Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis

Qian Niu,[1*] MS; Junyu Liu,[2*] MS; Masaya Kato,[1] MS; Yuki Shinohara,[1] MS; Natsuki Matsumura,[1] MS; Tomoki Aoyama,[1] MD, PhD; Momoko Nagai-Tanima,[1] PT, PhD

[1] Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2] Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, Japan

* Qian Niu and Junyu Liu contributed equally to this article.

Corresponding Author

Momoko Nagai-Tanima, PT, PhD
Department of Human Health Sciences, Graduate School of Medicine,
Kyoto University
53, Kawahara-cho, Shogoin
Sakyo-ku
Kyoto, 606-8507
Japan
Phone: 075-751-3964
Email: tanima.momoko.8s@kyoto-u.ac.jp

Abstract

Background: The global public health and socioeconomic impacts of COVID-19 have been substantial. To achieve herd immunity, widespread use of the vaccine is required, and it is therefore critical for government and public health agencies to understand public perceptions of the vaccine to help sustain subsequent vaccinations.
Objective: This study aims to explore the opinions and sentiments of tweets about COVID-19 vaccination among Twitter users in Japan, both before and at the beginning of the COVID-19 vaccination program.
Methods: We collected 144,101 Japanese tweets containing COVID-19 vaccine-related keywords from Japanese Twitter users between August 1, 2020, and June 30, 2021. Specifically, we identified temporal changes in the number of tweets and key events that triggered a surge in the number of tweets. In addition, we performed sentiment analysis, and calculated the correlation between different sentiments and the number of deaths, infections, and vaccinations. We also built latent Dirichlet allocation (LDA) topic models to identify commonly discussed topics in a large sample of tweets. We also provided a word cloud of high-frequency unigram and bigram tokens as additional evidence, and conducted further analysis on three different vaccine brands.
Results: The overall number of tweets has continued to increase since the start of mass vaccination in Japan. Sentiments were generally neutral, but negative sentiment was more significant than positive sentiment. Before and after the first vaccination in Japan, the correlations of negative/positive sentiment with death, infection, and vaccination cases changed significantly. Public concerns revolved around three themes: information on vaccine reservations and vaccinations in Japan; infection and mutation of COVID-19 in Japan; and prevention measures, vaccine development and supply, and vaccination status in other countries. Furthermore, public attention to the three brands of vaccines has a temporal shift as clinical trials move forward.
Conclusions: The number of tweets and changes in sentiment might be driven by major news events in relation to the COVID-19 vaccine, with negative sentiments dominating positive sentiments overall. Death and infection cases correlated significantly with negative sentiments, but the correlation fell after vaccinations began as morbidity and mortality decreased. The public's attention to different vaccine brands had a temporal change during their clinical trial process, and although the discussion points are slightly different, the core remains effective and secure.

Keywords: COVID-19; vaccination; public sentiment; Twitter; latent Dirichlet allocation; natural language processing; Japan

Introduction

A novel coronavirus causing COVID-19 was first identified in Wuhan, China, in December 2019 [1]. Globally, as of June 30, 2021, the cumulative number of cases was approximately 181 million, and the number of deaths was about 3.9 million [2]. There were 798,159 confirmed cases in Japan through June 30, 2021, and 14,740 deaths [3]. Since its outbreak, the pandemic has had significant adverse effects on individuals, governments, and the global economy [4].

Global efforts to develop and test COVID-19 vaccines have intensified since March 2020, when the pandemic was declared. COVID-19's spread can be limited to some extent with public health prevention measures, but a vaccine that can provide protective and sustained immunity is crucial to stop the outbreak. To reach herd immunity, at least 70% of the population needs to be vaccinated [5]. Global vaccination coverage of the COVID-19 pandemic must be increased urgently to contain it.

Compared with other developed countries, Japan has experienced significant delays in vaccination, with a vaccination rate of less than 4% until May 21, 2021 [6]. Our World in Data COVID-19 vaccination tracking [7] indicates that, as of June 2021, 68.15% of Canadians, 66.08% of UK adults, and 64.69% of Israeli adults who had received at least one dose of vaccine. However, this proportion was quite low in Japan (25.39%). Tokyo is preparing to host the Summer Olympic Games in 2021, yet only 13.6% of the population was fully vaccinated as of June 30. Hence, it is extremely important for Japan to understand the public's opinion about vaccination, and thereby their willingness to become vaccinated.

While traditional surveys are still helpful for eliciting public health perspectives, Internet forums are increasingly being utilized to debate and share perspectives on infectious disease outbreaks and other health topics [8]. The ongoing COVID-19 pandemic has led to an increase in the use of social media as a venue for addressing a variety of pandemic-related issues, including vaccines. However, social media users may come into contact with unfavorable attitudes and disinformation, which may affect individual opinions and result in vaccination hesitation or refusal [9]. Vaccine hesitancy is one of the ten main risks to world health according to the World Health Organization (WHO) [10,11]. Additionally, misinformation can undermine public confidence in science and public health authorities, resulting in a decline in vaccination uptake [12,13].

Previous research has linked social media activity to vaccination hesitancy and anti-vaccine movements [14–17]. Tangherlini et al. [14] analyzed vaccine hesitancy drivers on blogs and reported that parents may utilize these forums to

spread vaccine opposition views to other parents. Bonnevie et al. [12] studied the evolution of vaccine resistance in the United States by analyzing Twitter discussion topics and found that prominent Twitter accounts were responsible for a significant percentage of vaccine-opposition messages. Cossard et al. [18] investigated the extent to which vaccination discussions may result in vaccine reluctance, as well as how skeptics and proponents' networks are structured differently. Their results demonstrated that anti-vaccine community clusters are more linked than pro-vaccine community clusters. Lyu and colleagues concluded that public discussion is driven by major events, and vaccine sentiment around COVID-19 is increasingly positive, showing higher acceptance than previous vaccines [19]. The findings of several studies indicate positivity in Australia, the UK, and the US, but more positivity is needed to boost vaccination rates to the critical mass needed to attain herd immunity [20,21]. To further their research, Samira et al. [22] took extra effort to conduct a large-scale survey on people's views on vaccines. Although all these studies investigated people's reactions toward COVID-19 vaccines, there have been no studies on the general public's Twitter responses in Japan.

Addressing public vaccine hesitation and helping to boost public trust in vaccinations by investigating vaccination attitudes may lead to greater vaccination coverage [23]. Tailoring messaging using information about vaccination sentiments and views may lead to a successful vaccine promotion strategy. Some characteristics of the population, including demographics, education, religion, and familial status, influence which opinions are expressed for the COVID-19 vaccine [24].

Generating research on public opinion has proven challenging, and governments have tended to depend on surveys with limitations, such as non-representative samples. Twitter data may better represent people's actual thoughts than survey data, hence this is a good tool for conducting real-time studies of the public's views. To date, public opinion on COVID-19 vaccines in Japan needs to be deeply understood. The reluctance to be vaccinated against COVID-19 is particularly worrisome from a public health standpoint; thus, understanding vaccine hesitancy views is critical.

The purpose of this study was to examine the primary viewpoints and sentiments expressed in tweets about the COVID-19 vaccine among Japanese Twitter users. The findings of this research may assist governments and health organizations in planning, modifying, and implementing a timely vaccine promotion strategy that results in herd immunity.

Methods

Data Extraction and Preprocessing

A large-scale COVID-19 Twitter chatter dataset collected and maintained by Georgia State University's Panacea Lab [25] was used in this study. The tweet IDs, posting date and time, and the languages of all tweets are provided in the dataset. Countries of a few tweets were also available. In this study, 1,305,308 Japanese tweets were downloaded from August 1, 2020, to June 30, 2021. We first extracted the tweet IDs of tweets in the Japanese language and then collected the tweets using the Python package Tweepy. Before data cleaning, all the English words in the tweets were transferred from full-width to half-width forms, and lowercased. The tweet data were cleaned using selected keywords related to vaccination. After data cleaning, 144,101 tweets with selected keywords were chosen for further analysis. Official data on the number of deaths, infections, and vaccination cases were collected from the website of the Ministry of Health, Labor and Welfare [26], the Prime Minister's Office of Japan [27], and the Cabinet Secretariat [28]. The daily statistics were calculated from the original data.

Unigram and Bigram Token Analysis

Tokenization is a fundamental step in many natural-language processing (NLP) methods, especially for languages like Japanese, which do not have spaces between words. We tokenized all tweets and analyzed the unigram and bigram tokens. The website links, special characters, numbers, and "amp" (ampersands) were removed from the tweets before tokenization. Then, the Python packages SpaCy and GiNZA were used to remove the Japanese and English stopwords and implement word segmentation. The Python package scikit-learn was used to convert the segmented sentences into unigram and bigram tokens and to calculate the counts of tokens. The plots of the counts of unigrams and bigrams are shown in the Appendix Figure 1-2. We then built word networks on unigram tokens and bigram tokens separately. For unigram tokens, the top 200 tokens were selected. The cosine similarity (>0.3) [29] between each pair of selected tokens was calculated. A word network was created with all the selected tokens as vertices and the scores as the weights of the edges. For bigram tokens, the top 100 tokens were selected. A word network was created with all the words in the selected tokens as vertices, and with the count of tokens as the weights of the edges between the words.

Latent Dirichlet Allocation

LDA [30] is an unsupervised generative probabilistic model widely used in topic modeling. Many recent studies analyzing tweets about COVID-19 vaccines have applied LDA for topic modeling [19,21]. To apply the LDA method, we first generated the document-term matrix, which recorded the token frequencies in each tweet. All the tweets were put into a list, and each tweet was made into unigrams and reconnected by spaces between neighboring tokens. A document-term matrix was generated on the reconnected tweets. We performed the LDA algorithm on the document-term matrix using scikit-learn. The number of topics was set to three. We applied the LDA method to all the tweets, the tweets of different sentiments, and the tweets of different vaccine brands. The results were made into bar plots of different words and word clouds for convenience of analysis and visualization.

Sentiment Analysis

Lacking efficient sentiment analysis models for a mixture of Japanese and English, we used Amazon Web Services (AWS) for sentiment analysis. Among all the services in AWS, Amazon Comprehend [31] is a natural-language processing (NLP) service that uses machine learning to uncover information in unstructured data. A function calling the interface of Amazon Comprehend was deployed on AWS, and a storage bucket in AWS was used to store the texts and outputs of sentiment analysis. The sentiment analysis includes four labels: positive, negative, neutral, and mixed. For each tweet, the model predicts the label and provides the score for each label.

To determine the long-term influence of the statistical data on public attitudes, we calculated the Pearson correlation coefficient between the daily number of positive/negative tweets and the daily statistics of death, inflection, and vaccination cases. We calculated the correlations before and after the start of vaccination in Japan to determine whether vaccination had an influence on the correlations. The Pearson correlation coefficient r of two sets of data X and Y is

$$r = \frac{cov(X, Y)}{\sigma X \sigma Y} \qquad \text{(Eq 1)}$$

The closer the absolute value of the Pearson correlation coefficient is to 1, the stronger the linear correlation between X and Y. In this study, we used the Python package NumPy to make the correlation.

Peak Detection on Daily Trends

To provide an overview of the data, we plotted the trends of the daily number of total tweets and positive/negative tweets. For more precise analyses, the peaks, which are clearly outliers compared to the data of the neighboring days, were marked on the plots.

For more human-like but objective selection, the peaks were determined by an algorithm with reference to the Weber–Fechner law [32] instead of human observation. The peaks of each month were first selected using the following ratio:

$$p_{day} = \frac{n_{day}}{\bar{n}} \qquad \text{(Eq 2)}$$

where n_{day} is the number of tweets for the selected day, and $\bar{n}$ is the average number of daily tweets in the month, including the selected day. If the ratio p_{day} is higher than a threshold λ, the tweet number of that day is judged as a peak. We then selected the valid peaks from all the peaks. Peaks with less than n_{min} tweets and peaks within d days from the previous peak were abandoned. In this study, $\lambda = 1.8$, $d = 5$, and $n_{min} = 200$ for the total number of tweets, and $n_{min} = 40$ for the number of positive/negative tweets.

For all the detected peaks, we checked the tweets and provided the most mentioned vaccine-related news in the tweets that day.

Results

Overview of the Data

We calculated the number of vaccine-related tweets each day and constructed a curve to show the trends in Figure 1. Important news corresponding to the peaks, marking the milestones in Japan's vaccination, was marked on the curve. The Pearson correlation coefficient of daily tweet numbers with the daily death, infection, and vaccination cases are also calculated in Appendix Figure 3-5.

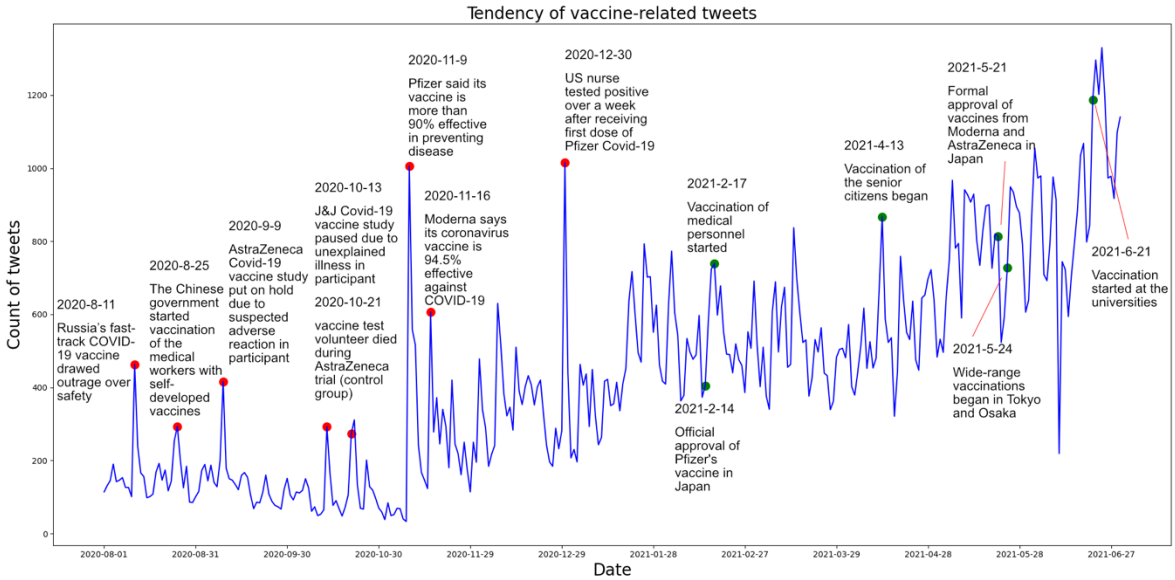

Figure 1. Tendency of vaccine-related tweets with key events collected between August 1, 2020, and June 30, 2021. The red points are selected peaks and green points are important events related to vaccination in Japan.

As shown in Figure 1, the number of daily vaccine-related tweets continually increased; the daily average for the month of June 2021 was 923. Before November 9, 2020, the number of daily vaccine-related tweets was around or below 200, with several peaks. On August 11, 2020 (n=463; n indicates the number of tweets), Russia approved the world's first COVID-19 vaccine [33]. On August 25, 2020 (n=293), the Chinese government initiated the vaccination of medical workers with self-developed vaccines. On September 9, 2020 (n=416), the AstraZeneca COVID-19 vaccine study was put on hold because of suspected adverse reactions in participants . Two relatively small peaks (n=292, n=274) were related to negative news about clinical trials in October. The first waves of sharp increases in the number of daily vaccine-related tweets were on November 9, 2020 (n=1006) and November 16, 2020 (n=606): Pfizer stated that its vaccine was 90% effective [35], and Moderna reported an effectiveness of 94.5%. The second wave of a surge in the number of daily tweets is December 30, 2020 (n=1015), when a US nurse tested positive over a week after receiving the first dose of Pfizer's COVID-19 vaccine [36]. No additional peaks were detected in the following days, but the daily number of vaccine-related tweets continued to increase. This may be related to vaccination in Japan, which is a long-term event across months. On February 14, 2021 (n=405) and May 21, 2021 (n=814), Japan successively recognized the use of Pfizer, Modena, and AstraZeneca vaccines. Japan has adopted a sequential approach to vaccination by opening appointments on February 17 (n=739), April 13 (n=867), May 24 (n=727), and June 21 (n=1187), 2021, for medical workers, senior citizens over 65 years of

age, residents of highly infected areas (Tokyo and Osaka), and university students.

In the Appendix Figure 3-5, the number of tweets per day correlated with the number of deaths, infections, and vaccinations to varying extents, 0.642, 0.405, and 0.686, respectively (0.715 after the first vaccination).

We also created the word networks of unigram and bigram tokens for all vaccine-related tweets. Figure 2 shows the network of word pairs in bigram tokens. From the graph, a group of words that were frequently used together were “Japan,” “Astra,” “Medical care,” “Mutation,” “Pfizer,” and “infection.” Moreover, the word “government” was linked to several tokens such as “Tokyo,” “Olympics,” “economy,” and “Prime minister.” There were a few word pairs, such as “reaction” and “harmful,” which were not connected to the main network and had a relatively small number of counts at the periphery of the graph.

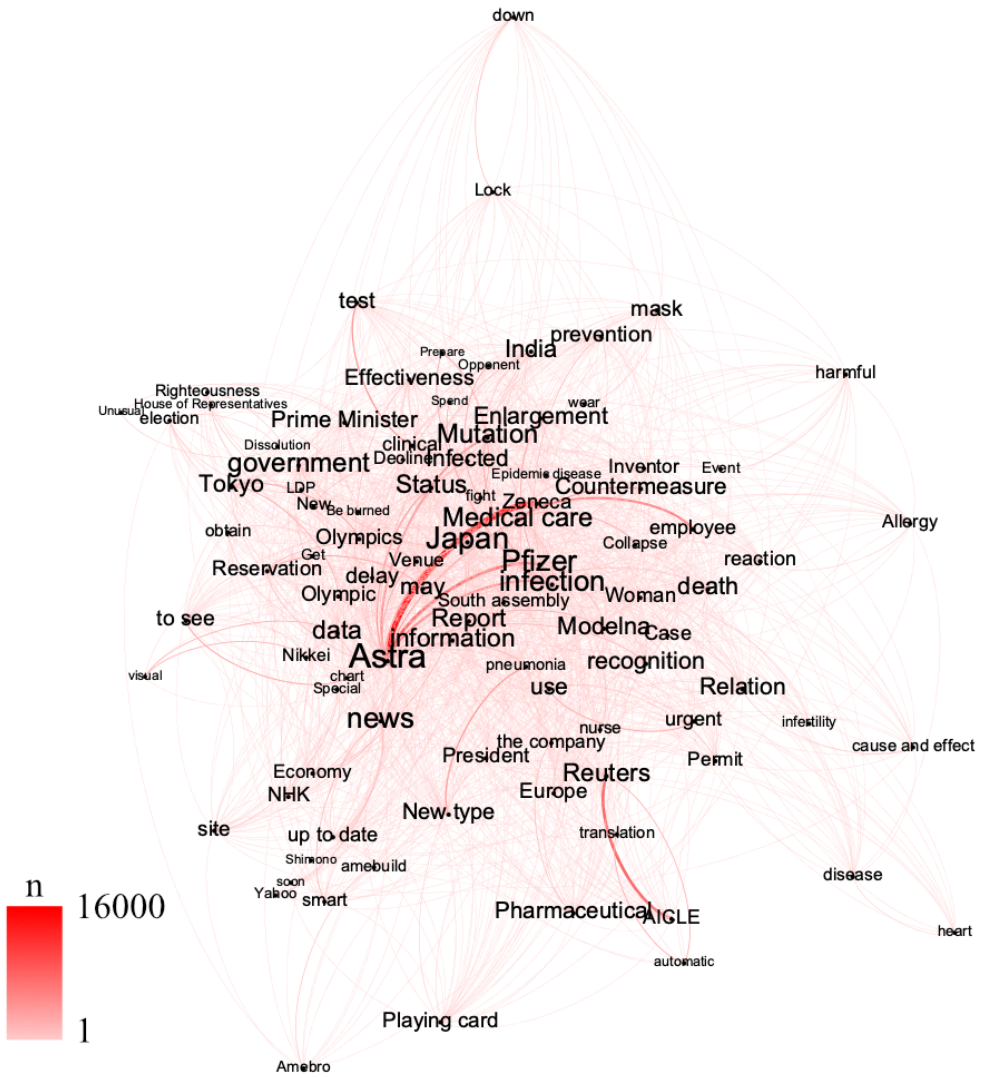

Figure 2. Network of top 100 words pairs in the corpus (English Version). The edges of the graph are the number of counts of word pairs, where thicker and more opaque lines indicate a higher number of counts.

We further examined the similarities between unigram tokens in Figure 3. A major network of words consisted of keywords associated with concerns of vaccines such as “news,” “mutation,” “infection,” “side effect,” “danger,” “senior citizens,” “clinical,” and “prevention.” Another noteworthy word network was composed of keywords that were related to countries: “United Kingdom,” “India,” “Germany,” “China,” “USA,” “Russia,” “Korea,” “Europe,” and “Brazil.”

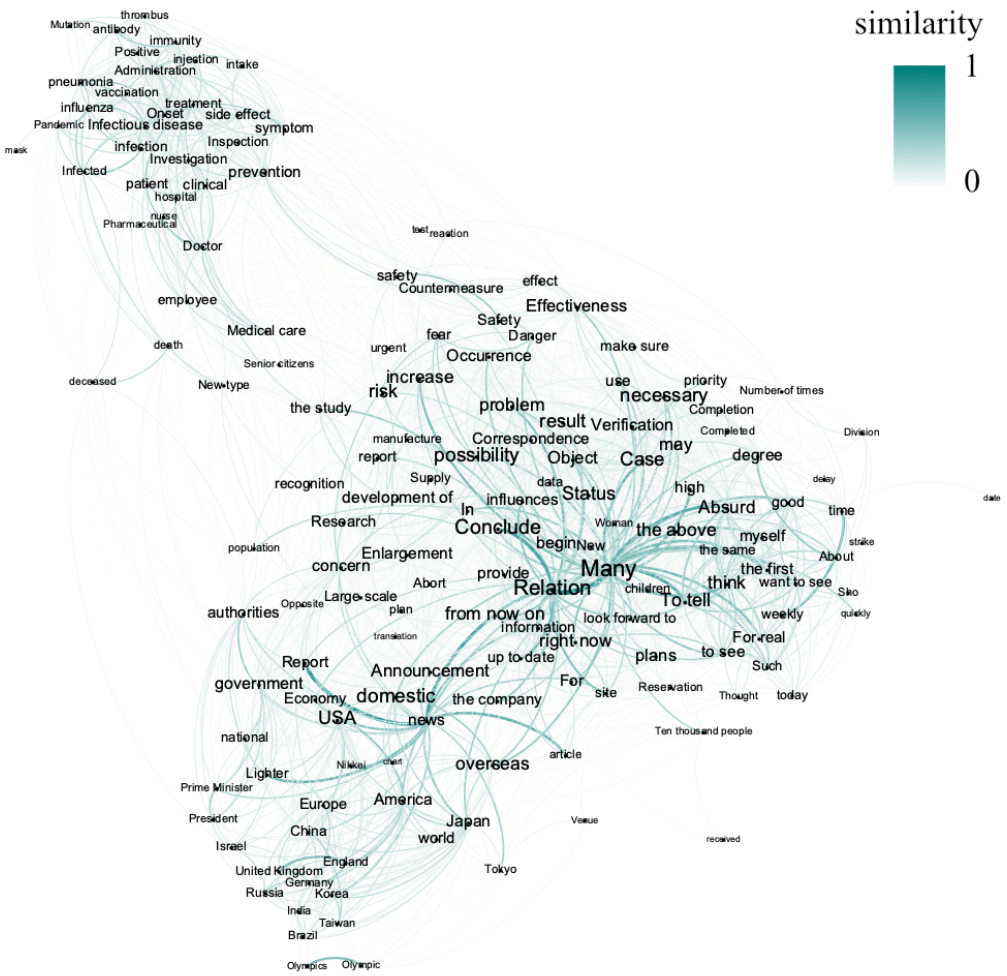

Figure 3. Network of similarity (>0.30) between top 200 unigram tokens (English Version). The edges in the graph are the similarity between unigram tokens, where greater width and higher opacity indicate stronger similarities.

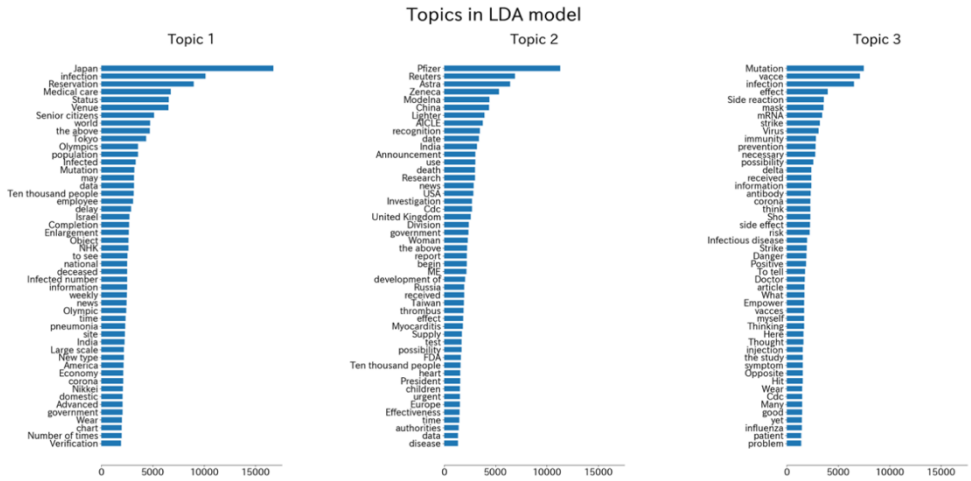

Figure 4. LDA topic modeling on all the vaccine-related tweets (English Version).

Topics and Attitudes toward the Vaccine

First, we generated the topics of all vaccine-related tweets. According to Figure 4 (the word cloud is in Appendix Figure 6), three topic themes were synthesized from the extracted tweets. Topic 1 is about the information on vaccine appointment and vaccination in Japan; Topic 2 relates to the infection and mutation, and prevention measures of COVID-19 in Japan; Topic 3 concerns vaccine development and supply, and vaccination status in other countries. We also provided the LDA modeling results of each sentiment in English as Appendix 15-22.

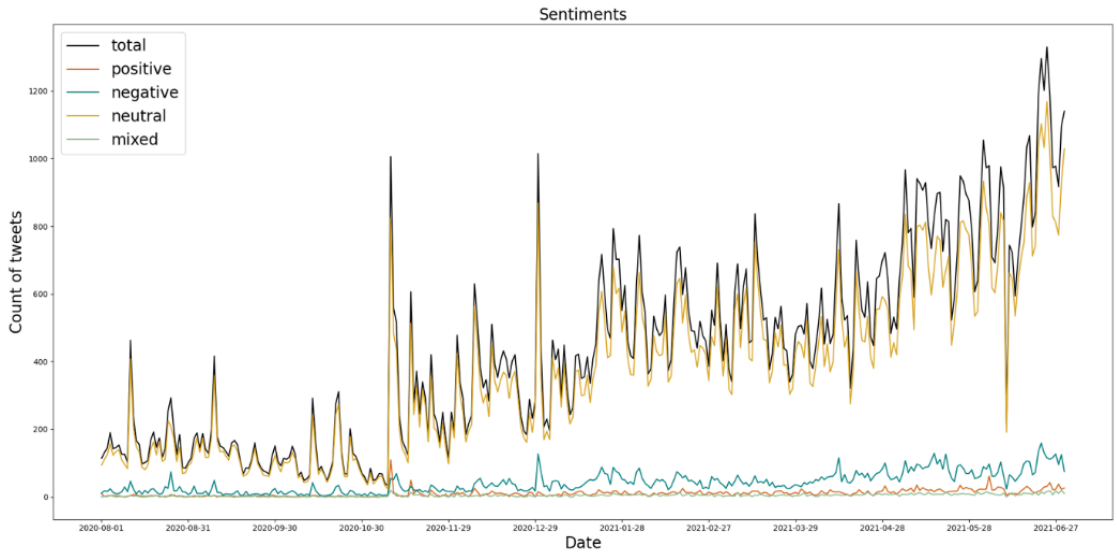

Figure 5. Tendency of different sentiments between August 1, 2020, and June 30, 2021.

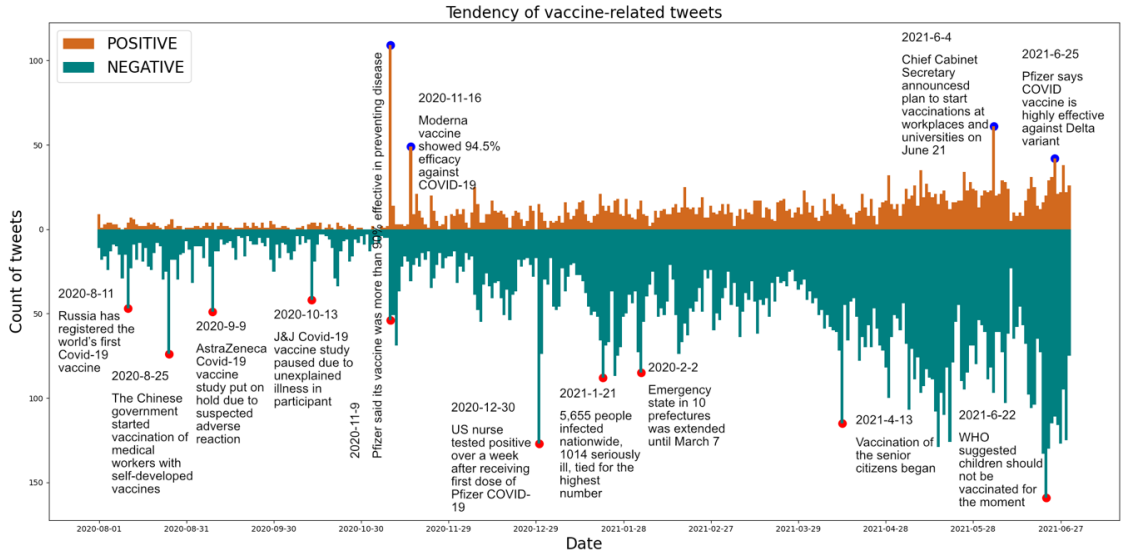

Figure 6. Trends of positive and negative sentiment with key events related to peaks between August 1, 2020, and June 30, 2021.

As Figure 5 shows, we applied sentiment analysis to all tweets and calculated the number of tweets per day for different sentiments. In Figure 6, we show the trends of positive and negative tweets. Sentiment was overwhelmingly neutral, with overall negative sentiment outweighing positive sentiment, but there was also great positive sentiment at some key points. On November 9 (n=83) and November 16 (n=41), 2020, Pfizer and Modena announced the vaccine's effectiveness at over 90% [35] and 94.5% [37], respectively. On June 4, 2021 (n=61), the chief cabinet secretary announced the opening of vaccination appointments for workplaces and universities starting on June 21. The peak of positive tweets on June 25, 2021 (n=42) is related to the effectiveness of the Delta variant of the virus. On the other hand, negative public sentiment erupted in response to events such as the approval of the world's first COVID-19 vaccine in Russia (August 11, 2020, n=47), the administration of China's self-developed vaccine for medical workers on August 25, 2020 (n=74); some negative news about clinical trials on September 9, 2020 (n=49) and October 13, 2020 (n=42); news of a US nurse infected with coronavirus even after the first dose of Pfizer vaccine, on December 30, 2020 (n=127); some negative news about the infection in Japan (January 21, 2021, n=88; February 2, 2021, n=85); the start of vaccines for senior citizens on April 13, 2021 (n=114); and June 22, 2021 (n=159), when the WHO suggested that children should not be vaccinated in the current stage. It should be noted that the positive news about clinical trials also caused a peak of negative attitudes on November 9, 2020 (n=49).

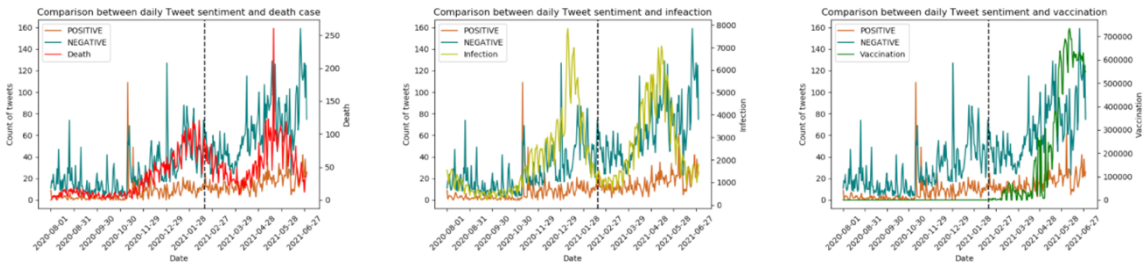

Figure 7. Trends of positive and negative sentiment together with the daily number of death, infection and vaccination cases. The dashed line indicates the first day of vaccination in Japan (February 17, 2020).

We compared the positive and negative sentiment-related tweets with the daily death, infection, and vaccination cases, and correlation coefficients were calculated both before and after the first vaccination on February 17, 2020 (dashed line in Figure 7). As Table 1 shows, negative sentiment was strongly correlated with death (r=0.691) and infection (r=0.500) cases before the first vaccination, but it decreased to 0.305 and 0.293, respectively. In addition, the correlation between negative sentiment and vaccination cases was slightly higher compared to positive sentiment after the first vaccination.

Table 1. The correlation of the positive and negative sentiment with the daily death, infection and vaccination cases.†

	Positive sentiment (*: r≥0.5)		Negative sentiment (*: r≥0.5)	
	Before	After	Before	After
Death cases	0.309	0.350	*0.691	0.305
Infection cases	0.242	0.147	*0.500	0.293
Vaccination cases	/	*0.532	/	*0.575

† The first vaccination in Japan was February 17, 2020, so we calculated the correlation before and after that day, respectively.

Topics and Attitudes toward Different Vaccines

As oJf une 30, 2021, the Japanese government has approved three vaccines, from Pfizer, Moderna, and AstraZeneca, so tweets related to each vaccine brand were extracted (Figure 8), and the daily tendencies of the three vaccine brands are shown in Appendix Figure 7). The numbers of related tweets, in descending order, are Pfizer (n=12,089), AstraZeneca (n=7,300), and Moderna (n=3,338),

and the most representative news related to spikes in tweet numbers was also consistent with Figures 1 and 6. Public concerns about the three vaccines had temporal variations. AstraZeneca vaccines showed an overall upward trend with peaks in September 2020 and March 2021, and a gradual decline after March. The Pfizer vaccine received greater attention than the others from October 2020 to February 2021, and reached a peak in November 2020. Moderna appeared less in discussions, but since March 2021, the number of related tweets has continued to grow along with the Pfizer vaccine.

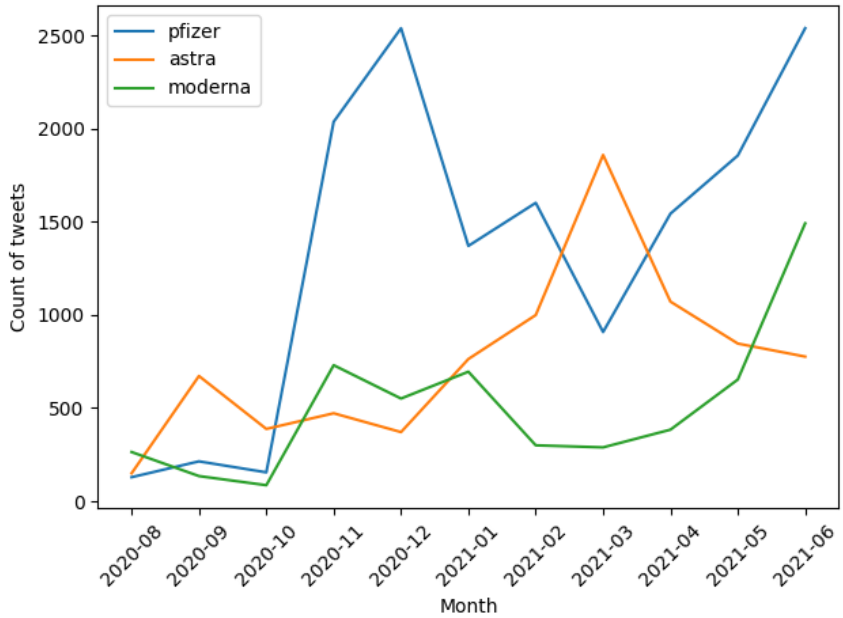

Figure 8. Trends of the Pfizer, AstraZeneca. and Moderna vaccines between August 1, 2020, and June 30, 2021.

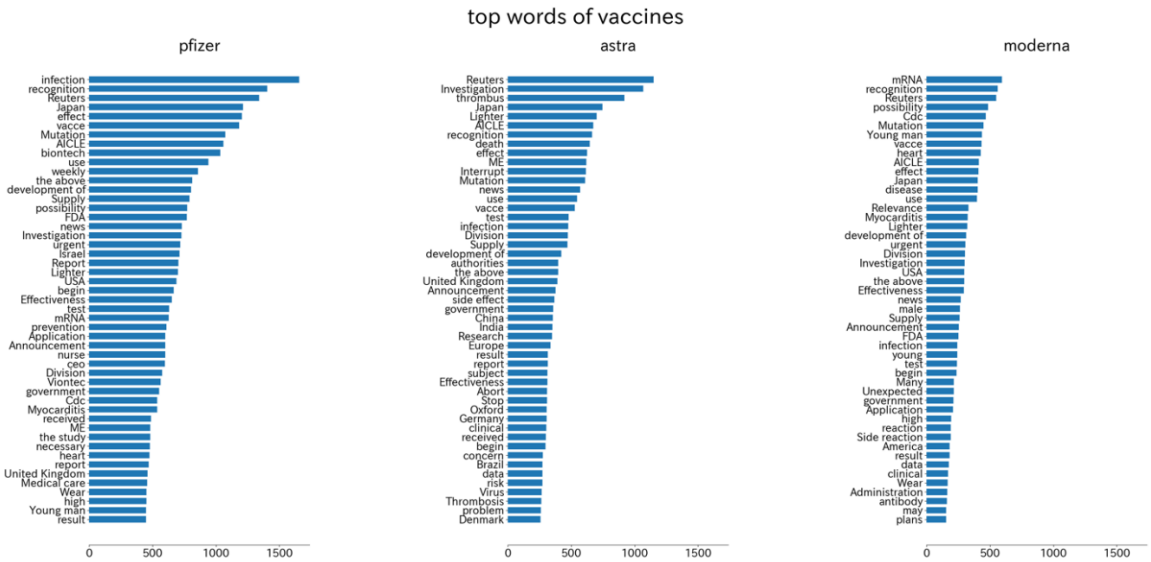

Figure 9. Top words associated with the Pfizer, AstraZeneca, and Moderna vaccines. The length of the bar indicates the number of times a word occurs.

We analyzed the top 50 unigrams in the tweets related to each vaccine brand. As shown in Figure 9 (the word clouds are in Appendix Figure 8), the discussion of all three vaccines were surrounding the keywords "Japan," "approve," "effect," "mutation," "clinical trials," "supply," "side effects," etc. Pfizer, the first vaccine approved by the Japanese government, was discussed more than AstraZeneca and Moderna put together. We also checked the LDA and tendency of positive and negative tweets for the three vaccine brands, as shown in Appendix 9-14. However, the daily number of tweets related to vaccine brands was too small, and it was difficult to perform further analysis according to the results.

Discussion

This study examines sentiment and topics over a long time span, covering discussions from August 11, 2020, through June 30, 2021, when multiple vaccines are available, and mass vaccination is underway in Japan and many other countries. We analyzed temporal changes in Japanese public sentiment toward the COVID-19 vaccine through Twitter, and identified, evaluated, and plotted temporal trends with key events driving the peak. In addition, we separately calculated the correlation between the number of tweets, positive/negative sentiments, and death, infection, and vaccination cases, and provided the hotspots of discussion through LDA models and word clouds. Importantly, we conducted LDA models and top word analysis for three vaccine brands (Pfizer, Moderna, and AstraZeneca), which are not covered in the prior literature.

We found that the number of tweets from the Japanese public on Twitter around vaccines was primarily driven by major events, such as vaccine development milestones, vaccination status, and new variants of the virus. Similar changes in sentiment and topics were found in another study on Facebook and Twitter posts in the United Kingdom and the United States [20]. In today's media ecology, the line between social media and mainstream media is no longer clear, and public health agencies can obtain public opinions and negative sentiments directly—so they should be give the public more sincere and reasonable explanations and policy revisions, to encourage and promote vaccination efficiency. In addition, we found that the daily number of tweets was highly correlated with the daily death cases (r=0.642), and showed correlation with the number of vaccinations (r=0.686); the correlation was 0.715 after the first vaccination. Shim et al. concluded that the daily number of tweets was not

correlated with infection cases in the Korean population [38]. Our results showed a correlation (r=0.405), which may be inaccurate due to several emergency states in Japan.

Our results showed that public sentiment in Japan changed during the pandemic, albeit due to certain vaccine-related events. Overall, negative emotions toward the COVID-19 vaccine predominated over positive emotions. This is not the same as the results of other recent studies on vaccination emotions [20–22]. Kwok et al. found that two-thirds of the emotions expressed toward the COVID-19 vaccine in Australia were positive. Similarly, Hussain et al. found that the overall sentiment for tweets and Facebook posts related to vaccination was positive in the United States and the United Kingdom. In addition, Samira et al. found that all six countries—the United States, Australia, India, the United Kingdom, Canada, and Ireland—showed more positive sentiment and higher levels of engagement with vaccines. In contrast, Shim et al. indicated similar proportions of positive and negative tweets before and after vaccination initiation; however, after an increase in infection numbers, the public bar showed prominently negative sentiments. Interestingly, our results indicated that the same event could trigger both a surge in positive and negative sentiments, as on November 9, 2020, when Pfizer reported that its vaccine was more than 90% effective, which triggered two opposite spikes in different sentiments, with the public expecting its effectiveness on the one hand and expressing concerns about its safety on the other. We presume this was caused by the accumulation of negative news from different vaccines during clinical trials. Moreover, negative sentiment, which was highly correlated with the number of deaths, declined once vaccinations for medical workers began, but they spiked rapidly as vaccinations for those over 65 years of age began and as the number of infections began to increase. Several studies have shown that the Pfizer vaccine reduces morbidity and mortality [39,40]; Figure 5 shows a downward trend in both deaths and infections after the first dose of the vaccine, contrary to the trend of a consistent increase in the number of tweets. Hence, we consider that before and after the first dose of vaccination, the decreased correlation may be attributed to the decrease in morbidity and mortality.

Concern among Japanese Twitter users about the COVID-19 vaccine revolves around three themes. Theme 1 indicates that people paid attention to the information that while vaccine appointments have been opened in Japan, the government is giving priority appointments to healthcare staff and seniors over 65 years old. However, the priority given to senior citizens generated widespread negative sentiment, with some Twitter users arguing that, as the outbreak continues to rise in Japan, priority should be given to those in the

public service sector rather than to senior citizens, who have less contact with people. After the start of large-scale vaccination appointments on May 24, frequent negative news stories, such as busy phone lines and websites crashing when people tried to make appointments, as well as incorrectly administered vaccinations, have caused people to complain and distrust the government. Theme 2 is that Twitter users are concerned about the infection and mutation of COVID-19. They believe that Japan's prevention and control measures are weak, and advocate ongoing, daily prevention, such as wearing masks and social distancing. In addition, people are concerned about the success of the Tokyo Olympics with the spread of the pandemic. Theme 3 focused on the development of vaccines and vaccination status in other countries. The vaccine brands approved in Japan were authorized for emergency use by the WHO and have been administered globally, but the rate of vaccination in Japan is much lower than that in other developed countries, such as Canada, the UK, and Israel. At the same time, Japan is concerned with the vaccination status of China and India, which have large populations. In addition, our results show that negative public sentiment stems from fears about the vaccine, doubts about effectiveness, concerns about side effects, dissatisfaction with the Japanese government, and speculations about conspiracy theories. Moreover, our word clouds, word pair counts, and word-to-word correlations provide additional evidence for visualization.

Since different institutions in Japan can administer different types of vaccines, such as Pfizer open to municipalities and Moderna to universities, understanding public opinion and sentiment about specific brands of vaccines can help vaccine administration move forward. Our analysis of the three vaccines (Pfizer, Moderna, and AstraZeneca) that have been recognized in Japan to date shows that centric public discussions and concerns about different types of vaccines are consistent, such as effectiveness, availability, vaccination information, and side effects. Compared to the other two vaccines, the public tended to focus on the effectiveness of Pfizer in preventing infection, as opposed to Moderna, which tended to focus more on its effectiveness against mutated viruses and its mRNA development technology. Our results show that the public is concerned about the benefits and risks of vaccination at the individual and societal levels, as well as the type of virus or disease, when deciding whether to vaccinate and choose a brand. Furthermore, we found that "fake news" or misleading headlines caused public panic and widespread negative sentiment. For example, on October 21, 2020, numerous media reported that the clinical trial of AstraZeneca vaccine led to the death of a Brazilian volunteer, which continued to trigger public panic even though the next day it was reported that the volunteer did *not* receive the vaccine [41,42]. In an information-exploding network, we regard mainstream

media as being responsible for ensuring objective and accurate news, and recommend that authorities should actively and quickly respond to inaccurate reports and fake news to establish public confidence in vaccines and build scientific understanding.

Negative sentiments about the COVID-19 vaccine predominate over positive ones in Japan. This means that more needs to be done to promote vaccination to achieve herd immunity and protect vulnerable and minority groups. Rigorous science that is easy to understand needs to replace biased, fabricated, or outdated information in the public domain.

Conclusions

This study identified the main topics and sentiments of the Japanese public as expressed on social media regarding COVID-19 vaccine-related issues, and how these topics and sentiments changed over time, with the aim of facilitating vaccination. It was suggested that the changes in the number of tweets and the sentiments expressed might be driven by major news events related to the COVID-19 vaccine. The same event can induce opposite sentiment peaks, and inaccurate news may trigger public panic and a distrust of vaccines. During the vaccine development process and the early stages of vaccination, the Japanese public expressed more negative sentiments than positive ones, with concerns about the vaccine's side effects, and about the mismanagement of the process by authorities. We also found that the public's attention to different vaccine brands changes temporarily with each brand's news, and while the discussion points are slightly different, the core remains effective and secure.

Future Directions

After Japan approved the use of the Pfizer vaccine on February 17, the number of tweets per day fluctuated but showed no overall trend, whereas the trend began rising steadily after Japan's approval of the vaccines by Moderna and AstraZeneca on May 21 and the start of large-scale vaccinations on May 24. Although the total amount of vaccine ordered in Japan is sufficient for the inoculation of all nationals, there were still regional vaccine shortages at the end of June, and those under age 65 can make appointments for vaccination only after August. Furthermore, the Tokyo Summer Olympics will be held in Japan for 17 days from July 23 to August 8, 2021. Considering these events, we will conduct further research on this issue.

Limitations

Although we attempted to analyze Japanese posts on Twitter to gauge public opinion and sentiment throughout Japan, our data may not be representative of the country's population at large, especially the elderly. According to statista[42] , in 2021, about 17% of the Twitter population worldwide is aged 50 or above, leaving 83% in the 49 or younger category. Goldstein et al. [43]showed that the cumulative rate of COVID-19 infection in the community is high among people younger than 35 years old, while transmission in middle or high schools is robust. The opinions of adults over 40 need to be further investigated by modifying the study design, but there is an urgent need to explore the opinions and acceptance of vaccination in younger populations. Moreover, according to Twitter's user privacy protection principles, our study did not further examine the demographic characteristics of users, such as age, gender, and geographic location. Since large-scale vaccination began in Japan on May 24, 2021, our nearly year-long data analysis includes only one month after that time; thus, we will continue to monitor and conduct long-term trend analysis.

Funding

The authors received no specific funding for this project.

Conflicts of Interest

None declared.

Ethical Approval

No ethical approval was required.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet Elsevier; 2020 Feb 15;395(10223):497–506. PMID:31986264

2. WHO Coronavirus (COVID-19) Dashboard [Internet]. [cited 2021 Jun 26]. Available from: https://covid19.who.int

3. Japan: WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data [Internet]. [cited 2021 Jun 26]. Available from: https://covid19.who.int

4. Debata B, Patnaik P, Mishra A. COVID-19 pandemic! It's impact on people, economy, and environment. J Public Aff 2020;20(4):e2372. [doi: 10.1002/pa.2372]

5. Aguas R, Corder RM, King JG, Gonçalves G, Ferreira MU, Gomes MGM. Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. medRxiv Cold Spring Harbor Laboratory Press; 2020 Nov 16;2020.07.23.20160762. [doi: 10.1101/2020.07.23.20160762]

6. Kosaka M, Hashimoto T, Ozaki A, Tanimoto T, Kami M. Delayed COVID-19 vaccine roll-out in Japan. The Lancet Elsevier; 2021 Jun 19;397(10292):2334–2335. PMID:34089658

7. Coronavirus (COVID-19) Vaccinations - Statistics and Research [Internet]. Our World Data. [cited 2021 Jun 26]. Available from: https://ourworldindata.org/covid-vaccinations

8. Peretti-Watel P, Seror V, Cortaredona S, Launay O, Raude J, Verger P, Fressard L, Beck F, Legleye S, L'Haridon O, Léger D, Ward JK. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. Lancet Infect Dis Elsevier; 2020 Jul 1;20(7):769–770. PMID:32445713

9. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, Sicilia-Montalvo JA, Machío-Regidor F. Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019. Vaccines Multidisciplinary Digital Publishing Institute; 2021 Jan;9(1):28. [doi: 10.3390/vaccines9010028]

10. Kunneman F, Lambooij M, Wong A, Bosch A van den, Mollema L. Monitoring stance towards vaccination in twitter messages. BMC Med Inform Decis Mak 2020 Feb 18;20(1):33. [doi: 10.1186/s12911-020-1046-y]

11. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious

diseases. Hum Vaccines Immunother 2020 Nov 1;16(11):2586–2593. PMID:32693678

12. Bonnevie E, Goldbarg J, Gallegos-Jeffrey AK, Rosenberg SD, Wartella E, Smyser J. Content Themes and Influential Voices Within Vaccine Opposition on Twitter, 2019. Am J Public Health American Public Health Association; 2020 Oct 1;110(S3):S326–S330. [doi: 10.2105/AJPH.2020.305901]

13. Hotez PJ, Nuzhath T, Colwell B. Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. Curr Opin Virol 2020 Apr 1;41:1–7. [doi: 10.1016/j.coviro.2020.01.001]

14. Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, Falahi M, Ebrahimzadeh E, Bastani R. "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. PMID:27876690

15. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health American Public Health Association; 2018 Oct 1;108(10):1378–1384. [doi: 10.2105/AJPH.2018.304567]

16. Burki T. Vaccine misinformation and social media. Lancet Digit Health Elsevier; 2019 Oct 1;1(6):e258–e259. [doi: 10.1016/S2589-7500(19)30136-0]

17. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, Zheng M, Manrique P, Wuchty S, Lupu Y. The online competition between pro- and anti-vaccination views. Nature 2020 Jun;582(7811):230–233. [doi: 10.1038/s41586-020-2281-1]

18. Cossard A, Morales GDF, Kalimeri K, Mejova Y, Paolotti D, Starnini M. Falling into the Echo Chamber: the Italian Vaccination Debate on Twitter. ArXiv200311906 Cs [Internet] 2020 Mar 26 [cited 2021 Jun 28]; Available from: http://arxiv.org/abs/2003.11906

19. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Internet Res 2021 Jun 29;23(6):e24435. PMID:34115608

20. 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 Apr 5;23(4):e26627. PMID:33724919

21. 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 [Internet] 2021 May 19 [cited 2021 Jun 6];23(5). PMID:33886492

22. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An Analysis of COVID-19 Vaccine Sentiments and Opinions on Twitter. Int J Infect Dis IJID Off Publ Int Soc Infect Dis 2021 May 27; PMID:34052407

23. Ferrer RA, Ellis EM. Moving beyond categorization to understand affective influences on real world health decisions. Soc Personal Psychol Compass 2019;13(11):e12502. [doi: 10.1111/spc3.12502]

24. Lyu H, Wang J, Wu W, Duong V, Zhang X, Dye TD, Luo J. Social Media Study of Public Opinions on Potential COVID-19 Vaccines: Informing Dissent, Disparities, and Dissemination. ArXiv201202165 Cs [Internet] 2020 Dec 15 [cited 2021 Jun 28]; Available from: http://arxiv.org/abs/2012.02165

25. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, Artemova K, Tutubalina E, Chowell G. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. ArXiv200403688 Cs [Internet] 2020 Nov 13 [cited 2021 Jul 4]; Available from: http://arxiv.org/abs/2004.03688

26. Situation report [Internet]. [cited 2021 Jul 4]. Available from: https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou_00006.html

27. Ongoing Topics | Prime Minister of Japan and His Cabinet [Internet]. [cited 2021 Jul 4]. Available from: https://japan.kantei.go.jp/ongoingtopics/vaccine.html

28. COVID-19 Information and Resouces [Internet]. [cited 2021 Jul 4]. Available from: https://corona.go.jp/en/

29. Nguyen HV, Bai L. Cosine Similarity Metric Learning for Face Verification. In: Kimmel R, Klette R, Sugimoto A, editors. Comput Vis – ACCV 2010 Berlin, Heidelberg: Springer; 2011. p. 709–720. [doi: 10.1007/978-3-642-19309-5_55]

30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 1;3(null):993–1022.

31. Amazon Comprehend（テキストのインサイトや関係性を検出）| AWS [Internet]. Amaz Web Serv Inc. [cited 2021 Jul 4]. Available from: https://aws.amazon.com/jp/comprehend/

32. Hecht S. THE VISUAL DISCRIMINATION OF INTENSITY AND THE WEBER-FECHNER LAW. J Gen Physiol 1924 Nov 20;7(2):235–267. [doi: 10.1085/jgp.7.2.235]

33. Russia plans “world’s first approved” COVID-19 vaccine by Aug. 12 | The Japan Times [Internet]. [cited 2021 Jul 4]. Available from:

https://www.japantimes.co.jp/news/2020/07/30/world/science-health-world/russia-worlds-first-approved-vaccine/

34. Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study | pfpfizeruscom [Internet]. [cited 2021 Jul 4]. Available from: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against

35. Vaccinated U.S. Nurse Contracts COVID-19, Expert Says Pfizer Shot Needed More Time to Work - ABC | Top News | US News [Internet]. US News World Rep. [cited 2021 Jul 6]. Available from: //www.usnews.com/news/top-news/articles/2020-12-30/california-nurse-tests-positive-over-a-week-after-receiving-pfizer-covid-19-vaccine-abc

36. Moderna's COVID-19 Vaccine Candidate Meets its Primary Efficacy Endpoint in the First Interim Analysis of the Phase 3 COVE Study | Moderna, Inc. [Internet]. [cited 2021 Jul 5]. Available from: https://investors.modernatx.com/news-releases/news-release-details/modernas-covid-19-vaccine-candidate-meets-its-primary-efficacy/

37. Shim J-G, Ryu K-H, Lee SH, Cho E-A, Lee YJ, Ahn JH. Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea. Int J Environ Res Public Health 2021 Jun 18;18(12):6549. PMID:34207016

38. Bernal JL, Andrews N, Gower C, Robertson C, Stowe J, Tessier E, Simmons R, Cottrell S, Roberts R, O'Doherty M, Brown K, Cameron C, Stockton D, McMenamin J, Ramsay M. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. BMJ British Medical Journal Publishing Group; 2021 May 13;373:n1088. PMID:33985964

39. Wise J. Covid-19: Pfizer BioNTech vaccine reduced cases by 94% in Israel, shows peer reviewed study. BMJ British Medical Journal Publishing Group; 2021 Feb 25;372:n567. PMID:33632676

40. Covid-19 vaccine volunteer in Brazil's AstraZeneca trial dies - CNN [Internet]. [cited 2021 Jul 16]. Available from: https://edition.cnn.com/2020/10/21/health/covid-19-vaccine-trial-brazil-astrazeneca-death-bn/index.html

41. Deceased AstraZeneca Trial Volunteer Didn't Receive Vaccine. Bloomberg.com [Internet] 2020 Oct 21 [cited 2021 Jul 16]; Available from: https://www.bloomberg.com/news/articles/2020-10-21/astrazeneca-dips-on-report-of-brazil-death-in-vaccine-trial

42. Global Twitter user age distribution 2021 [Internet]. Statista. [cited 2021 Jul 19]. Available from: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

43. Goldstein E, Lipsitch M, Cevik M. On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community. J Infect Dis 2021 Feb 1;223(3):362–369. [doi: 10.1093/infdis/jiaa691]