

Original Paper

The Effect of Fear of Infection and Sufficient Vaccine Reservation Information on Rapid COVID-19 Vaccination in Japan: Evidence From a Retrospective Twitter Analysis

Qian Niu^{1*}, MS; Junyu Liu^{2*}, ME; Masaya Kato¹, MS; Momoko Nagai-Tanima¹, PhD; Tomoki Aoyama¹, MD, PhD

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

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

*these authors contributed equally

Corresponding Author:

Momoko Nagai-Tanima, PhD

Department of Human Health Sciences

Graduate School of Medicine

Kyoto University

53, Kawahara-Cho, Shogoin

Sakyo-Ku

Kyoto, 606-8507

Japan

Phone: 81 075 751 3964

Email: tanima.momoko.8s@kyoto-u.ac.jp

Abstract

Background: The global public health and socioeconomic impacts of the COVID-19 pandemic have been substantial, rendering herd immunity by COVID-19 vaccination an important factor for protecting people and retrieving the economy. Among all the countries, Japan became one of the countries with the highest COVID-19 vaccination rates in several months, although vaccine confidence in Japan is the lowest worldwide.

Objective: We attempted to find the reasons for rapid COVID-19 vaccination in Japan given its lowest vaccine confidence levels worldwide, through Twitter analysis.

Methods: We downloaded COVID-19-related Japanese tweets from a large-scale public COVID-19 Twitter chatter data set within the timeline of February 1 and September 30, 2021. The daily number of vaccination cases was collected from the official website of the Prime Minister's Office of Japan. After preprocessing, we applied unigram and bigram token analysis and then calculated the cross-correlation and Pearson correlation coefficient (r) between the term frequency and daily vaccination cases. We then identified vaccine sentiments and emotions of tweets and used the topic modeling to look deeper into the dominant emotions.

Results: We selected 190,697 vaccine-related tweets after filtering. Through n-gram token analysis, we discovered the top unigrams and bigrams over the whole period. In all the combinations of the top 6 unigrams, tweets with both keywords "reserve" and "venue" showed the largest correlation with daily vaccination cases ($r=0.912$; $P<.001$). On sentiment analysis, negative sentiment overwhelmed positive sentiment, and fear was the dominant emotion across the period. For the latent Dirichlet allocation model on tweets with fear emotion, the two topics were identified as "infect" and "vaccine confidence." The expectation of the number of tweets generated from topic "infect" was larger than that generated from topic "vaccine confidence."

Conclusions: Our work indicates that awareness of the danger of COVID-19 might increase the willingness to get vaccinated. With a sufficient vaccine supply, effective delivery of vaccine reservation information may be an important factor for people to get vaccinated. We did not find evidence for increased vaccine confidence in Japan during the period of our study. We recommend policy makers to share accurate and prompt information about the infectious diseases and vaccination and to make efforts on smoother delivery of vaccine reservation information.

(*J Med Internet Res* 2022;24(6):e37466) doi: [10.2196/37466](https://doi.org/10.2196/37466)

KEYWORDS

COVID-19; vaccine hesitancy; Japan; social media; text mining

Introduction

COVID-19 has spread worldwide since its first case in December 2019 and has become a public health emergency of international concern [1]. Until September 30, 2021, Japan experienced 5 waves of the COVID-19 pandemic [2]. The surge of COVID-19 in Japan occurred during the Tokyo Olympics, bringing the cumulative number of COVID-19 cases to 1,556,998 when the Games finished. However, with the lifting of the fourth national state of emergency on September 30, 2021, the pandemic was effectively contained nationwide, and the number of new confirmed cases abruptly decreased. The high vaccination rate in Japan was considered to have caused a decline in the community infections during the fifth wave [3].

A high vaccination rate is thought to be promoted by high vaccine confidence [4]. According to the US Centers of Disease Control and Prevention, “Vaccine confidence is the belief that vaccines work, are safe, and are part of a trustworthy medical system” [5]. A global survey that did not include Japan showed that the potential acceptance of a COVID-19 vaccine largely varied among countries [6]. Japan ranks among the countries with the lowest vaccine confidence worldwide according to a survey in 2020 [7]. Another survey conducted before large-scale vaccination in Japan indicated that Japan ranked last with regard to confidence in COVID-19 vaccines among 15 countries [8]. Gordon and Reich [9] explained the historical reasons for low vaccine confidence in Japan. Kunitoki et al [10] proposed that barriers to vaccine access and use mainly result from effective public communication and called for rebuilding vaccine confidence in Japan.

However, Japan’s speed of vaccination has been impressive since large-scale vaccination was opened up (May 24, 2021). Japan’s first-dose vaccination rate was approximately 6.8% by June 1, 2021, and over 70% of the population accepted at least one dose until September 30, 2021 [11]. Notably, vaccination was not mandatory and was administered only with the recipient’s consent. A survey of multiple countries reported the coexistence of a high level of uncertainty about the safety of COVID-19 vaccines and a high willingness to get vaccinated [12], which indicates that Japan may not be a special case. The reason for the contradiction between the rapid growth in the COVID-19 vaccination rate and low vaccine confidence in Japan is worth studying and maybe instructive for propelling worldwide vaccination against infectious diseases.

Twitter is a widespread social media platform that has attracted the increasing attention of public health researchers because of its advantages of large amounts, real-time availability, and ease of public searching and access [13]. With a large amount of real-time COVID-19–related posts, Twitter has been widely used for public opinion mining toward COVID-19 during the pandemic, providing policy makers with substantiated evidence [12,14,15]. Lyu et al [14] reported the trend of topics and sentiments of English tweets for approximately 11 months since the World Health Organization declared COVID-19 a pandemic.

Yousefinaghani et al [12] reported the dominance of positive sentiments and more vaccine objection and hesitancy than vaccine interest. Huangfu et al [15] reported the results of topic modeling and sentiment analysis of tweets between December 8, 2020, and April 8, 2021. Eibensteiner et al [16] reported willingness to vaccinate despite the safety concerns of vaccines, according to a survey on a Twitter poll. Besides, Twitter is the most popular social media platform in Japan [17], owning 58.2 million users as of October 2021 [18], making Twitter analysis more powerful for COVID-19 research in Japan. A Twitter analysis by Niu et al [19] reported that the Japanese public’s negative sentiment overwhelmed the positive sentiment toward the COVID-19 vaccine before and at the beginning of the large-scale vaccination campaign.

This retrospective study aimed to identify public sentiments and concerns associated with rapid COVID-19 vaccination in Japan. We hypothesized that the increase in vaccination rates might be due to subjective factors including increased public confidence in vaccines (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery reservation–related vaccine information (O2). To test these hypotheses, we collected vaccine-related tweets posted between February 1 and September 30, 2021. Then, we preprocessed the collected tweets and conducted a unigram token analysis, sentiment analysis, and topic modeling.

Methods**Overview**

In previous works of large-scale Twitter analyses, after preprocessing, there are mainly 4 types of natural language processing (NLP) methods: n-gram token analysis [12,15,20,21], sentiment analysis [12,14,15,20–25], topic modeling [12,14,15,20,22–25], and geographical analysis [22,24]. The geographical analysis is less important in our work because the range of our research is a whole country instead of subareas. In this work, we followed previous works in applying n-gram token analysis, sentiment analysis, and topic modeling. Code in this work will be shared on the web [26].

Data Collection and Preprocessing

The data used in this study were obtained from a large-scale public COVID-19 Twitter chatter data set [27] updated by the Georgia State University’s Panacea Lab. The data set provided the IDs, posting time, and the languages of all the tweets were provided in the data set. We downloaded COVID-19–related Japanese tweets between February 1, 2021, the month the first person was vaccinated, and September 30, 2021, when the first-dose vaccination rate exceeded 70%. In addition, data on the number of vaccination cases were collected from the official website of the Prime Minister’s Office of Japan (PMOJ) [28].

The downloaded tweets were then cleaned and processed. Retweets were filtered using the Python package tweepy. Tweets that included no keywords related to vaccines were deleted. The keywords used in the filtering are listed in [Multimedia Appendix](#)

1. It is worth noting that the three vaccine brands (Pfizer, Moderna, and AstraZeneca) that were approved by the Japanese government were included in the keywords. Other vaccine brands were excluded because we attempted to focus more on the brands adopted in the vaccination process. Frequent misspellings (eg, “Modelna”) was also included in the keywords. Weblinks, special characters, emojis, and “amp” (ampersands) were removed, and all full-width English characters were converted to half-width, lowercase characters.

For convenience, all Japanese words in our results were directly presented in English translations. The English-Japanese translation table is provided in [Multimedia Appendix 1](#). In order to minimize the influence of difference between languages, all the translations in our results were carried out as the last step by directly replacing the Japanese words in the graphs with corresponding English words; therefore, they would not influence the statistical results.

Unigram and Bigram Token Analysis

Tokenization is necessary before many other NLP tasks, especially for many non-Latin languages, such as Japanese. We removed the predefined English and Japanese stop words in the Python packages NLTK [29] and SpaCy [30] and tokenized all collected vaccine-related tweets using the Python package SpaCy into unigrams or bigrams for statistical analysis, as reported by Kwok et al [27]. We sorted the unigram tokens or bigram tokens in descending order of term frequency over the entire period. Similar to Liu et al [24], we used the pruned exact linear time (PELT) algorithm [31] to find the first change point of the term frequency. Unigrams before the first change point were regarded as top unigrams, and the term frequencies of the unigrams after the change point were significantly lower than those of the top unigrams. Similar processes were carried out for bigrams. To eliminate the difference in the number of days between months, the monthly term frequency was defined by dividing the total term frequency by the number of days each month for each top unigram or bigram.

Correlation coefficients were widely used in social media analysis. In Google Trends analysis, correlations were calculated between reported cases of infectious disease and the trends of search for relevant keywords. In Twitter analyses, correlations between the daily cases of infection or death and the number of related tweets or sentiment scores, were also investigated [24,25]. In this work, correlation analyses were adopted to find out the factors from the top unigrams that are most related to the COVID-19 vaccination campaign. We first calculated the cross-correlations between the number of tweets containing the top unigrams or bigrams and the vaccination cases and then observed the time lags when maximum cross-correlation appeared for each unigram and bigram. Pearson correlation coefficients (r) between top unigrams or bigrams and the vaccination cases were also calculated.

Sentiment Analysis

After n-gram analysis, sentiment analyses were often used to explore the real-time public attitudes in social media analysis related to COVID-19 vaccination, which may reflect the acceptance of COVID-19 vaccines and related policies

[12,14,20,22,24]. The trend of negative sentiments may provide potential evidence for vaccine hesitancy [23]. In this work, sentiment analysis was applied to all vaccine-related tweets. Cloud services were used in this study because there were no reliable public models for sentiment analysis in the Japanese language. We selected Amazon Web Services (AWS) for consistency with previous work [29]. The tweets were divided into positive, negative, neutral, or a mixture of positive and negative tweets using the AWS. Fine-grained emotions were also explored using the Japanese version of the NRC Emotion Lexicon [32]. The NRC Emotion Lexicon is a dictionary of words and their associated scores for eight emotions: anticipation, trust, joy, surprise, anger, disgust, fear, and sadness. The positive and negative tweets were tokenized, and the degree of valence (DOV) for the eight emotions was calculated by adding up the scores for the unigrams that appeared in the NRC Emotion Lexicon. Finally, we calculated the daily average DOV by dividing the number of positive and negative tweets on that day to show the trend of each emotion.

Topic Modeling

Topic modeling were applied to identify fine-grained information from tweets of different sentiments [12,15,24]. Based on the sentiment analysis results, we summarized the topics to look deeper into the dominant emotion in the tweets. Latent Dirichlet allocation (LDA) is often used in tweet topic modeling studies [14,15,20,22,23]. In this study, LDA regards tweets as being generated from different topics, and each topic generates tweets with a Dirichlet distribution. A Python package scikit-learn was used to determine the best number of topics. Log likelihood was adopted as the metric for selection, and 5-fold cross correlation was applied to avoid overfitting. As shown in [Multimedia Appendix 1](#), we chose 2 as the number of topics for LDA modeling, which showed the highest log likelihood score. We used scikit-learn for LDA topic modeling and displayed the top 10 keywords and their weights related to each topic. The weights were the pseudocounts of the keywords in a topic. The themes of topics were summarized from the top 10 keywords by 3 volunteers. The volunteers were first asked to work out the themes of the topics independently, and then they had a meeting to finally reach an agreement on the themes.

We then checked the trends of tweets related to different topics. Defining the i -th tweet in all collected tweets as d_i , and the j -th topic of the LDA model as t_j , the probability of a tweet d_i coming from t_j was calculated using the fitted LDA model as p_{ij} . For tweets posted each day, the expectation of the number of tweets generated from topic j was calculated by summing p_{ij} on that day. The ratio between the expected number of tweets generated from each topic was also plotted to show the trend of public attention under dominant emotion.

Ethics Approval

This study used publicly available and accessible tweets collected by Georgia State University's Panacea Lab, allowing free download. We assert that our analysis is compliant with Twitter's usage policy in aggregate form without identifying specific individuals who published the Twitter posts. Furthermore, the number of vaccination cases downloaded from

the PMOJ are open government data. Therefore, the activities described do not meet the requirements of human subject research and did not require review by an institutional review board.

Results

Data Summary

We downloaded 979,636 Japanese tweets posted between February 1 and September 30, 2021, according to the ID and region information in the data set. After filtering, 190,697 vaccine-related tweets were selected. As a result, the total number of vaccine-related tweets increased from 14,758 tweets in February to 34,692 in August and then decreased to 27,824 in September.

Unigram and Bigram Token Analysis

The change point of unigram term frequencies detected by the PELT algorithm was 6, and the top 6 unigrams were Japanese words for “infection,” “Japan,” “reserve,” “Pfizer,” “venue,” and “mutation.” The unigram “side effects,” related to the safety of vaccines, ranked eighth overall. The unigrams “infect,” “reserve,” and “venue” gradually ranked in the top 3 from February to September, as shown in Figure 1.

The change point of bigram term frequencies detected by the PELT algorithm was 5, and the top 5 were Japanese bigrams

for “Astra + Zeneca,” “reserve + available,” “article + Reuters,” “venue + reserve,” and “medical-care + workers.” The bigrams “reserve + available” and “venue + reserve” ranked in the top-2 from June to September, and the ranking of “Astra + Zeneca” decreased since May, as shown in Figure 2.

Regarding correlation analysis of unigrams, the time lags for “reserve” and “venue” were 0, and the vaccination cases led the number of tweets containing “infection” for 5 days. After calculating r between the daily number of tweets containing each top unigram and vaccination cases, significant r values ($P < .001$) were obtained for all unigrams except “mutation.” The largest r value for the daily vaccination cases was from unigrams “infection” ($r = 0.746$), “reserve” ($r = 0.829$), and “venue” ($r = 0.908$). We then checked the daily number of tweets containing all the combinations of the 3 unigrams showing a strong correlation and found the highest r value ($r = 0.912$; $P < .001$) for tweets containing both “reserve” and “venue.” By randomly selecting 5 days and checking the source of all the tweets on those days, we found that the 95% CI of tweets containing both “reserve” and “venue” posted by official accounts or mainstream media was 96.0%-100%. The trend of tweets containing both unigrams “reserve” and “venue” compared with the daily vaccination cases is shown in Figure 3.

Figure 1. Translation of the top 10 unigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.

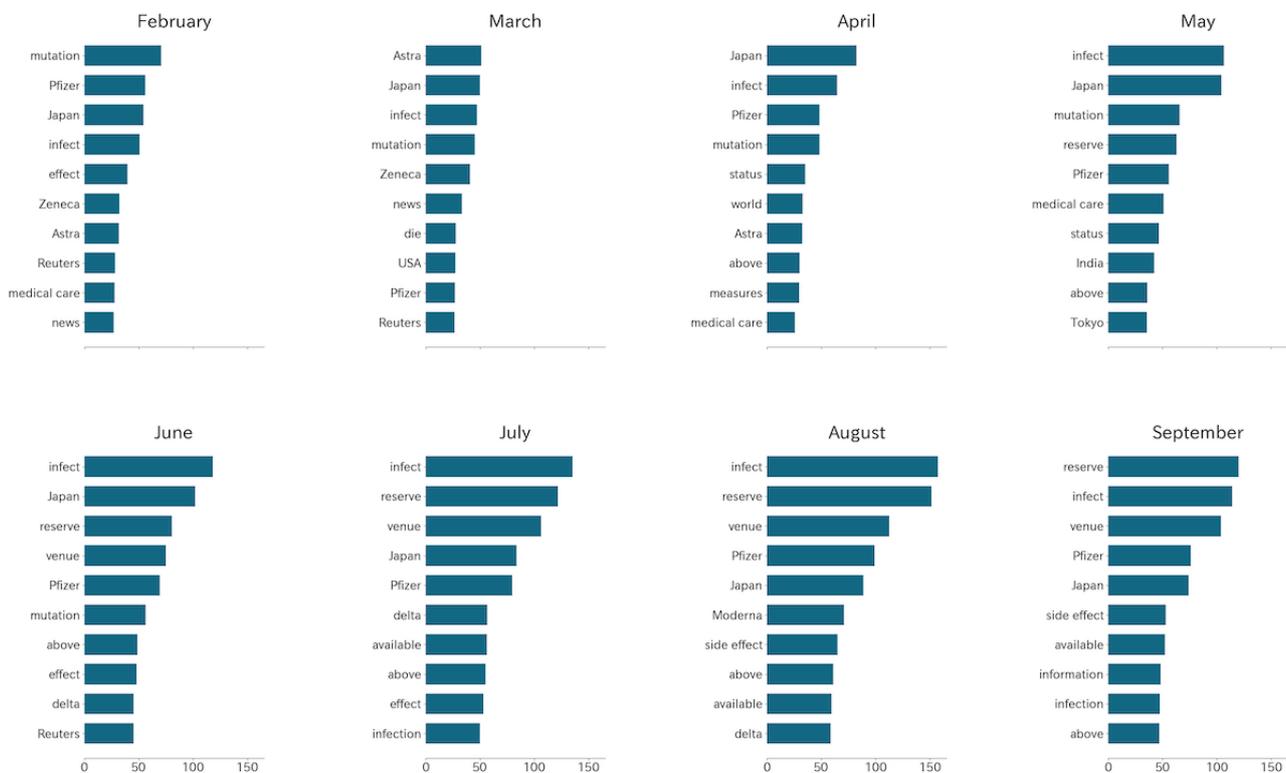


Figure 2. Translation of the top 10 bigrams of each month. The lengths of the bars represent the monthly term frequencies in tweets of each month.

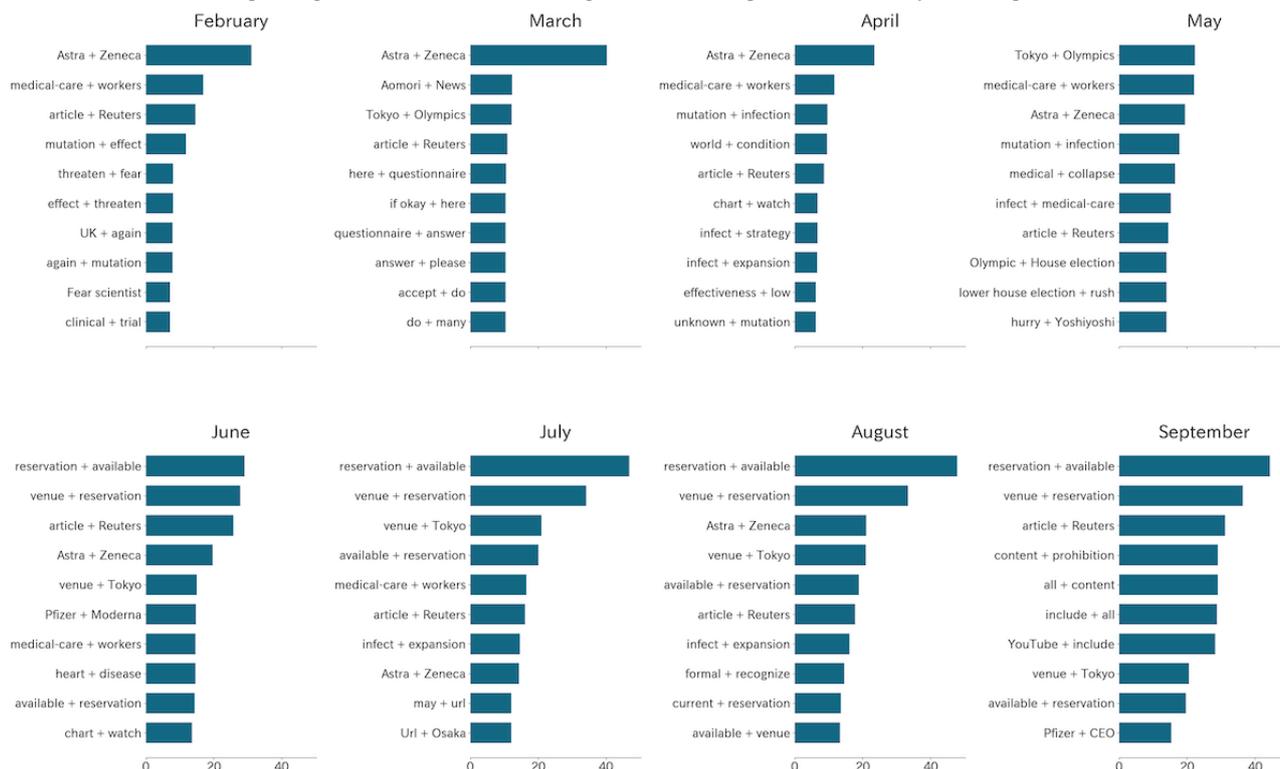
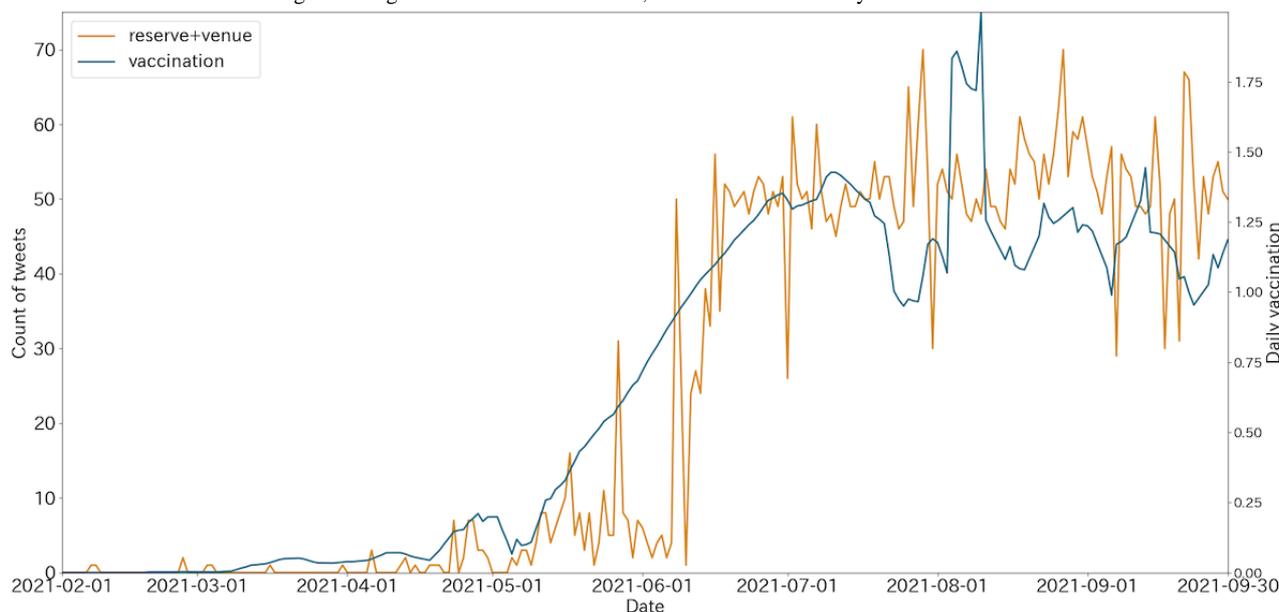


Figure 3. Trend of tweets containing both unigrams “reserve” and “venue,” and the curve for daily first-dose vaccination cases.



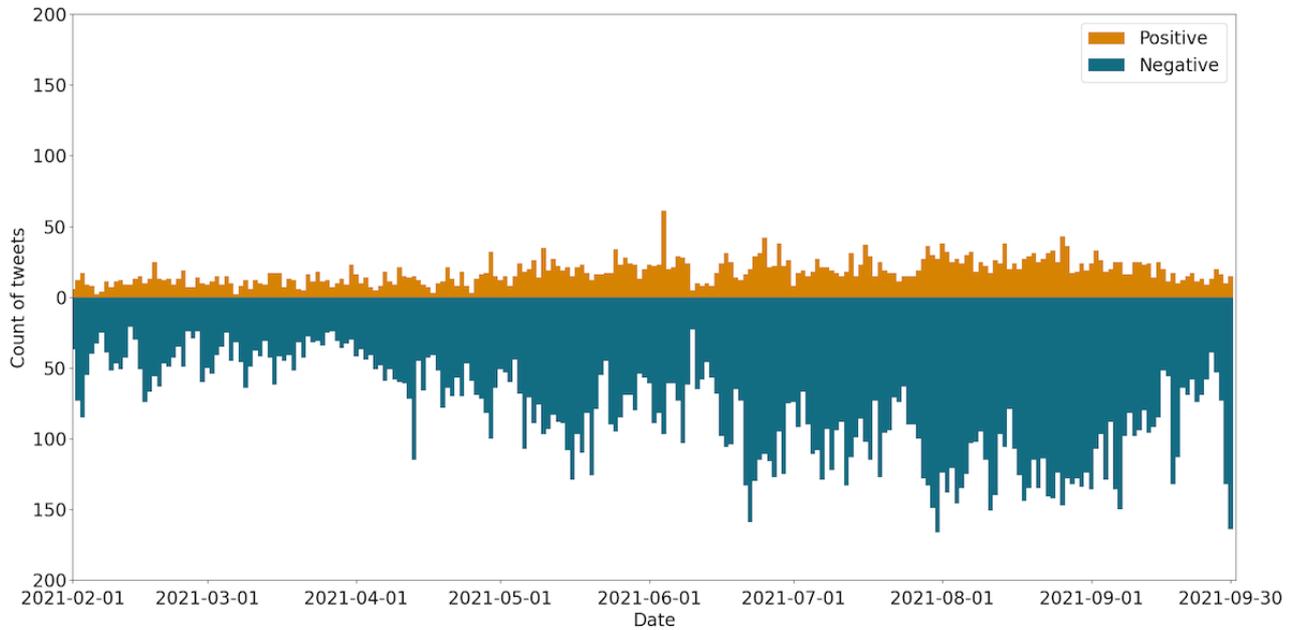
As for bigrams, the bigram “venue + reserve” overlapped with the unigram analysis and was excluded from this part. The time lags for bigrams “reserve + available” and “article + Reuters” were 0, and vaccination cases led “Astra + Zeneca” and “medical-care + workers” for 116 and 63 days, respectively. The bigrams “reserve + available” and “article + Reuters” had the highest cross-correlations than the others. The bigrams “Astra + Zeneca” ($r=-0.331$), “reserve + available” ($r=0.908$), and “article + Reuters” ($r=0.229$) showed significant correlations ($P<.001$) except for “medical-care + workers” ($r=-0.055$). On manual evaluation by 3 volunteers, we found that 95.4% of the

tweets that contain the bigrams “reserve + available” were the same as those of containing the combination of unigrams “venue” and “reserve.”

Sentiment Analysis

For all tweets, 4453 (2.3%) were positive, 19,340 (10.1%) were negative, 164,687 (86.4%) were neutral, and 2217 (1.2%) were mixed positive and negative sentiments. A comparison between the daily numbers of tweets marked as positive and negative is shown in Figure 4. Negative sentiments overwhelmed positive sentiments for all days.

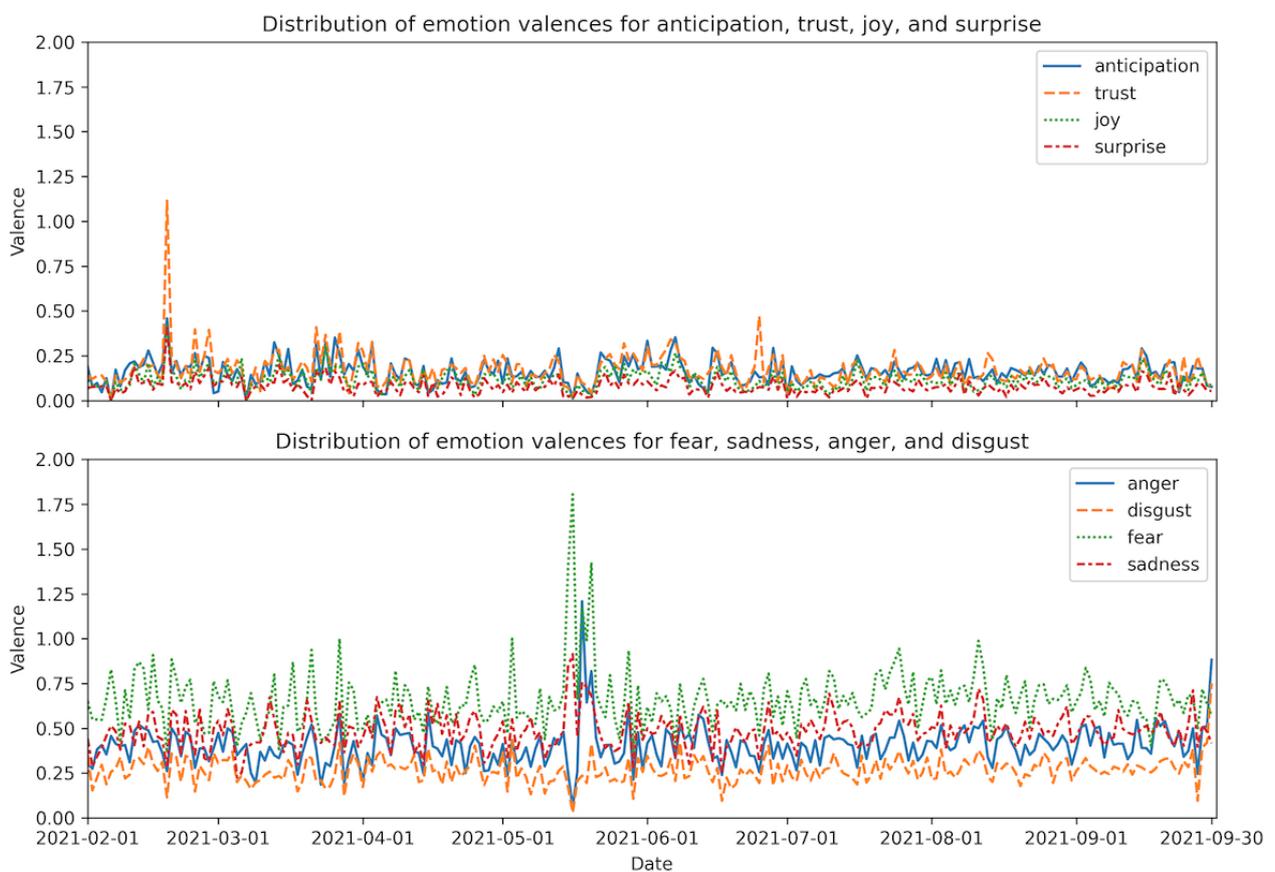
Figure 4. Comparison between the daily number of tweets marked positive (orange) and negative (green).



The DOVs for the 8 emotions are shown in Figure 5. The daily average DOV of anger (0.404), disgust (0.268), fear (0.659), sadness (0.486), overwhelmed anticipation (0.163), trust (0.173), joy (0.118), and surprise (0.081). Fear was the dominant emotion

during this period. Here, we defined the peaks of emotion as larger than 3 times the daily average DOV for that emotion. Trust peaked (1.114) on February 18, 2021. From May 13 to 18, 2021, there were several peaks of fear.

Figure 5. Daily average degree of valence of 8 emotions in the vaccine-related tweets.



Topic Modeling

The top 10 keywords for each LDA topic are shown in Figure 6. The theme of topic 1 is “infect,” and that of topic 2 is “vaccine confidence.” It is also noticeable that the weight of “infect” (14,895) in topic-1 was over 3 times that of the second keyword “Japan” (4359), but the weight of “Pfizer” (4348) in topic 2 was only 15.5% larger than the second keyword “die” (3763).

The ratio between the expectation of the number of “infect”-related tweets and “vaccine confidence”-related tweets

is shown in Figure 7. The total expectation of the number of tweets generated from topic 1 (“infect,” n=30,288) is larger than that generated from topic 2 (“vaccine confidence,” n=27,572), and the mean ratio between the expectation of the daily number of tweets generated from topics 1 and 2 is significantly larger than 1 ($P<.01$). On 68.2% of days, the expectation of the number of tweets generated from “infect” was larger than that generated from “vaccine confidence.”

Figure 6. Top 10 keywords of 2 topics by latent Dirichlet allocation modeling. The bars represent the weights, which can be regarded as the pseudocounts of the keywords in each topic.

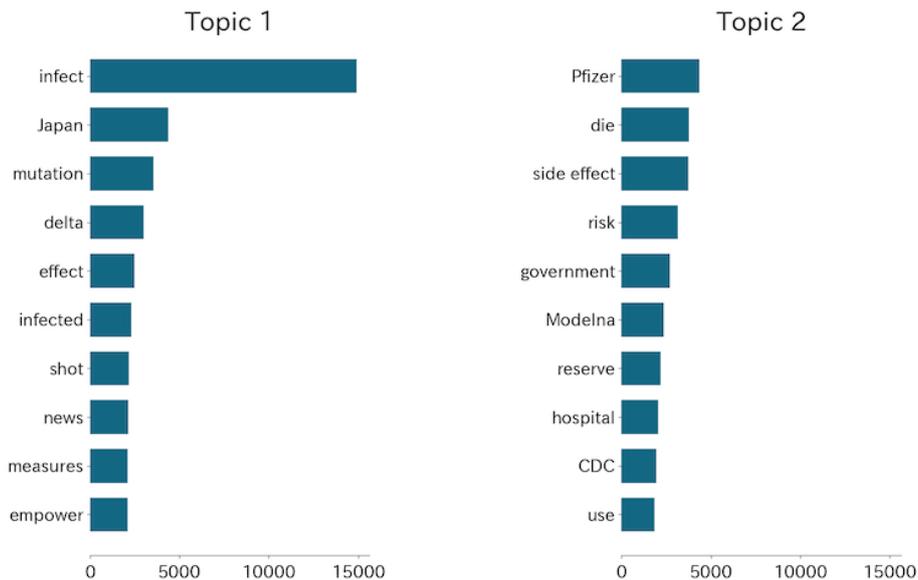
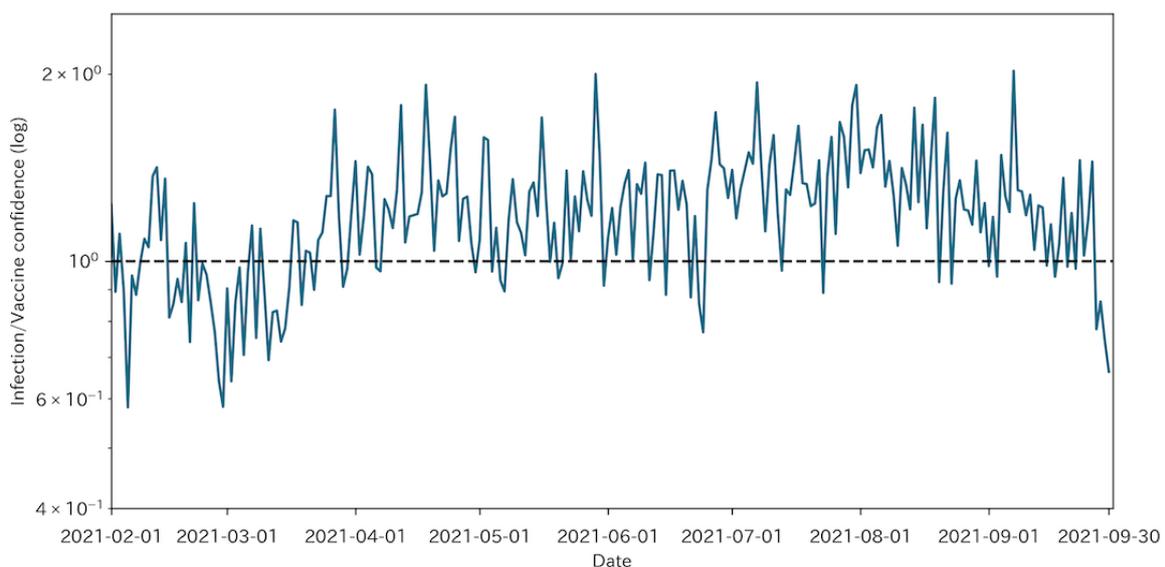


Figure 7. Ratio between the expectation of the number of “infection”-related tweets and “vaccine confidence”-related tweets.



Discussion

Principal Findings

A high vaccination rate is thought to be promoted by high vaccine confidence [16,33-36], but Japan achieved a high

vaccination rate in several months, with the lowest vaccine confidence in the world. This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which may be instructive for propelling worldwide vaccination for infectious diseases. Based on previous studies [16,34-37], we hypothesized that subjective factors, including increased

vaccine confidence (S1) and fear of infection (S2), and objective factors including adequate vaccine supply (O1) and effective delivery of reservation-related vaccine information (O2). Our results indicate that hypotheses S2 and O2 might have driven the public to be vaccinated. No evidence supporting hypothesis S1 was found in our results. Evidence for hypothesis O1 can be found in the history of vaccine supply on the official website of the PMOJ (Prime Minister of Japan and his Cabinet) and is not discussed in this paper.

Several results support hypothesis S2. In the unigram token analysis shown in [Figure 1](#), the keyword “infect” ranked among the top 3, except in February, and ranked first from May to August during Japan’s fourth and fifth wave infections. The keywords “venue” and “reserve” also ranked up from May. No keywords related to increased vaccine confidence were found. The sentiment analysis shown in [Figure 4](#) showed that negative sentiment overwhelmed positive sentiments, consistent with the results of [Chen et al \[35\]](#) that Japan showed dissatisfaction compared with neighboring countries. Combined with our result that “infect” was the top keyword and “side effect” ranked eighth in the unigram token analysis, our results support that the Japanese public was more concerned about infection than the side effects of COVID-19 vaccines.

More evidence for hypothesis S2 was obtained from the topic modeling results. From the keywords of topic 1 (“infect”), we can see that the public was concerned about the infection and death rate. The mutated virus and empowered cases also led to fear. [Willis et al \[37\]](#) found that less fear of infection may lead to a lower willingness to be vaccinated, which is complementary to our results. From the keywords of topic 2 (“vaccine confidence”), we can see that the side effects of the vaccines were the most concerning, but the following keywords were related to the effectiveness of vaccines on the mutated virus, reservation of vaccines, and medical care conditions. Previous surveys in different countries have indicated that fear of vaccine safety is the key factor for low vaccine acceptance [[38,39](#)]. Furthermore, the “side effect” weight in topic 2 was much less than that of “infect” in topic 1. The top keywords in the two topics indicated that people were more concerned about COVID-19 rather than the side effects of vaccines. [Bendau et al \[40\]](#) reported a significant positive correlation between fears of infection and vaccine acceptance and a significant negative correlation between fear of vaccine safety and vaccine acceptance. Therefore, it is important to distinguish the mainstream fear emotion to determine the reason for the high vaccination rate. [Figure 7](#) provides details about the ratio between the expected number of tweets related to “infect” and “vaccine confidence.” In most cases, the ratio was larger than 1, indicating that the public was more concerned about infection rather than the safety and effectiveness of vaccines. Higher ratios were observed in April and from July to end-September, which were periods of Japan’s fourth and fifth waves of infection. There was also a relatively long period of less than one ratio from mid-February to mid-March, which was the period when the vaccines were less effective against the mutated virus (February 10), severe side effects of the AstraZeneca vaccine were observed (March 12), and several side effects were observed in Japan (February 21, March 7, and March 10).

However, the ratio soon increased because of the fourth wave of infections. This example also proved that fear of infection overcame the vaccine safety concern.

We also provide evidence of a strong relationship between vaccination and hypothesis O2. Bigram analysis in [Figure 2](#) showed that “reservation + available” ranked first since June, shortly after large-scale vaccination started, which might reflect the strong concerns about vaccine reservation by the public. Unigram token analysis in [Figure 3](#) showed that tweets including the keywords “reserve” and “venue” were significantly highly correlated ($r>0.9$; $P<.01$) with the daily number of vaccination cases in Japan, and most of them were from government official accounts. The bigram “reservation + available” also showed a high correlation ($r>0.9$; $P<.01$) with the daily vaccination cases. Because reservation information should always lead to the actual vaccination, this result indicated that in addition to sufficient vaccine supply, reservation information delivery might also be important in large-scale vaccination. Furthermore, the time lag for the maximum cross-correlation was 0, which may indicate the efficiency of the reservation information posted on Twitter. Our results were consistent with [Fu’s \[41\]](#) finding that inflexible information systems for vaccine reservation can impair immunization services in the community.

We did not find any evidence for hypothesis S1. [Macaraan](#) reported a shift from hesitancy to confidence toward the COVID-19 vaccination program among Filipinos [[36](#)]. [Okubo \[42\]](#) reported a shift from hesitancy to confidence but also admitted that the shift might come from the differences in the survey metrics in previous studies [[43](#)]. Following these studies, we looked for a similar shift in sentiment or emotions from negative to positive, but negative sentiments overwhelmed positive sentiments as shown in [Figure 4](#), and fear dominates all the emotions in [Figure 5](#). The positive emotions “anticipation,” “trust,” and “joy” did not increase during the entire period. These two results made it difficult to conclude increased vaccine confidence.

Our results were partially related to the 5 C model (confidence, competence, convenience, calculation, and collective responsibility) measuring vaccine hesitancy [[36,44](#)]. Confidence and complacency are two subjective measures that are directly related to individuals. In our work, the LDA theme “vaccine confidence” belonged to “confidence,” and “fear of infection” belonged to “complacency.” In Japan, fear of infection may drive a high vaccination rate. The delivery of reservation information may be an extension to “convenience,” which was previously defined as “physical availability, affordability and willingness-to-pay, geographical accessibility, ability to understand (language and health literacy), and appeal of immunization service affect uptake” [[45](#)]. Our work indicates that information about vaccination reservations should also be considered for the convenience of vaccination.

Limitations

We admit that our research might have some potential limitations: (1) the imbalance of the demographics of Twitter users in Japan [[46](#)] may cause bias in the results; (2) the status of the user on a certain day (at home or not, other events on that day, etc) may also bias the data set [[47](#)]; (3) owing to the lack

of a reliable public model for sentiment analysis in the Japanese language, the cloud service AWS was used for sentiment analysis; (4) filtering keywords may include irrelevant or missing related tweets; (5) antivaccine tweets, especially rumors, were not distinguished or analyzed separately in this study. However, feature works can be combined with classical surveys to train the sentiment analysis model and model to distinguish rumors from tweets to overcome these limitations.

Conclusions

This retrospective study aimed to determine the reasons for the fast vaccination process in Japan, which might be instructive for propelling worldwide vaccination toward infectious diseases.

In conclusion, our work indicated that awareness of the danger of COVID-19 increased the willingness to be vaccinated; with a sufficient supply of vaccines, effective reservation information delivery might provide more opportunities for people to be vaccinated. Models measuring vaccine hesitancy might also need to add efficiency in delivering reservation information as a metric. Based on our findings, we recommend public health policy makers and the government to share accurate and prompt information about the infectious diseases and vaccination. Furthermore, efforts on tied cooperation among multilevel relevant organizations and new media operations may help achieve smoother delivery of vaccine reservation information.

Acknowledgments

This work was supported by the JST SPRING (grant number JPMJSP2110).

Authors' Contributions

QN and JL performed analyses and drafted the manuscript. All authors conceived the study, interpreted the results, and revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The keywords used for the selection of vaccine-related tweets and the corresponding translation; English translations used in our paper and the corresponding original Japanese words; Mean Log likelihood scores for different LDA topic numbers using five-fold cross-validation.

[\[DOCX File, 29 KB-Multimedia Appendix 1\]](#)

References

1. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. *Soc Sci Med* 2020 Oct;263:113356 [FREE Full text] [doi: [10.1016/j.socscimed.2020.113356](https://doi.org/10.1016/j.socscimed.2020.113356)] [Medline: [32967786](https://pubmed.ncbi.nlm.nih.gov/32967786/)]
2. Experts warn of 6th wave as COVID cases decrease in Japan but rise overseas. *The Mainichi*. 2021 Oct 30. URL: <https://mainichi.jp/english/articles/20211030/p2a/00m/0na/020000c> [accessed 2022-02-22]
3. Fifth wave had fewer big clusters possibly due to vaccinations. *The Asahi Shimbun*. 2021 Oct 15. URL: <https://www.asahi.com/ajw/articles/14461519> [accessed 2022-02-22]
4. Larson H. Vaccine confidence and public trust as drivers of vaccine failure. *Int J Infect Dis* 2014 Apr;21:50. [doi: [10.1016/j.ijid.2014.03.522](https://doi.org/10.1016/j.ijid.2014.03.522)]
5. What Is Vaccine Confidence? Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html> [accessed 2022-02-22]
6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. *Nat Med* 2021 Feb 20;27(2):225-228 [FREE Full text] [doi: [10.1038/s41591-020-1124-9](https://doi.org/10.1038/s41591-020-1124-9)] [Medline: [33082575](https://pubmed.ncbi.nlm.nih.gov/33082575/)]
7. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. *The Lancet* 2020 Sep;396(10255):898-908. [doi: [10.1016/s0140-6736\(20\)31558-0](https://doi.org/10.1016/s0140-6736(20)31558-0)]
8. Mahase E. Covid-19: UK has highest vaccine confidence and Japan and South Korea the lowest, survey finds. *BMJ* 2021 Jun 04;373:n1439. [doi: [10.1136/bmj.n1439](https://doi.org/10.1136/bmj.n1439)] [Medline: [34088703](https://pubmed.ncbi.nlm.nih.gov/34088703/)]
9. Gordon A, Reich MR. The Puzzle of Vaccine Hesitancy in Japan. *J Jpn Stud* 2021;47(2):411-436. [doi: [10.1353/jjs.2021.0047](https://doi.org/10.1353/jjs.2021.0047)]
10. Kunitoki K, Funato M, Mitsunami M, Kinoshita T, Reich MR. Access to HPV vaccination in Japan: Increasing social trust to regain vaccine confidence. *Vaccine* 2021 Oct 01;39(41):6104-6110 [FREE Full text] [doi: [10.1016/j.vaccine.2021.08.085](https://doi.org/10.1016/j.vaccine.2021.08.085)] [Medline: [34507858](https://pubmed.ncbi.nlm.nih.gov/34507858/)]
11. Ritchie H, Mathieu E, Rod s-Guirao L, Appel C, Giattino C, Ortiz-Ospina E. Coronavirus Pandemic (COVID-19). *Our World in Data*. URL: <https://ourworldindata.org/covid-vaccinations> [accessed 2022-01-22]

12. 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* 2021 Jul;108:256-262 [FREE Full text] [doi: [10.1016/j.ijid.2021.05.059](https://doi.org/10.1016/j.ijid.2021.05.059)] [Medline: [34052407](https://pubmed.ncbi.nlm.nih.gov/34052407/)]
13. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health* 2017 Jan;107(1):e1-e8. [doi: [10.2105/AJPH.2016.303512](https://doi.org/10.2105/AJPH.2016.303512)] [Medline: [27854532](https://pubmed.ncbi.nlm.nih.gov/27854532/)]
14. 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 [FREE Full text] [doi: [10.2196/24435](https://doi.org/10.2196/24435)] [Medline: [34115608](https://pubmed.ncbi.nlm.nih.gov/34115608/)]
15. Huangfu L, Mo Y, Zhang P, Zeng DD, He S. COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling. *J Med Internet Res* 2022 Feb 08;24(2):e31726 [FREE Full text] [doi: [10.2196/31726](https://doi.org/10.2196/31726)] [Medline: [34783665](https://pubmed.ncbi.nlm.nih.gov/34783665/)]
16. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. *J Med Internet Res* 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: [10.2196/28973](https://doi.org/10.2196/28973)] [Medline: [33872185](https://pubmed.ncbi.nlm.nih.gov/33872185/)]
17. Social Media Stats Japan. StatCounter. URL: <https://gs.statcounter.com/social-media-stats/all/japan> [accessed 2022-01-22]
18. Leading countries based on number of Twitter users as of January 2022. Statista. URL: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> [accessed 2022-01-22]
19. Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T, et al. Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. *JMIR Infodemiology* 2022;2(1):e32335 [FREE Full text] [doi: [10.2196/32335](https://doi.org/10.2196/32335)] [Medline: [35578643](https://pubmed.ncbi.nlm.nih.gov/35578643/)]
20. 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/)]
21. Yang X, Sornlertlamvanich V. Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis. 2021 Presented at: 2021 International Electronics Symposium (IES); September 29-30, 2021; Surabaya. [doi: [10.1109/ies53407.2021.9594036](https://doi.org/10.1109/ies53407.2021.9594036)]
22. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, et al. Public Perception of COVID-19 Vaccines on Twitter in the United States. medRxiv Preprint posted online October 18, 2021. [FREE Full text] [doi: [10.1101/2021.10.16.21265097](https://doi.org/10.1101/2021.10.16.21265097)] [Medline: [34704100](https://pubmed.ncbi.nlm.nih.gov/34704100/)]
23. Monselise M, Chang C, Ferreira G, Yang R, Yang CC. Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. *J Med Internet Res* 2021 Oct 21;23(10):e30765 [FREE Full text] [doi: [10.2196/30765](https://doi.org/10.2196/30765)] [Medline: [34581682](https://pubmed.ncbi.nlm.nih.gov/34581682/)]
24. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine* 2021 Sep 15;39(39):5499-5505 [FREE Full text] [doi: [10.1016/j.vaccine.2021.08.058](https://doi.org/10.1016/j.vaccine.2021.08.058)] [Medline: [34452774](https://pubmed.ncbi.nlm.nih.gov/34452774/)]
25. Shim J, Ryu K, Lee SH, Cho E, 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 [FREE Full text] [doi: [10.3390/ijerph18126549](https://doi.org/10.3390/ijerph18126549)] [Medline: [34207016](https://pubmed.ncbi.nlm.nih.gov/34207016/)]
26. COVID_fear. GitHub. URL: https://github.com/juniorliu95/COVID_fear [accessed 2022-06-06]
27. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. arXiv Preprint posted online April 7, 2020. [Medline: [32550247](https://pubmed.ncbi.nlm.nih.gov/32550247/)]
28. Prime Minister of Japan and His Cabinet. URL: <https://japan.kantei.go.jp/> [accessed 2022-02-22]
29. Bird S, Klein E, Loper E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Newton, MA: O'Reilly Media, Inc; 2009.
30. Engbreth G. *Bug Fixes*. In: *PHP 8 Revealed*. Berkeley, CA: Apress; 2021:87-99.
31. Dorcas Wambui G. The Power of the Pruned Exact Linear Time (PELT) Test in Multiple Changepoint Detection. *AJTAS* 2015;4(6):581. [doi: [10.11648/j.ajtas.20150406.30](https://doi.org/10.11648/j.ajtas.20150406.30)]
32. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013;29:465. [doi: [10.1111/j.1467-8640.2012.00460.x](https://doi.org/10.1111/j.1467-8640.2012.00460.x)]
33. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique* 2014 Apr 30;17(1):61-84. [doi: [10.3166/dn.17.1.61-84](https://doi.org/10.3166/dn.17.1.61-84)]
34. Mori H, Naito T. A rapid increase in the COVID-19 vaccination rate during the Olympic and Paralympic Games 2021 in Japan. *Hum Vaccin Immunother* 2022 Dec 31;18(1):2010440-2010442. [doi: [10.1080/21645515.2021.2010440](https://doi.org/10.1080/21645515.2021.2010440)] [Medline: [34893009](https://pubmed.ncbi.nlm.nih.gov/34893009/)]
35. Chen CW, Lee S, Dong MC, Taniguchi M. What factors drive the satisfaction of citizens with governments' responses to COVID-19? *Int J Infect Dis* 2021 Jan;102:327-331 [FREE Full text] [doi: [10.1016/j.ijid.2020.10.050](https://doi.org/10.1016/j.ijid.2020.10.050)] [Medline: [33115678](https://pubmed.ncbi.nlm.nih.gov/33115678/)]
36. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. *PLoS One* 2018 Dec 7;13(12):e0208601 [FREE Full text] [doi: [10.1371/journal.pone.0208601](https://doi.org/10.1371/journal.pone.0208601)] [Medline: [30532274](https://pubmed.ncbi.nlm.nih.gov/30532274/)]
37. Willis DE, Andersen JA, Bryant-Moore K, Selig JP, Long CR, Felix HC, et al. COVID-19 vaccine hesitancy: Race/ethnicity, trust, and fear. *Clin Transl Sci* 2021 Nov;14(6):2200-2207 [FREE Full text] [doi: [10.1111/cts.13077](https://doi.org/10.1111/cts.13077)] [Medline: [34213073](https://pubmed.ncbi.nlm.nih.gov/34213073/)]

38. Pugliese-Garcia M, Heyerdahl LW, Mwamba C, Nkwemu S, Chilengi R, Demolis R, et al. Factors influencing vaccine acceptance and hesitancy in three informal settlements in Lusaka, Zambia. *Vaccine* 2018 Sep 05;36(37):5617-5624 [FREE Full text] [doi: [10.1016/j.vaccine.2018.07.042](https://doi.org/10.1016/j.vaccine.2018.07.042)] [Medline: [30087047](https://pubmed.ncbi.nlm.nih.gov/30087047/)]
39. Robertson E, Reeve KS, Niedzwiedz CL, Moore J, Blake M, Green M, et al. Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study. *Brain Behav Immun* 2021 May;94:41-50 [FREE Full text] [doi: [10.1016/j.bbi.2021.03.008](https://doi.org/10.1016/j.bbi.2021.03.008)] [Medline: [33713824](https://pubmed.ncbi.nlm.nih.gov/33713824/)]
40. Bendau A, Plag J, Petzold MB, Ströhle A. COVID-19 vaccine hesitancy and related fears and anxiety. *Int Immunopharmacol* 2021 Aug;97:107724 [FREE Full text] [doi: [10.1016/j.intimp.2021.107724](https://doi.org/10.1016/j.intimp.2021.107724)] [Medline: [33951558](https://pubmed.ncbi.nlm.nih.gov/33951558/)]
41. Fu C. Milestone and challenges: lessons from defective vaccine incidents in China. *Hum Vaccin Immunother* 2020 Sep 06;16(1):80-80 [FREE Full text] [doi: [10.1080/21645515.2019.1646579](https://doi.org/10.1080/21645515.2019.1646579)] [Medline: [31339795](https://pubmed.ncbi.nlm.nih.gov/31339795/)]
42. Okubo R, Yoshioka T, Ohfuji S, Matsuo T, Tabuchi T. COVID-19 Vaccine Hesitancy and Its Associated Factors in Japan. *Vaccines (Basel)* 2021 Jun 17;9(6):662 [FREE Full text] [doi: [10.3390/vaccines9060662](https://doi.org/10.3390/vaccines9060662)] [Medline: [34204465](https://pubmed.ncbi.nlm.nih.gov/34204465/)]
43. Machida M, Nakamura I, Kojima T, Saito R, Nakaya T, Hanibuchi T, et al. Acceptance of a COVID-19 Vaccine in Japan during the COVID-19 Pandemic. *Vaccines (Basel)* 2021 Mar 03;9(3) [FREE Full text] [doi: [10.3390/vaccines9030210](https://doi.org/10.3390/vaccines9030210)] [Medline: [33802285](https://pubmed.ncbi.nlm.nih.gov/33802285/)]
44. Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batouré O, Anya BM, et al. Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future? *Hum Vaccin Immunother* 2022 Dec 31;18(1):1-3 [FREE Full text] [doi: [10.1080/21645515.2021.1893062](https://doi.org/10.1080/21645515.2021.1893062)] [Medline: [33684019](https://pubmed.ncbi.nlm.nih.gov/33684019/)]
45. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. *Vaccine* 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: [10.1016/j.vaccine.2015.04.036](https://doi.org/10.1016/j.vaccine.2015.04.036)] [Medline: [25896383](https://pubmed.ncbi.nlm.nih.gov/25896383/)]
46. Statista. URL: <https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/> [accessed 2022-04-30]
47. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PLoS One* 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: [10.1371/journal.pone.0198857](https://doi.org/10.1371/journal.pone.0198857)] [Medline: [29902270](https://pubmed.ncbi.nlm.nih.gov/29902270/)]

Abbreviations

amp: ampersands
AWS: Amazon Web Services
DOV: degree of valence
LDA: latent Dirichlet allocation
NLP: natural language processing
PELT: pruned exact linear time
PMOJ: Prime Minister's Office of Japan

Edited by M Gisondi, J Faust; submitted 22.02.22; peer-reviewed by C Tsagkaris, R Gore, V Ritschl; comments to author 19.04.22; revised version received 09.05.22; accepted 30.05.22; published 09.06.22

Please cite as:

Niu Q, Liu J, Kato M, Nagai-Tanima M, Aoyama T

The Effect of Fear of Infection and Sufficient Vaccine Reservation Information on Rapid COVID-19 Vaccination in Japan: Evidence From a Retrospective Twitter Analysis

J Med Internet Res 2022;24(6):e37466

URL: <https://www.jmir.org/2022/6/e37466>

doi: [10.2196/37466](https://doi.org/10.2196/37466)

PMID: [35649182](https://pubmed.ncbi.nlm.nih.gov/35649182/)

©Qian Niu, Junyu Liu, Masaya Kato, Momoko Nagai-Tanima, Tomoki Aoyama. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.06.2022. 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 the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.