Negative Emotions

This theme had 2 topics: *Negative Emotions (toward Trump and big pharmaceutical companies)* and *Trump-related frustrations*. Negative emotions, including anger and disappointment, toward Donald Trump or big pharmaceutical companies, were presented, as those Twitter users believed that Trump and Big Pharma were trying to profit from the pandemic. Additionally, negative emotions were expressed toward Trump explicitly owing to claims he made that are believed to have been mistaken, such as the claim that receiving the influenza vaccine would prevent COVID-19 (eg, “He [Trump] actually believed that a flu shot could fight COVID-19. I do not understand how people with brains elected this guy”).

Prayers and Calls for Positivity

The 2 topics under this theme were *Emotional Expressions/Prayers* and *Calls for Positivity*. Tweets allocated within these 2 topics included messages that aimed to encourage others during the pandemic, expressed hopes for and needs for effective coronavirus vaccines, and hopes for an end to the pandemic (eg, “Let us hope that the COVID-19 situation will be resolved when we have a vaccine/cure for it!”).

Efficacy of the Vaccine and Treatment

The 2 topics under this theme were *Efficacy of Vaccine* and *Efficacy of Treatment/Prevention*. These topics stressed the uncertainties of how well the vaccines for coronavirus work as well as the effectiveness of the current treatment and prevention strategies (eg, “I have learnt from some journals that medicines such as chloroquine, hydroxychloroquine, and azithromycin could be used as treatment or prophylaxis of COVID-19. I hope such treatments can help buying us time while getting a vaccine for COVID-19”).

Conspiracies About Coronavirus and Its Vaccines

There were different conspiracies about coronavirus and its vaccines on social media (RQ3). Many of these were related to the companies developing coronavirus vaccines, stock markets, as well as the government. For example, some tweets were claiming that the coronavirus vaccine would contain a microchip that would allow the government or company to track the vaccine receivers (eg, “Once the COVID-19 vaccines are launched, people will be motivated by fear to receive the vaccines that have microchips in it.”). There were also claims that the US government spread the coronavirus deliberately and withheld the coronavirus vaccines (eg, “I think the US has the cure already because it invented this bioweapon. It does not want other parties to wreck its cautiously crafted plans for devastation and racketeering”).

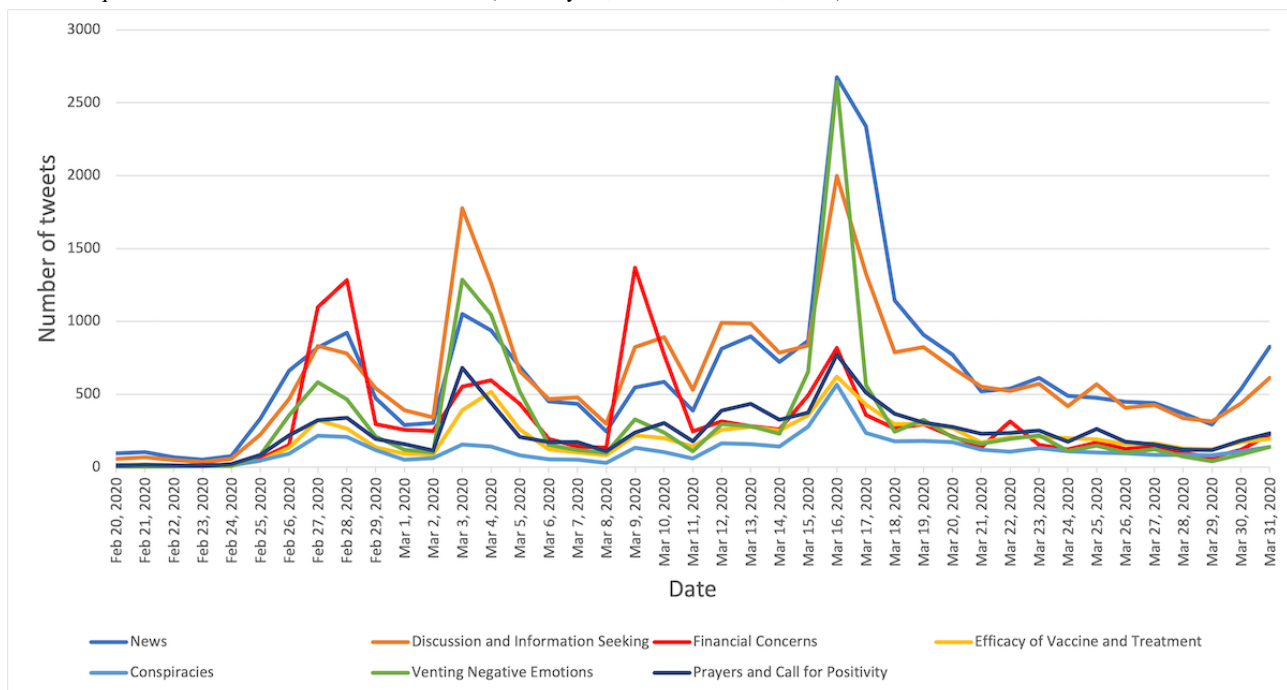
Themes Across Time

Figure 2 shows the changes in the coronavirus vaccine-related discussions on Twitter based on the themes identified across the data collection period, that is, February 20, 2020 to March 31, 2020 (RQ2). Among the 7 themes identified, *News Related to Coronavirus and Vaccine Development* and *General Discussion and Seeking of Information on Coronavirus* were

the most frequently presented overall. Coronavirus vaccine-related discourses on Twitter were promoted by breaking news or announcements and speeches made by the governments and political elites. As shown in Figure 2, there were several peaks in the coronavirus vaccine-related discussions at the early stage of the outbreak. The discussions were elevated on March 16, 2020, which corresponded to the trending news that the Trump administration was attempting to offer large sums of money to a German company in exchange for exclusive access to a possible coronavirus vaccine on March 15, 2020. Coronavirus vaccine discourses regarding *Financial Concerns* reached the peak and exceeded other themes on February 28, 2020 and March 9, 2020. The rising discussions about the prices and affordability of the coronavirus vaccines were related to Health and Human Services Secretary Alex Azar's refusal of promising affordable coronavirus vaccines for all US citizens on February 27, 2020 and Bernie Sanders'

promises of free coronavirus vaccine for all Americans on March 9, 2020, respectively. We observed some co-occurring patterns across the themes during the same peaks and time periods. General discussions and information on coronavirus highly mirrored the themes of news related to coronavirus and vaccine developments. The expression of negative emotions also increased when the discussions of these 2 themes reached a spike. There were also some observed differences in the themes across time. The efficacy of the vaccine and treatment, conspiracies about the coronavirus and vaccines, and prayers and calls for positivity appeared more periodically, while other themes (ie, news related to coronavirus and vaccine developments, general discussion and information on coronavirus, financial concerns, venting negative emotions) were more episodic, featured with several peaks instigated by breaking news or events related to vaccine developments.

Figure 2. Frequencies of themes of the tweets over time (February 20, 2020 to March 31, 2020).



Differences in the Follower Numbers and Engagement Level

Analyses of Follower Numbers and Engagement Level

H1a and H1b hypothesized that conspiracy tweets' contributors had more followers, and conspiracy tweets received higher levels of engagement than tweets with other themes. To examine the differences between themes in the number of followers and levels of engagement, further analyses were performed. First, the results of classification of LDA topic modeling were attached in the original data (which contained the metadata, including the number of followers and engagement metric of each tweet).

The data were then entered into the SPSS software (IBM Corp). Next, we created a categorical variable according to the themes and used it as the independent variable to examine the differences in the number of followers and levels of engagement across themes by using the Kruskal-Wallis H test. Post-hoc analysis was also performed using the Bonferroni-corrected Dunn test. As the hypotheses focus on the difference between conspiracy tweets and the other tweets with themes that presented attitudes and concerns toward the coronavirus vaccines, tweets labeled as news or discussion/information seeking were excluded from the analyses. Table 2 presents the median and mean rank of numbers of followers and levels of engagement.

Table 2. Median and mean ranks of the followers and engagement among themes.

Themes	Followers		Engagement	
	Median	Mean rank	Median	Mean rank
Financial concerns	685	24,411.87 ^a	1	20,836.52 ^a
Efficacy of vaccines and treatments	740	24,957.68 ^{a,b}	1	20,784.34 ^a
Conspiracies about coronavirus and vaccines	770	25,095.54 ^b	1	20,631.67 ^a
Venting negative emotions	616	23,615.43 ^c	0	19,275.31 ^b
Prayers/calls for positivity	620	23,776.11 ^c	1	20,807.51 ^a

^{a-c}Same superscripts in the same column indicate no significant statistical differences ($P>.05$); different superscripts in the same column indicate significant statistical differences ($P<.05$).

Followers

The results of our study suggested that there were significant differences in the number of followers between different themes ($\chi^2_4=77.8$, $P<.001$). The post-hoc test further suggested that conspiracy tweets were more likely to be posted by users with a large number of followers than the tweets classified as *Venting Negative Emotions* ($P<.001$), *Prayers and Calls for Positivity* ($P<.001$), and *Financial Concerns* ($P=.04$). Tweets that discussed the efficacy of vaccines and treatments were also more likely to be posted by users with a large number of followers than the tweets classified as *Venting Negative Emotions* ($P<.001$) and *Prayers and Call for Positivity* ($P<.001$). Similarly, tweets expressing financial concerns were more likely to be posted by users with a large number of followers than the tweets classified as *Venting Negative Emotions* ($P<.001$) and *Prayers and Calls for Positivity* ($P=.007$). As conspiracy tweets were more likely to be posted by users with more followers than the tweets identified as *Financial Concerns*, *Venting Negative Emotions*, and *Prayers and Calls for Positivity*, H1a was partially supported.

Engagement Levels

The results of our study suggested that there were significant differences in the levels of engagement between different themes ($\chi^2_4=155.8$, $P<.001$). The post-hoc test further suggested that tweets classified as *Venting Negative Emotions* significantly received lower levels of engagement than tweets classified as *Conspiracies about Coronavirus and Its Vaccines* ($P<.001$), *Efficacy of Vaccines and Treatments* ($P<.001$), *Prayers and Calls for Positivity* ($P<.001$), and *Financial Concerns* ($P<.001$). As conspiracy tweets only received higher levels of engagement than the tweets classified as *Venting Negative Emotions*, H1b was partially supported.

Discussion

Principal Findings

This study examined how American Twitter users discussed coronavirus vaccines during the initial stage of the COVID-19 pandemic. Using the technique of topic modeling, this study identified 7 themes in Twitter discussions. While approximately one-fourth of the tweets were about news updates related to coronavirus and vaccine developments, the remaining tweets

consisted of general discussion and information seeking on coronavirus, expressions of financial concerns, disclosures of negative emotions, prayers and calling for positivity, discussions of vaccine and treatment efficacy, and conspiracy theories. In a close association with news or events related to vaccine developments, some themes demonstrated episodic changes and high degrees of co-occurrences. However, the themes of conspiracies about coronavirus and vaccines, prayers and calling for positivity, and efficacy of vaccines and treatments appeared in more periodic patterns. This study enriches our understanding of the public concerns related to vaccines during the early stage of the outbreak, and these shared concerns can inform public health organizations and professionals for more tailored health messages and vaccination policies.

Our results suggest that during the early stage of the pandemic, Twitter discussions related to coronavirus vaccines were centered on general policy issues and were largely mixed with political discussions. Two contextual factors presumably contributed to such characteristics. First, because key stakeholders did not quickly achieve a consensus on containment measures in the initial phase of the pandemic, vaccines were often staged in the public discourses as a potential remedy [12]. It is also understandable that when there was no specific vaccine available, individuals and communities addressed the vaccine issues from a policy-related perspective by discussing the investment and cost aspects of vaccination. Second, the discussions on coronavirus vaccines were situated in the political discourses during the presidential election. A topic revealed from this study was negative emotions toward Donald Trump and explicitly for his claim of using influenza shots to prevent coronavirus infections. Other COVID-19 studies also similarly demonstrated that Donald Trump and other politicians deeply influenced the vaccine discussions and even contributed to the spread of misinformation [34]. This was not surprising as vaccination is one of the politicized health controversies [52,69,70], and it was a strategic effort to feature the vaccine in political discourses. However, political disagreement over vaccines could be detrimental because they were often associated with vaccine hesitancy, reduced confidence in scientific and health facts [71], and decreased policy support for immunizations [72]. Recent research suggested that as different vaccines passed phase trials and were made available to the public, the discussions over vaccine efficacy and safety sharply increased in the United States [73].

Consistent with other studies that examined coronavirus vaccine sentiments and attitudes on social media over different periods of the pandemic [34,74,75], our study indicates that the public had mixed opinions and emotions over coronavirus vaccines, which may create significant barriers to reaching the vaccine-induced herd immunity. Antivaccination arguments and conspiracy theories were one of the major sources for vaccination opposition, although they did not constitute a large part of social media discussions. However, this small proportion of tweets was contributed by Twitter users with more influence. They also demonstrated higher engagement levels, thus resulting in echo chamber effects among small-size subnetworks [52]. It is observed that most themes demonstrated peaks and troughs over time but some themes (eg, *Conspiracies* and *Efficacy of Vaccines and Treatments*) were more periodic and some themes (eg, *Venting Negative Emotions*) were more episodic. We speculate that conspiracies and efficacy concerns were largely about unconfirmed but expected issues (eg, pharmacy conspiracies apply for all the vaccines); thus, such discussions were likely to merge periodically. However, unexpected events (eg, Trump's claim of using influenza shots to prevent coronavirus) will stimulate heated discussions, leading to a peak in the data. When these events were later addressed by the authority's responses, the discussions gradually vanished.

Practical Implications

This study offers several practical implications for addressing the infodemic at the early stage of outbreaks or health crises. First, public health professionals should timely and appropriately address the public needs for vaccine-related information. Our analysis revealed that many Twitter discussions were by people seeking more information or expressing concerns on coronavirus and vaccines. Such surges in information demand should be addressed by supplying with appropriate information that is easy to follow.

Second, health communication may differentiate communication strategies for episodic and periodic themes. As indicated by the results, episodic themes (eg, financial concerns, venting negative emotions) tended to emerge when breaking news or unexpected events occurred. Quick and appropriate responses to these events would effectively reassure the public and eliminate "epidemics of fear" [76]. For periodic themes such as conspiracy theories and efficacy concerns, regular surveillance and tailored responses can counterbalance the negative effects of these themes.

Health organizations and health professionals should make more systematic and organized efforts to address antivaccination content and other vaccine-related misinformation. Together with other studies [47-49], this study indicated that antivaccination content and misinformation about vaccines were contributed by closely connected communities and followed several clear and predictable patterns. When coronavirus vaccines were still under development, antivaccination content had been spreading on the internet along with these recurring conspiracy themes, which indicates that the battle with conspiracies and antivaccination messages is a long fight. A prebunking approach could effectively reduce the negative outcomes of conspiracy theories and misinformation about

vaccines [77]. For example, recent research shows that attitudinal inoculation (eg, prewarning the audiences with common vaccine-related conspiracy theories) can develop resistance to the influence of vaccine conspiracy theories at a later stage [78].

Last but not the least, social media influencers (ie, accounts with many followers) play an important role in the spread of vaccine-related opinions. Fact-checking the content published by social media influencers may effectively limit the spread of conspiracy theories, which requires efforts from both social media platforms and the influencers themselves [79,80]. Twitter recently made some initial moves by introducing a labeling and striking system to identify and remove COVID-19 misinformation [81]. In a related vein, social media influencers are also encouraged by social media and health organizations to enhance their health literacy and their capacities for fact-checking before they self-proclaim as vaccine activists or public health activists on social media platforms.

Limitations

This study has the following limitations. First, the study findings are limited to the Twitter discussions during the first phase of the COVID-19 pandemic in the United States. The public's concerns might have changed over time as the development of vaccines progressed. The recent suspension of a coronavirus vaccine owing to adverse effects has brought a lot of discussions on vaccine safety [9]. The peculiar political environment (eg, presidential election year) may also have contributed to the patterns of the results. Further research is encouraged to look at discussions related to coronavirus vaccines on different social media forms and in different countries. Longitudinal studies and comparisons across countries or regions are particularly preferred to examine the dynamics and heterogeneity in the spread of information and opinions. This study only captured the Twitter discussions during the first stage of the COVID-19 pandemic. Future research could employ larger data sets from Twitter or other social media platforms, especially the latest data sets, to reveal the bigger picture of public concerns over coronavirus vaccines.

There are also some limitations in this analysis. For example, we relied on keyword inquiry to extract vaccine tweets from a database, but we cannot guarantee that all posts were related to coronavirus vaccine conversations. Some outliers might have been included in the data. When interpreting the topic themes, although the 3 authors independently coded the 20 topics, the intercoder reliability was not calculated owing to the small number of topics revealed from the LDA results. The analysis also did not distinguish the nature of Twitter accounts, which may be a mixture of personal, organizational, and bot accounts. Bot accounts may have contributed to a certain portion of the Twitter discussions, but we did not estimate the potential bot traffic. Because Twitter data did not account for users' demographics, while we had a limited understanding of the types of users engaged in the discussion (ie, the number of followers), we do not know more details about the users who were contributing to the discourse. Provided the difficulty of manual classification, future studies should seek to apply more sophisticated machine learning techniques to identify the types

of Twitter accounts—ideally, the characteristics of personal accounts (eg, political ideology). Such knowledge will allow us to go beyond the aggregated data to look at individual users.

Conclusion

Overall, the spread of information and opinions on social media platforms during the early stage of the outbreak has profoundly affected individuals' beliefs and attitudes toward vaccines and, ultimately, their vaccination decisions. During the early stage

of the COVID-19 pandemic in the United States, Twitter discussions related to coronavirus vaccines were centered on general policy issues and were largely mixed with political discussions. The public discussions demonstrated mixed concerns for coronavirus vaccines even before the vaccines were available, and some concerns appeared periodically. These issues call for more preparatory work to cope with the infodemic challenge and to handle infectious breaks in the future.

Acknowledgments

This project is funded by the Faculty Research Fund, City University of Hong Kong (grant 9618020).

Authors' Contributions

LCJ and THC conceptualized and designed the study. THC and MS collected and analyzed the data. All authors drafted the manuscript and revised the final manuscript.

Conflicts of Interest

None declared.

References

1. COVID-19 live updates. Worldometer. URL: <https://www.worldometers.info/coronavirus/> [accessed 2020-12-30]
2. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. *JAMA* 2020 May 26;323(20):2011-2012. [doi: [10.1001/jama.2020.4469](https://doi.org/10.1001/jama.2020.4469)] [Medline: [32202611](https://pubmed.ncbi.nlm.nih.gov/32202611/)]
3. Chan A, Nickson C, Rudolph J, Lee A, Joynt G. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. *Anaesthesia* 2020 Dec;75(12):1579-1582 [FREE Full text] [doi: [10.1111/anae.15057](https://doi.org/10.1111/anae.15057)] [Medline: [32227594](https://pubmed.ncbi.nlm.nih.gov/32227594/)]
4. Callaway E. Coronavirus vaccines: five key questions as trials begin. *Nature* 2020 Mar;579(7800):481. [doi: [10.1038/d41586-020-00798-8](https://doi.org/10.1038/d41586-020-00798-8)] [Medline: [32203367](https://pubmed.ncbi.nlm.nih.gov/32203367/)]
5. Blasi P, King D, Henrikson N. HPV Vaccine Public Awareness Campaigns: An Environmental Scan. *Health Promot Pract* 2015 Nov;16(6):897-905. [doi: [10.1177/1524839915596133](https://doi.org/10.1177/1524839915596133)] [Medline: [26220277](https://pubmed.ncbi.nlm.nih.gov/26220277/)]
6. Manning ML, Davis J. Journal Club: Twitter as a source of vaccination information: content drivers and what they're saying. *Am J Infect Control* 2013 Jun;41(6):571-572. [doi: [10.1016/j.ajic.2013.02.003](https://doi.org/10.1016/j.ajic.2013.02.003)] [Medline: [23726549](https://pubmed.ncbi.nlm.nih.gov/23726549/)]
7. Chen N, Murphy S. Examining the role of media coverage and trust in public health agencies in H1N1 influenza prevention. *International Public Health Journal*. 2011. URL: <https://www.proquest.com/docview/1727485229/abstract/B20E36AA192647FCPO/1?accountid=10134> [accessed 2021-09-07]
8. Chen W, Stoecker C. Mass media coverage and influenza vaccine uptake. *Vaccine* 2020 Jan 10;38(2):271-277. [doi: [10.1016/j.vaccine.2019.10.019](https://doi.org/10.1016/j.vaccine.2019.10.019)] [Medline: [31699506](https://pubmed.ncbi.nlm.nih.gov/31699506/)]
9. Moran MB, Chatterjee JS, Frank LB, Murphy ST, Zhao N, Chen N, et al. Individual, Cultural and Structural Predictors of Vaccine Safety Confidence and Influenza Vaccination Among Hispanic Female Subgroups. *J Immigr Minor Health* 2017 Aug;19(4):790-800 [FREE Full text] [doi: [10.1007/s10903-016-0428-9](https://doi.org/10.1007/s10903-016-0428-9)] [Medline: [27154236](https://pubmed.ncbi.nlm.nih.gov/27154236/)]
10. Shropshire AM, Brent-Hotchkiss R, Andrews UK. Mass media campaign impacts influenza vaccine obtainment of university students. *J Am Coll Health* 2013;61(8):435-443. [doi: [10.1080/07448481.2013.830619](https://doi.org/10.1080/07448481.2013.830619)] [Medline: [24152021](https://pubmed.ncbi.nlm.nih.gov/24152021/)]
11. Xu Z, Ellis L, Laffidy M. News Frames and News Exposure Predicting Flu Vaccination Uptake: Evidence from U.S. Newspapers, 2011-2018 Using Computational Methods. *Health Commun* 2020 Sep 14:1-9. [doi: [10.1080/10410236.2020.1818958](https://doi.org/10.1080/10410236.2020.1818958)] [Medline: [32927970](https://pubmed.ncbi.nlm.nih.gov/32927970/)]
12. Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. Social media can have an impact on how we manage and investigate the COVID-19 pandemic. *J Clin Epidemiol* 2020 Nov;127:198-201 [FREE Full text] [doi: [10.1016/j.jclinepi.2020.06.028](https://doi.org/10.1016/j.jclinepi.2020.06.028)] [Medline: [32603686](https://pubmed.ncbi.nlm.nih.gov/32603686/)]
13. Zhang Y, Tambo E, Djuikoue IC, Tazemda GK, Fotsing MF, Zhou XN. Early stage risk communication and community engagement (RCCE) strategies and measures against the coronavirus disease 2019 (COVID-19) pandemic crisis. *Glob Health J* 2021 Mar;5(1):44-50 [FREE Full text] [doi: [10.1016/j.glohj.2021.02.009](https://doi.org/10.1016/j.glohj.2021.02.009)] [Medline: [33850632](https://pubmed.ncbi.nlm.nih.gov/33850632/)]
14. Lal A, Ashworth HC, Dada S, Hoemeke L, Tambo E. Optimizing Pandemic Preparedness and Response Through Health Information Systems: Lessons Learned From Ebola to COVID-19. *Disaster Med Public Health Prep* 2020 Oct 02:1-8 [FREE Full text] [doi: [10.1017/dmp.2020.361](https://doi.org/10.1017/dmp.2020.361)] [Medline: [33004102](https://pubmed.ncbi.nlm.nih.gov/33004102/)]
15. Mocatta G, Hawley E. The coronavirus crisis as tipping point: communicating the environment in a time of pandemic. *Media International Australia* 2020 Aug 17;177(1):119-124. [doi: [10.1177/1329878x20950030](https://doi.org/10.1177/1329878x20950030)]

16. Apuke O, Omar B. Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics* 2021 Jan;56:101475 [FREE Full text] [doi: [10.1016/j.tele.2020.101475](https://doi.org/10.1016/j.tele.2020.101475)]
17. Freckelton Qc I. COVID-19: Fear, quackery, false representations and the law. *Int J Law Psychiatry* 2020;72:101611 [FREE Full text] [doi: [10.1016/j.ijlp.2020.101611](https://doi.org/10.1016/j.ijlp.2020.101611)] [Medline: [32911444](https://pubmed.ncbi.nlm.nih.gov/32911444/)]
18. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol Sci* 2020 Jul;31(7):770-780 [FREE Full text] [doi: [10.1177/0956797620939054](https://doi.org/10.1177/0956797620939054)] [Medline: [32603243](https://pubmed.ncbi.nlm.nih.gov/32603243/)]
19. Calvillo DP, Ross BJ, Garcia RJB, Smelter TJ, Rutchick AM. Political Ideology Predicts Perceptions of the Threat of COVID-19 (and Susceptibility to Fake News About It). *Social Psychological and Personality Science* 2020 Jul 22;11(8):1119-1128. [doi: [10.1177/1948550620940539](https://doi.org/10.1177/1948550620940539)]
20. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and Extending a Typology to Identify Vaccine Misinformation on Twitter. *Am J Public Health* 2020 Oct;110(S3):S331-S339. [doi: [10.2105/AJPH.2020.305940](https://doi.org/10.2105/AJPH.2020.305940)] [Medline: [33001737](https://pubmed.ncbi.nlm.nih.gov/33001737/)]
21. Rothgerber H, Wilson T, Whaley D, Rosenfeld DL, Humphrey M, Moore AL, et al. Politicizing the COVID-19 Pandemic: Ideological Differences in Adherence to Social Distancing. *PsyArXiv*. Preprint posted online April 20, 2020 [FREE Full text] [doi: [10.31234/osf.io/k23cv](https://doi.org/10.31234/osf.io/k23cv)]
22. Munich security conference. World Health Organization. 2020 Feb 15. URL: <https://www.who.int/dg/speeches/detail/munich-security-conference> [accessed 2020-12-30]
23. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. *J Med Internet Res* 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: [10.2196/21820](https://doi.org/10.2196/21820)] [Medline: [32589589](https://pubmed.ncbi.nlm.nih.gov/32589589/)]
24. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. *J Med Internet Res* 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: [10.2196/19659](https://doi.org/10.2196/19659)] [Medline: [32558655](https://pubmed.ncbi.nlm.nih.gov/32558655/)]
25. Zarocostas J. How to fight an infodemic. *The Lancet* 2020 Feb;395(10225):676 [FREE Full text] [doi: [10.1016/s0140-6736\(20\)30461-x](https://doi.org/10.1016/s0140-6736(20)30461-x)]
26. Abdul-Mageed M, Diab M, Kübler S. SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language* 2014 Jan;28(1):20-37 [FREE Full text] [doi: [10.1016/j.csl.2013.03.001](https://doi.org/10.1016/j.csl.2013.03.001)]
27. Chan M, Jamieson K, Albarracin D. Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States. *Vaccine* 2020 Sep 11;38(40):6236-6247 [FREE Full text] [doi: [10.1016/j.vaccine.2020.07.054](https://doi.org/10.1016/j.vaccine.2020.07.054)] [Medline: [32792251](https://pubmed.ncbi.nlm.nih.gov/32792251/)]
28. Hill HA, Elam-Evans LD, Yankey D, Singleton JA, Kang Y. Vaccination Coverage Among Children Aged 19-35 Months - United States, 2017. *MMWR Morb Mortal Wkly Rep* 2018 Oct 12;67(40):1123-1128 [FREE Full text] [doi: [10.15585/mmwr.mm6740a4](https://doi.org/10.15585/mmwr.mm6740a4)] [Medline: [30307907](https://pubmed.ncbi.nlm.nih.gov/30307907/)]
29. More than 140,000 die from measles as cases surge worldwide. World Health Organization. 2019 Dec 05. URL: <https://www.who.int/news/item/05-12-2019-more-than-140-000-die-from-measles-as-cases-surge-worldwide> [accessed 2020-12-30]
30. Beguerisse-Díaz M, McLennan AK, Garduño-Hernández G, Barahona M, Ulijaszek SJ. The 'who' and 'what' of #diabetes on Twitter. *Digit Health* 2017;3:2055207616688841 [FREE Full text] [doi: [10.1177/2055207616688841](https://doi.org/10.1177/2055207616688841)] [Medline: [29942579](https://pubmed.ncbi.nlm.nih.gov/29942579/)]
31. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. *Nature* 2020 Jun;582(7811):230-233. [doi: [10.1038/s41586-020-2281-1](https://doi.org/10.1038/s41586-020-2281-1)] [Medline: [32499650](https://pubmed.ncbi.nlm.nih.gov/32499650/)]
32. Dredze M, Broniatowski D, Hilyard K. Zika vaccine misconceptions: A social media analysis. *Vaccine* 2016 Jun 24;34(30):3441-3442 [FREE Full text] [doi: [10.1016/j.vaccine.2016.05.008](https://doi.org/10.1016/j.vaccine.2016.05.008)] [Medline: [27216759](https://pubmed.ncbi.nlm.nih.gov/27216759/)]
33. Yiannakoulis N, Slavik C, Chase M. Expressions of pro- and anti-vaccine sentiment on YouTube. *Vaccine* 2019 Apr 03;37(15):2057-2064. [doi: [10.1016/j.vaccine.2019.03.001](https://doi.org/10.1016/j.vaccine.2019.03.001)] [Medline: [30862365](https://pubmed.ncbi.nlm.nih.gov/30862365/)]
34. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. *PLoS One* 2021;16(3):e0247642 [FREE Full text] [doi: [10.1371/journal.pone.0247642](https://doi.org/10.1371/journal.pone.0247642)] [Medline: [33657152](https://pubmed.ncbi.nlm.nih.gov/33657152/)]
35. Walter D, Ophir Y, Jamieson KH. Russian Twitter Accounts and the Partisan Polarization of Vaccine Discourse, 2015-2017. *Am J Public Health* 2020 May;110(5):718-724. [doi: [10.2105/AJPH.2019.305564](https://doi.org/10.2105/AJPH.2019.305564)] [Medline: [32191516](https://pubmed.ncbi.nlm.nih.gov/32191516/)]
36. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
37. Ceron A, Curini L, Iacus SM, Porro G. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society* 2013 Apr 04;16(2):340-358. [doi: [10.1177/1461444813480466](https://doi.org/10.1177/1461444813480466)]
38. Paul M, Dredze M. Discovering health topics in social media using topic models. *PLoS One* 2014;9(8):e103408 [FREE Full text] [doi: [10.1371/journal.pone.0103408](https://doi.org/10.1371/journal.pone.0103408)] [Medline: [25084530](https://pubmed.ncbi.nlm.nih.gov/25084530/)]
39. Stieglitz S, Dang-Xuan L. Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems* 2014 Dec 08;29(4):217-248. [doi: [10.2753/MIS0742-1222290408](https://doi.org/10.2753/MIS0742-1222290408)]

40. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. *Psychol Sci* 2015 Feb;26(2):159-169 [FREE Full text] [doi: [10.1177/0956797614557867](https://doi.org/10.1177/0956797614557867)] [Medline: [25605707](https://pubmed.ncbi.nlm.nih.gov/25605707/)]
41. Gibbs M, Meese J, Arnold M, Nansen B, Carter M. Funeral and Instagram: death, social media, and platform vernacular. *Information, Communication & Society* 2014 Dec 15;18(3):255-268. [doi: [10.1080/1369118X.2014.987152](https://doi.org/10.1080/1369118X.2014.987152)]
42. Calvo Gallardo E, Fernandez de Arroyabe JC, Arranz N. Preventing Internal COVID-19 Outbreaks within Businesses and Institutions: A Methodology Based on Social Networks Analysis for Supporting Occupational Health and Safety Services Decision Making. *Sustainability* 2020 Jun 06;12(11):4655. [doi: [10.3390/su12114655](https://doi.org/10.3390/su12114655)]
43. Scanfeld D, Scanfeld V, Larson E. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010 Apr;38(3):182-188 [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](https://pubmed.ncbi.nlm.nih.gov/20347636/)]
44. Valente T, Gallaher P, Mouttapa M. Using social networks to understand and prevent substance use: a transdisciplinary perspective. *Subst Use Misuse* 2004;39(10-12):1685-1712. [doi: [10.1081/ja-200033210](https://doi.org/10.1081/ja-200033210)] [Medline: [15587948](https://pubmed.ncbi.nlm.nih.gov/15587948/)]
45. Wood MJ. Propagating and Debunking Conspiracy Theories on Twitter During the 2015-2016 Zika Virus Outbreak. *Cyberpsychol Behav Soc Netw* 2018 Aug;21(8):485-490 [FREE Full text] [doi: [10.1089/cyber.2017.0669](https://doi.org/10.1089/cyber.2017.0669)] [Medline: [30020821](https://pubmed.ncbi.nlm.nih.gov/30020821/)]
46. Nawa N, Kogaki S, Takahashi K, Ishida H, Baden H, Katsuragi S, et al. Analysis of public concerns about influenza vaccinations by mining a massive online question dataset in Japan. *Vaccine* 2016 Jun 08;34(27):3207-3213. [doi: [10.1016/j.vaccine.2016.01.008](https://doi.org/10.1016/j.vaccine.2016.01.008)] [Medline: [26776467](https://pubmed.ncbi.nlm.nih.gov/26776467/)]
47. Pruss D, Fujinuma Y, Daughton A, Paul M, Arnot B, Albers Szafir D, et al. Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLoS One* 2019;14(5):e0216922 [FREE Full text] [doi: [10.1371/journal.pone.0216922](https://doi.org/10.1371/journal.pone.0216922)] [Medline: [31120935](https://pubmed.ncbi.nlm.nih.gov/31120935/)]
48. Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, et al. "Mommy Blogs" and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites. *JMIR Public Health Surveill* 2016 Nov 22;2(2):e166 [FREE Full text] [doi: [10.2196/publichealth.6586](https://doi.org/10.2196/publichealth.6586)] [Medline: [27876690](https://pubmed.ncbi.nlm.nih.gov/27876690/)]
49. Hu D, Martin C, Dredze M, Broniatowski D. Chinese social media suggest decreased vaccine acceptance in China: An observational study on Weibo following the 2018 Changchun Changsheng vaccine incident. *Vaccine* 2020 Mar 17;38(13):2764-2770 [FREE Full text] [doi: [10.1016/j.vaccine.2020.02.027](https://doi.org/10.1016/j.vaccine.2020.02.027)] [Medline: [32093982](https://pubmed.ncbi.nlm.nih.gov/32093982/)]
50. Stephens M. A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. *Dialogues in Human Geography* 2020 Jun 23;10(2):276-281. [doi: [10.1177/2043820620935683](https://doi.org/10.1177/2043820620935683)]
51. 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 May 06;22(5):e19458 [FREE Full text] [doi: [10.2196/19458](https://doi.org/10.2196/19458)] [Medline: [32352383](https://pubmed.ncbi.nlm.nih.gov/32352383/)]
52. Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS. COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* 2020 Oct 05;22(10):e22374 [FREE Full text] [doi: [10.2196/22374](https://doi.org/10.2196/22374)] [Medline: [32936771](https://pubmed.ncbi.nlm.nih.gov/32936771/)]
53. Shahrezaye M, Meckel M, Steinacker L, Suter V. COVID-19's (mis)information ecosystem on Twitter: How partisanship boosts the spread of conspiracy narratives on German speaking Twitter. In: *Advances in Information and Communication. Future of Information and Communication Conference 2021*. Cham: Springer; Sep 27, 2020:1060-1073.
54. Chan C, Shumaker L, Maler S. Confirmed coronavirus cases in U.S. reach 100,000: Reuters tally. Reuters. 2020 Mar 28. URL: <https://www.reuters.com/article/us-health-coronavirus-usa-cases-idUSKBN21E3DA> [accessed 2020-12-30]
55. Meltwater. URL: <https://www.meltwater.com/en> [accessed 2021-09-07]
56. Campbell JC, Hindle A, Stroulia E. Latent dirichlet allocation: extracting topics from software engineering data. In: Bird C, Menzies T, Zimmermann T, editors. *The Art and Science of Analyzing Software Data*. 225 Wyman Street, Waltham, MA 02451, USA: Elsevier; 2015:139-159.
57. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* 2013 Dec;41(6):570-606 [FREE Full text] [doi: [10.1016/j.poetic.2013.08.004](https://doi.org/10.1016/j.poetic.2013.08.004)]
58. Zhao X, Zhan M, Jie C. Examining multiplicity and dynamics of publics' crisis narratives with large-scale Twitter data. *Public Relations Review* 2018 Nov;44(4):619-632 [FREE Full text] [doi: [10.1016/j.pubrev.2018.07.004](https://doi.org/10.1016/j.pubrev.2018.07.004)]
59. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. *J Med Internet Res* 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: [10.2196/22590](https://doi.org/10.2196/22590)] [Medline: [32750001](https://pubmed.ncbi.nlm.nih.gov/32750001/)]
60. Tang J, Meng Z, Nguyen X, Mei Q, Zhang M. Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis. 2014 Presented at: *Proceedings of the 31st International Conference on Machine Learning*; June 22-24; Beijing, China p. 190-198 URL: <http://proceedings.mlr.press/v32/tang14.pdf>
61. Grün B, Hornik K. topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software* 2011;40(13):1-30 [FREE Full text] [doi: [10.18637/jss.v040.i13](https://doi.org/10.18637/jss.v040.i13)]
62. Natural Language Toolkit. URL: <https://www.nltk.org/> [accessed 2021-09-07]
63. Citing spaCy v2 #1555. URL: <https://github.com/explosion/spaCy/issues/1555> [accessed 2021-09-07]

64. Lutz C, Carr W, Cohn A, Rodriguez L. Understanding barriers and predictors of maternal immunization: Identifying gaps through an exploratory literature review. *Vaccine* 2018 Nov 26;36(49):7445-7455. [doi: [10.1016/j.vaccine.2018.10.046](https://doi.org/10.1016/j.vaccine.2018.10.046)] [Medline: [30377064](https://pubmed.ncbi.nlm.nih.gov/30377064/)]
65. Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. 2011 Presented at: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing; July 27–31; Edinburgh, Scotland, UK p. 262-272 URL: <https://dl.acm.org/doi/10.5555/2145432.2145462>
66. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *J Med Internet Res* 2016 Aug 29;18(8):e232 [FREE Full text] [doi: [10.2196/jmir.6045](https://doi.org/10.2196/jmir.6045)] [Medline: [27573910](https://pubmed.ncbi.nlm.nih.gov/27573910/)]
67. Mehrotra R, Sanner S, Buntine W, Xie L. Improving LDA topic models for microblogs via tweet pooling and automatic labeling. 2013 Presented at: SIGIR'13; July 28-August 1; Dublin, Ireland URL: <http://users.cecs.anu.edu.au/~ssanner/Papers/sigir13.pdf>
68. Smith N, Graham T. Mapping the anti-vaccination movement on Facebook. *Information, Communication & Society* 2017 Dec 27;22(9):1310-1327. [doi: [10.1080/1369118x.2017.1418406](https://doi.org/10.1080/1369118x.2017.1418406)]
69. Fowler EF, Gollust SE. The Content and Effect of Politicized Health Controversies. *The ANNALS of the American Academy of Political and Social Science* 2015 Feb 08;658(1):155-171. [doi: [10.1177/0002716214555505](https://doi.org/10.1177/0002716214555505)]
70. Saulsberry L, Fowler E, Nagler R, Gollust S. Perceptions of politicization and HPV vaccine policy support. *Vaccine* 2019 Aug 14;37(35):5121-5128. [doi: [10.1016/j.vaccine.2019.05.062](https://doi.org/10.1016/j.vaccine.2019.05.062)] [Medline: [31296376](https://pubmed.ncbi.nlm.nih.gov/31296376/)]
71. Iyengar S, Massey DS. Scientific communication in a post-truth society. 2018 Nov 26 Presented at: The Arthur M. Sackler Colloquium of the National Academy of Sciences, "The Science of Science Communication III"; November 16-17; Washington DC, USA p. 7656-7661. [doi: [10.1073/pnas.1805868115](https://doi.org/10.1073/pnas.1805868115)]
72. Gollust SE, Dempsey AF, Lantz PM, Ubel PA, Fowler EF. Controversy undermines support for state mandates on the human papillomavirus vaccine. *Health Aff (Millwood)* 2010 Nov;29(11):2041-2046. [doi: [10.1377/hlthaff.2010.0174](https://doi.org/10.1377/hlthaff.2010.0174)] [Medline: [21041746](https://pubmed.ncbi.nlm.nih.gov/21041746/)]
73. Dutta S, Kumar A, Dutta M, Walsh C. Tracking COVID-19 vaccine hesitancy and logistical challenges: A machine learning approach. *PLoS One* 2021;16(6):e0252332 [FREE Full text] [doi: [10.1371/journal.pone.0252332](https://doi.org/10.1371/journal.pone.0252332)] [Medline: [34077467](https://pubmed.ncbi.nlm.nih.gov/34077467/)]
74. Rahul K, Jindal BR, Singh K, Meel P. Analysing Public Sentiments Regarding COVID-19 Vaccine on Twitter. 2021 Presented at: 7th International Conference on Advanced Computing and Communication Systems (ICACCS); 19-20 March; Coimbatore, India p. 488-493 URL: <https://ieeexplore.ieee.org/document/9441693> [doi: [10.1109/ICACCS51430.2021.944169](https://doi.org/10.1109/ICACCS51430.2021.944169)]
75. Kwok SWH, Vadde SK, Wang G. Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis. *J Med Internet Res* 2021 May 19;23(5):e26953 [FREE Full text] [doi: [10.2196/26953](https://doi.org/10.2196/26953)] [Medline: [33886492](https://pubmed.ncbi.nlm.nih.gov/33886492/)]
76. Eysenbach G. SARS and population health technology. *J Med Internet Res* 2003;5(2):e14 [FREE Full text] [doi: [10.2196/jmir.5.2.e14](https://doi.org/10.2196/jmir.5.2.e14)] [Medline: [12857670](https://pubmed.ncbi.nlm.nih.gov/12857670/)]
77. Hameleers M. Separating truth from lies: comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society* 2020 May 18:1-17. [doi: [10.1080/1369118x.2020.1764603](https://doi.org/10.1080/1369118x.2020.1764603)]
78. van der Linden S, Dixon G, Clarke C, Cook J. Inoculating against COVID-19 vaccine misinformation. *EClinicalMedicine* 2021 Mar;33:100772 [FREE Full text] [doi: [10.1016/j.eclinm.2021.100772](https://doi.org/10.1016/j.eclinm.2021.100772)] [Medline: [33655205](https://pubmed.ncbi.nlm.nih.gov/33655205/)]
79. Trethewey SP. Strategies to combat medical misinformation on social media. *Postgrad Med J* 2020 Jan;96(1131):4-6 [FREE Full text] [doi: [10.1136/postgradmedj-2019-137201](https://doi.org/10.1136/postgradmedj-2019-137201)] [Medline: [31732511](https://pubmed.ncbi.nlm.nih.gov/31732511/)]
80. Jamison AM, Broniatowski DA, Dredze M, Sangraula A, Smith MC, Quinn SC. Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic' on Twitter. *Harvard Kennedy School Misinformation Review* 2020 Sep;1(3):1-22 [FREE Full text] [doi: [10.37016/mr-2020-38](https://doi.org/10.37016/mr-2020-38)]
81. Updates to our work on COVID-19 vaccine misinformation. *Twitter Safety*. 2021 Mar 01. URL: https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation.html [accessed 2021-04-27]

Abbreviations

LDA: Latent Dirichlet Allocation

Edited by T Mackey; submitted 10.11.20; peer-reviewed by Y Lama, M Agrawal, S Doan, M Das; comments to author 11.12.20; revised version received 30.12.20; accepted 21.06.21; published 14.09.21

Please cite as:

Jiang LC, Chu TH, Sun M

Characterization of Vaccine Tweets During the Early Stage of the COVID-19 Outbreak in the United States: Topic Modeling Analysis
JMIR Infodemiology 2021;1(1):e25636

URL: <https://infodemiology.jmir.org/2021/1/e25636>

doi: [10.2196/25636](https://doi.org/10.2196/25636)

PMID: [34604707](https://pubmed.ncbi.nlm.nih.gov/34604707/)

©Li Crystal Jiang, Tsz Hang Chu, Mengru Sun. Originally published in JMIR Infodemiology (<https://infodemiology.jmir.org>), 14.09.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on <https://infodemiology.jmir.org/>, as well as this copyright and license information must be included.