

1 **Fear of Infection and Sufficient Vaccine Reservation Information Might Drive Rapid**
2 **Coronavirus Disease 2019 Vaccination in Japan: Evidence from Twitter Analysis**

3
4 Qian Niu,^{1*} MS; Junyu Liu,^{2*} ME; Masaya Kato,¹ BS; Tomoki Aoyama,¹ MD, PhD; Momoko
5 Nagai-Tanima,¹ PhD

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

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

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

11
12
13
14
15 **Corresponding Author:**

16 Momoko Nagai-Tanima, PT, PhD

17 Department of Human Health Sciences,

18 Graduate School of Medicine,

19 Kyoto University

20 53, Kawahara-Cho, Shogoin

21 Sakyo-Ku

22 Kyoto, 606-8507

23 Japan

24 Phone: 075-751-3964

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

44 **Abstract**

45 **Background:** The global public health and socioeconomic impacts of coronavirus disease 2019
46 (COVID-19) have been substantial, making herd immunity by COVID-19 vaccination an
47 important factor for protecting people and retrieving the economy. Among all the countries,
48 Japan became one of the countries with the highest COVID-19 vaccination rate in several
49 months, although the vaccine confidence in Japan is the lowest worldwide.

50
51 **Objective:** We attempted to find the reasons for the rapid coronavirus disease 2019 (COVID-19)
52 vaccination in Japan under the lowest vaccine confidence in the world by Twitter analysis.

53
54 **Materials and methods:** We downloaded COVID-19 related Japanese tweets from a large-scale
55 public COVID-19 Twitter chatter dataset within the timeline of February 1, 2021 and September
56 30, 2021. The daily number of vaccination cases was collected from the official website of the
57 Prime Minister's Office of Japan. After preprocessing, we applied unigram and bigram token
58 analysis, then calculated the cross correlation and Pearson correlation coefficient (r) between the
59 term frequency and daily vaccination cases. Then we identified vaccine sentiments and emotions
60 of tweets and used the topic modeling to look deeper into the dominant emotions.

61
62 **Results:** We selected 190,697 vaccine-related tweets after filtering. By n-gram token analysis, we
63 discovered the top unigrams and bigrams over the whole period. In all the combinations of the top
64 six unigrams, tweets with both keywords “reserve” and “venue” showed the largest $r = 0.912$ ($P <$
65 0.001) with the daily vaccination cases. In sentiment analysis, negative sentiment overwhelmed
66 positive sentiment, and fear was the dominant emotion across the period. For the latent Dirichlet
67 allocation model on tweets with fear emotion, the two topics were identified as “infect” and
68 “vaccine confidence”. The expectation of the number of tweets generated from topic “infect” was
69 larger than “vaccine confidence.”

70
71 **Conclusion:** Our work indicated that awareness of the danger of COVID-19 might increase the
72 willingness to get vaccinated; With sufficient vaccine supply, effective vaccine reservation
73 information delivery may be an important factor for people to get vaccinated; We didn't find
74 evidence for increased vaccine confidence in Japan during the period in our research. We
75 recommend policymakers to share fair and prompt information about the infectious diseases and
76 vaccination, and make efforts on smoother delivery of vaccine-reservation information.

77
78
79 **Keywords:** COVID-19, vaccine hesitancy, Japan, social media, text mining

80 1. Introduction

81 Coronavirus disease 2019 (COVID-19) has spread worldwide since its first case in December 2019
82 and has become a public health emergency of international concern [1]. Until September 30, 2021,
83 Japan experienced five waves of the COVID-19 pandemic [2]. The surge of COVID-19 in Japan
84 occurred during the Tokyo Olympics, bringing the cumulative number of COVID-19 cases to
85 1,556,998 when the Games finished. However, with the lifting of the fourth national state of
86 emergency on September 30, 2021, the nationwide pandemic was effectively contained, and the
87 number of new confirmed cases abruptly decreased. The high vaccination rate in Japan was
88 considered an important factor in stopping the fifth wave [3].

89
90 A high vaccination rate is thought to be promoted by high vaccine confidence [4]. Vaccine
91 confidence is the belief that vaccines work, are safe, and are part of a trustworthy medical system
92 [5]. A global survey that did not include Japan showed that the potential acceptance of a COVID-
93 19 vaccine largely varied among countries [6]. Japan ranks among the countries with the lowest
94 vaccine confidence worldwide according to a survey in 2020 [7]. Another survey conducted before
95 the large-scale vaccination in Japan indicated that Japan ranked last in confidence in COVID-19
96 vaccines among 15 countries [8]. Gordon and Reich explained the historical reasons for low
97 vaccine confidence in Japan [9]. Kunitoki et al. proposed that barriers to vaccine access and use
98 are mainly from effective public communication and called for rebuilding vaccine confidence in
99 Japan [10].

100
101 However, Japan's speed of vaccination has been impressive since the large-scale vaccination (May
102 24, 2021). Japan's first dose vaccination rate was approximately 6.8% by June 1, 2021, and over
103 70% of the population accepted at least one dose until September 30, 2021 [11]. Notably,
104 vaccination was not mandatory and was administered only with the recipient's consent. A survey
105 of multiple countries reported the coexistence of a high level of uncertainty about the safety of
106 COVID-19 vaccines and a high willingness to get vaccinated [12], which indicates that Japan may
107 not be a special case. The reason for the contradiction between the rapid growth of the COVID-19
108 vaccination rate and low vaccine confidence in Japan is worth studying and maybe instructive for
109 propelling worldwide vaccination towards infectious diseases.

110
111 Twitter is a widespread social media platform that has attracted the increasing attention of public
112 health researchers because of its advantages of large amounts, real-time availability, and ease of
113 public searching and accessing [13]. With a large amount of real-time COVID-19-related posts,
114 Twitter has been widely used for public opinion mining towards COVID-19 during the pandemic,
115 providing policymakers with substantiated evidence [12,14,15]. Lyu et al. reported the trend of
116 topics and sentiments of English tweets for approximately 11 months since the World Health
117 Organization declared the COVID-19 pandemic [14]. Yousefinaghani et al. reported the
118 dominance of positive sentiments and more vaccine objection and hesitancy than vaccine interest
119 [12]. Huangfu et al. reported the results of topic modeling and sentiment analysis of tweets between
120 December 8, 2020 and April 8, 2021 [15]. Eibensteiner et al. reported willingness to vaccinate
121 despite the safety concerns of vaccines, according to a survey on Twitter Poll [16]. Besides, Twitter
122 is the most popular social media platform in Japan [17], owning 58.2 million users as of October
123 2021[18], making Twitter analysis more powerful for COVID-19 research in Japan. A Twitter
124 analysis by Niu et al. reported that the Japanese public's negative sentiment overwhelmed the

125 positive sentiment towards the COVID-19 vaccine before and at the beginning of the large-scale
126 vaccination campaign [19].

127
128 This retrospective study aimed to identify the reasons for the rapid COVID-19 vaccination in Japan
129 by analyzing tweets. We hypothesized that the increase in vaccination rates might be due to
130 subjective factors, including S1) increased public confidence in vaccines and S2) fear of infection,
131 and objective factors, including O1) adequate vaccine supply and O2) effective delivery
132 reservation-related vaccine information. To test these hypotheses, we collected vaccine-related
133 tweets posted between February 1, 2021, and September 30, 2021. Then, we preprocessed the
134 collected tweets and conducted a unigram token analysis, sentiment analysis, and topic modeling.

135 **2. Method**

136 In previous works of large-scale Twitter analyses, after preprocessing, there are mainly four
137 types of NLP methods: n-gram token analysis [12,15,20,21], sentiment analysis [12,14,15,20–
138 25], topic modelling [12,14,15,20,22–25], and geographical analysis [22,24]. The geographical
139 analysis is less important in our work because the range of our research is a whole country
140 instead of subareas. In this work, we followed the previous works in applying the n-gram token
141 analysis, sentiment analysis, and topic modeling.

142 **2.1 Data collection and preprocessing**

143 The data used in this study were obtained from a large-scale public COVID-19 Twitter chatter
144 dataset [26] updated by the Georgia State University's Panacea Lab. The dataset provided the
145 identifications (IDs), posting time, and the languages of all the tweets were provided in the dataset.
146 We downloaded COVID-related Japanese tweets between February 1, 2021, the month the first
147 person was vaccinated, and September 30, 2021, when the first-dose vaccination rate exceeded
148 70%. In addition, data on the number of vaccination cases were collected from the official website
149 of the Prime Minister's Office of Japan (PMOJ) [27].

150
151 The downloaded tweets were then cleaned and processed. Retweets were filtered using the Python
152 package tweepy. Tweets that included no keywords related to vaccines were deleted. It worth
153 notice that the three vaccine brands (Pfizer, Moderna, and AstraZeneca) that were approved by the
154 Japanese government were included in the keywords. Other vaccine brands were excluded because
155 we tried to focus more on the brands adopted in the vaccination process. Frequent misspellings
156 (e.g. Modelna) was also included in the keywords. Weblinks, special characters, emojis, and "amp"
157 (ampersands) were removed, and all full-width English characters were transferred to half-width
158 and lowercase.

159
160 For convenience, all Japanese words in our results were directly presented in English translations.
161 In order to minimize the influence of difference between languages, all the translations in our
162 results were done as the last step by directly replacing the Japanese words in the graphs with
163 corresponding English words, and therefore won't influence the statistical results.

164

165 **2.2 Unigram and bigram token analysis**

166 Tokenization is necessary before many other natural language processing (NLP) tasks, especially
167 for many non-Latin languages, such as Japanese. We removed the pre-defined English, and
168 Japanese stop words in the Python packages NLTK [28] and SpaCy [29], and tokenized all
169 collected vaccine-related tweets using the Python package SpaCy into unigrams or bigrams for
170 statistical analysis, as in the work of Kwok et al. [26]. We sorted the unigram tokens or bigram
171 tokens in descending order of term frequency over the entire period. Similar to Liu et al. [24], we
172 used the pruned exact linear time (PELT) algorithm [30] to find the first change point of the term
173 frequency. Unigrams before the first change point were regarded as top unigrams, and the term
174 frequencies of the unigrams after the change point were significantly lower than the top unigrams.
175 Similar processes were done for bigrams. To eliminate the difference in the number of days
176 between months, the monthly term frequency was defined by dividing the total term frequency by
177 the number of days each month for each top unigram or bigram.

178
179 Correlation coefficients were widely used in social media analysis. In Google Trends analysis,
180 correlations were calculated between reported cases of infectious disease and the trends of search
181 for relevant keywords. In Twitter analyses, correlations between the daily infectious or death cases,
182 and the number of related tweets or sentiment scores, were also investigated [24,25]. In this work,
183 correlation analyses were adopted to find out the factors from the top unigrams that are most related
184 to COVID-19 vaccination campaign. we first calculated the cross correlations between the number
185 of tweets containing the top unigrams or bigrams and the vaccination cases, then observed the time
186 lags when maximum cross correlation appeared for each unigram and bigram. Pearson correlation
187 coefficients (r) between top unigrams or bigrams and the vaccination cases were also calculated.
188

189 **2.3 Sentiment analysis**

190 After n-gram analysis, sentiment analyses were often used to explore the real-time public attitudes
191 in social media analysis related to COVID-19 vaccination, which may reflect the acceptance of
192 COVID-19 vaccines and related policies [12,14,20,22,24]. Trend of negative sentiments may
193 provide potential evidence for vaccine hesitancy [23]. In this work, sentiment analysis was applied
194 to all vaccine-related tweets. Cloud services were used in this study because there were no reliable
195 public models for sentiment analysis in the Japanese language. We selected the Amazon Web
196 Service (AWS) for consistency with previous work [28]. The tweets were divided into positive,
197 negative, neutral, or a mixture of positive and negative tweets using the AWS. Fine-grained
198 emotions were also explored using the Japanese version of the NRC Emotion Lexicon [31]. The
199 NRC Emotion Lexicon is a dictionary of words and their associated scores for eight emotions:
200 anticipation, trust, joy, surprise, anger, disgust, fear, and sadness. The positive and negative tweets
201 were tokenized, and the degree of valence (DOV) for the eight emotions was calculated by adding
202 up the scores for the unigrams that appeared in the NRC Emotion Lexicon. Finally, we calculated
203 the daily average DOV by dividing the number of positive and negative tweets on that day to show
204 the trend of each emotion.

205 **2.4 Topic modeling**

206 Topic modeling were applied to identify fine-grained information from tweets of different
207 sentiments [12,15,24]. Based on the sentiment analysis results, we summarized the topics to look

208 deeper into the dominant emotion in the tweets. Latent Dirichlet allocation (LDA) is often used in
209 tweets topic modeling studies [14,15,20,22,23]. In this study, LDA regards tweets as being
210 generated from different topics, and each topic generates tweets with a Dirichlet distribution. A
211 Python package scikit-learn was used to determine the best number of topics. Log likelihood was
212 adopted as the metric for selection, and five-fold cross correlation was applied to avoid overfitting.
213 According to figure 1 in the Appendix, we chose two as the number of topics for LDA modelling,
214 which showed highest log likelihood score. We used scikit-learn for the LDA topic modeling and
215 displayed the top ten keywords and their weights related to each topic. The weights were the
216 pseudo-counts of the keywords in a topic. The themes of topics were summarized from the top ten
217 keywords by three volunteers. The volunteers were first asked to work out the themes of the topics
218 independently, and then they had a meeting to finally reach an agreement on the themes.

219
220 We then checked the trends of tweets related to different topics. Defining the i -th tweet in all
221 collected tweets as d_i and the j -th topic of the LDA model as t_j , the probability of a tweet d_i
222 coming from t_j was calculated using the fitted LDA model as p_{ij} . For tweets posted each day, the
223 expectation of the number of tweets generated from topic j was calculated by summing p_{ij} on that
224 day. The ratio between the expected number of tweets generated from each topic was also plotted
225 to show the trend of public attention under dominant emotion.

226 **3. Results**

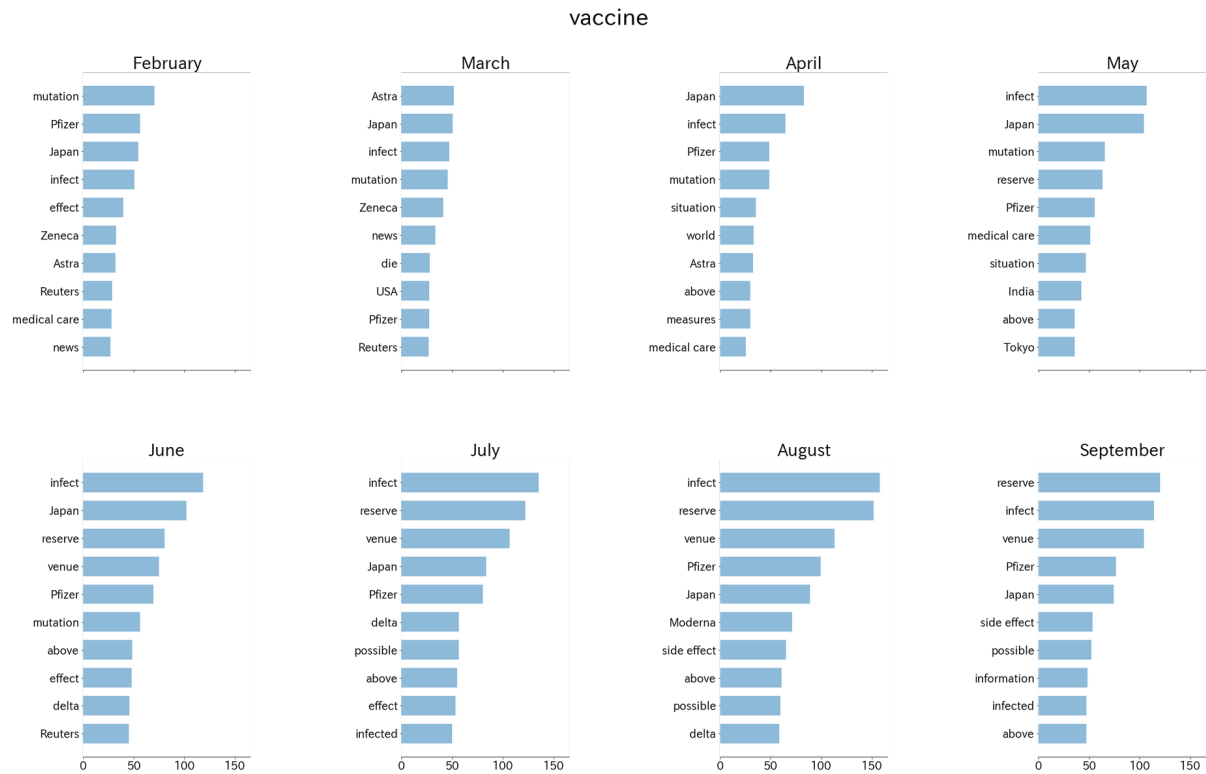
227 **3.1 Data summary**

228 We downloaded 979,636 Japanese tweets posted between February 1, 2021, and September 30,
229 2021, according to the ID and region information in the dataset. After filtering, 190,697 vaccine-
230 related tweets were selected. As a result, the total number of vaccine-related tweets increased from
231 14,758 tweets in February to 34,692 in August and then decreased to 27,824 in September.

232 **3.2 Unigram and bigram token analysis**

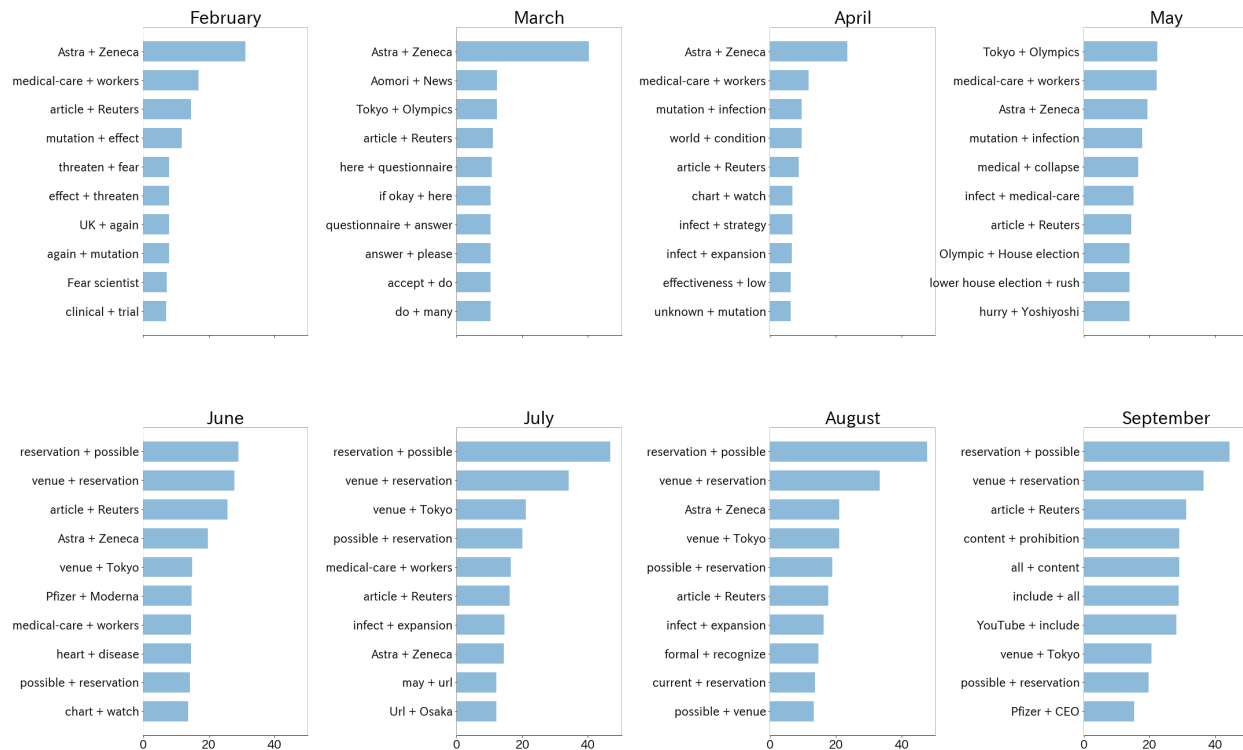
233 The change point of unigram term frequencies detected by the PELT algorithm was six, and the
234 top six unigrams were Japanese words for “infection,” “Japan,” “reserve,” “Pfizer,” “venue,” and
235 “mutation.” The unigram “side effects”, related to the safety of vaccines, ranked eighth overall.
236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

238



239
 240 Fig. 1. Translation of top ten unigrams of each month. The lengths of the bars represent the monthly term
 241 frequencies in tweets of each month.
 242

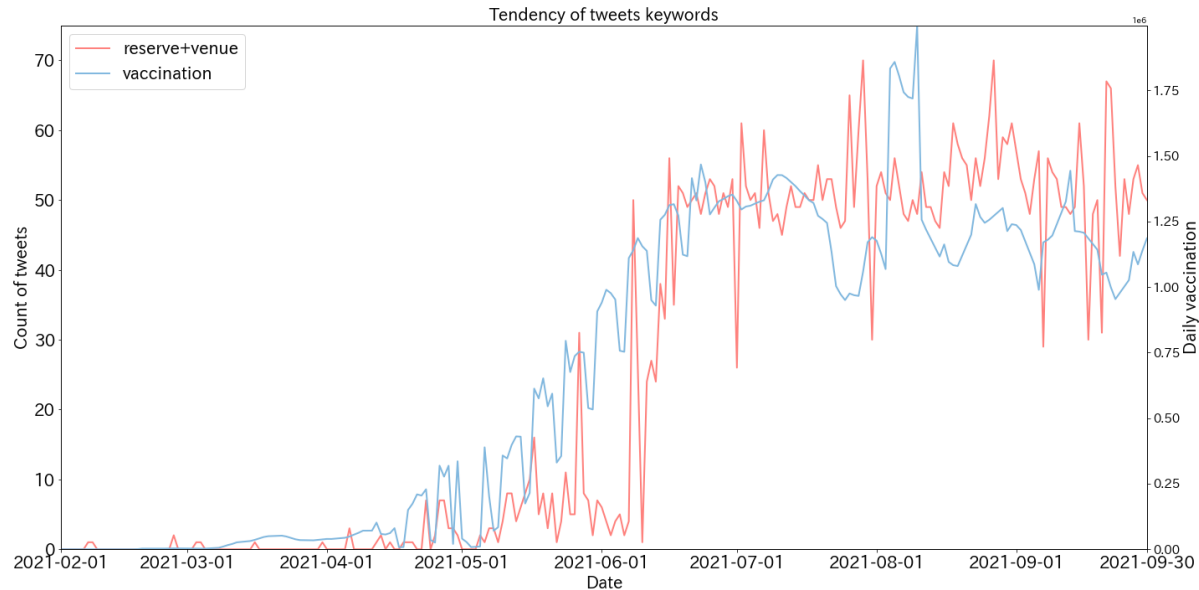
243 The change point of bigram term frequencies detected by the PELT algorithm was five, and the
 244 top five were Japanese bigrams for “Astra + Zeneca,” “reserve + possible,” “article + Reuters,”
 245 “venue + reserve,” and “medical-care + workers.” The bigrams “reserve + possible” and “venue
 246 + reserve” ranked in the top-2 from June to September, and the ranking of “Astra + Zeneca”
 247 decreased since May, as shown in Fig. 2.



248
249 Fig. 2. Translation of top ten bigrams of each month. The lengths of the bars represent the monthly term frequencies
250 in tweets of each month.
251

252 As for correlation analysis of unigrams, the time lags for “reserve” and “venue” were 0, and the
253 vaccination cases led the number of tweets containing “infection” for 5 days. After calculating r
254 between the daily number of tweets containing each top unigram and vaccination cases, significant
255 r ($P < 0.001$) was found for all unigrams except “mutation.” The largest r for the daily vaccination
256 cases was from unigrams “infection” ($r = 0.746$), “reserve” ($r = 0.829$), and “venue” ($r = 0.908$).
257 We then checked the daily number of tweets containing all the combinations of the three unigrams
258 showing a strong correlation and found the highest r ($r = 0.912$, $P < 0.001$) for tweets containing
259 both “reserve” and “venue.” By randomly selecting 5 days and checking the source of all the tweets
260 on those days, we found the percentage of tweets containing both “reserve” and “venue” posted
261 by official accounts or mainstream media was 96.0%–100% (95% confidence interval [CI]). The
262 trend of tweets containing both unigrams “reserve” and “venue” compared with the daily
263 vaccination cases is shown in Fig. 3.
264

265 As for bigrams, the bigram “venue + reserve” overlapped with the unigram analysis and was
266 excluded from this part. The time lags for bigrams “reserve + possible” and “article + Reuters”
267 were 0, and vaccination cases led “Astra + Zeneca” and “medical-care + workers” for 116 days
268 and 63 days respectively. The bigrams “reserve + possible” and “article + Reuters” got the highest
269 cross correlations than the rest. The bigrams “Astra + Zeneca” ($r = -0.331$), “reserve + possible”
270 ($r = 0.908$), and “article + Reuters” ($r = 0.229$) showed significant r ($P < 0.001$) except “medical-care
271 + workers” ($r = -0.055$). By three volunteers’ manually checking, we found 95.4% of the tweets
272 contains the bigrams “reserve + possible” were the same as those of the combination of unigrams
273 “venue” and “reserve”.
274

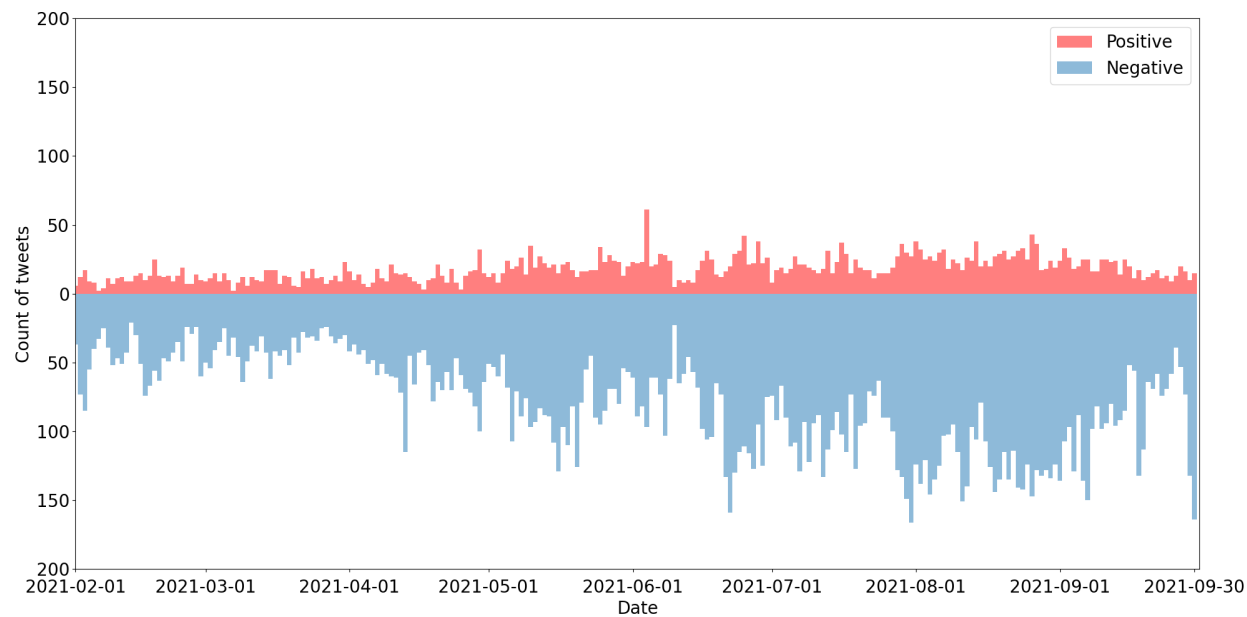


275
276 Fig. 3. Trend of tweets containing both unigrams “reserve” and “venue,” and the curve for daily first dose
277 vaccination cases
278

279 3.3 Sentiment analysis

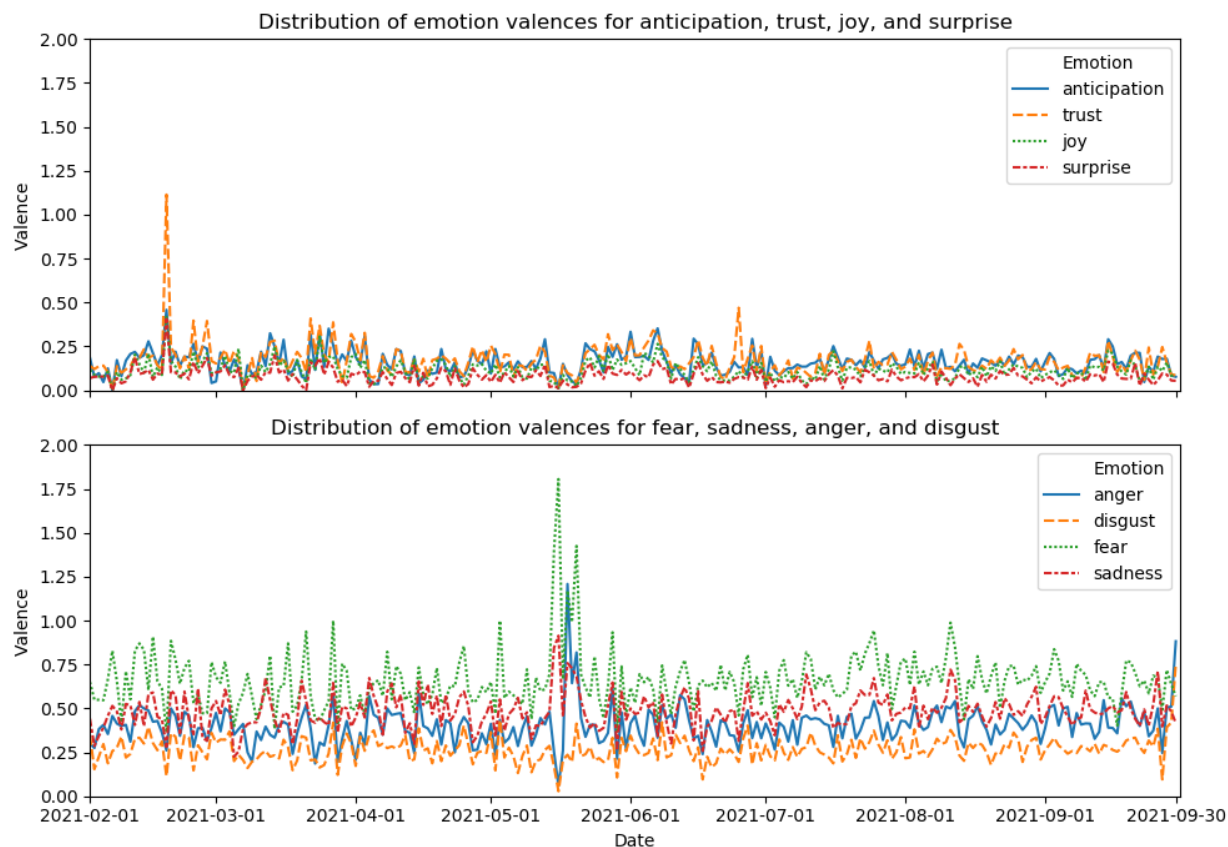
280 For all tweets, 4,453 (2.3%) were positive, 19,340 (10.1%) were negative, 164,687 (86.4%) were
281 neutral, and 2,217 (1.2%) were mixed positive and negative sentiments. A comparison between
282 the daily numbers of tweets marked as positive and negative is shown in Fig. 4. Negative
283 sentiments overwhelmed positive sentiments for all days.
284

285 The DOVs for the eight emotions are shown in Fig. 5. The daily average DOV of anger (0.404),
286 disgust (0.268), fear (0.659), sadness (0.486), overwhelmed anticipation (0.163), trust (0.173), joy
287 (0.118), and surprise (0.081). Fear was the dominant emotion during this period. Here, we defined
288 the peaks of emotion as larger than three times the daily average DOV for that emotion. Trust
289 peaked (1.114) on February 18, 2021. From May 13, 2021 to May 18, 2021, there were several
290 peaks of fear.
291
292



293
294
295

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

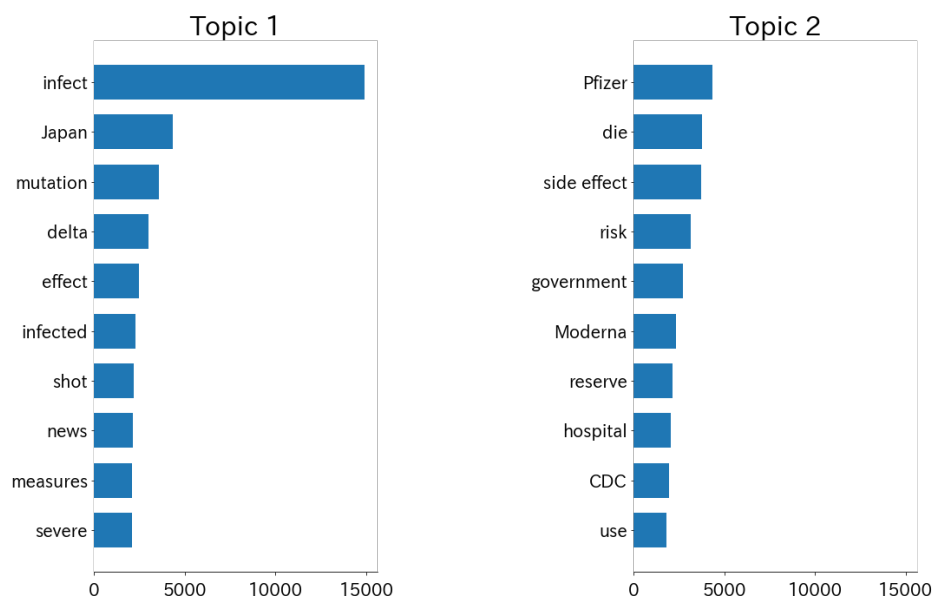


296
297
298

Fig. 5. Daily average DOV of eight emotions in the vaccine-related tweets

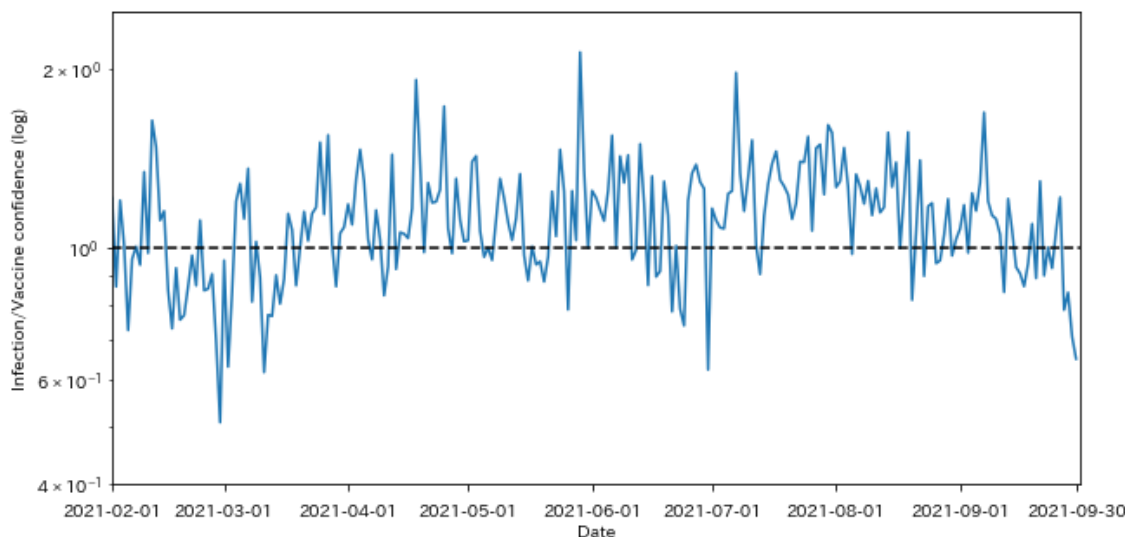
299 3.4 Topic modeling

300 The top ten keywords for each LDA topic are shown in Fig. 6. The theme of topic-1 is “infect,”
301 and topic-2 is “vaccine confidence.” It is also noticeable that the weight of “infect” (14,895) in
302 topic-1 was over three times the weight of the second keyword “Japan” (4,359), but the weight of
303 “Pfizer” (4,348) in topic-2 was only 15.5% larger than the second keyword “die” (3,763).
304



305 Fig. 6. Top ten keywords of two topics by LDA modeling. The bars represent the weights, which can be regarded as
306 the pseudo counts of the keywords in each topic.
307

308
309 The ratio between the expectation of the number of “infect”-related tweets and “vaccine
310 confidence” vaccine confidence-related tweets is shown in Fig. 7. The total expectation of the
311 number of tweets generated from topic-1 (“infect,” $n=30,288$) is larger than topic-2 (“vaccine
312 confidence,” $n=27,572$), and the mean ratio between the expectation of the daily number of tweets
313 generated from topics-1 and -2 is significantly larger than 1 ($P < 0.01$). On 68.2% of days, the
314 expectation of the number of tweets generated from “infect” was larger than “vaccine confidence.”
315



316
317 Fig. 7. Ratio between the expectation of the number of “infection”-related tweets and “vaccine
318 confidence”-related tweets
319

320 4. Discussion

321 A high vaccination rate is thought to be promoted by high vaccine confidence [16,36–39], but
322 Japan achieved a high vaccination rate in several months, with the lowest vaccine confidence in
323 the world. This retrospective study aimed to determine the reasons for the fast vaccination process
324 in Japan, which may be instructive for propelling worldwide vaccination towards infectious
325 diseases. Based on previous studies [16,37–40], we hypothesized that subjective factors, including
326 S1) increased vaccine confidence and S2) fear of infection, and objective factors, including O1)
327 adequate vaccine supply and O2) effective delivery of reservation-related vaccine information.
328 Our results indicated that hypotheses S2 and O2 might have driven the public to be vaccinated.
329 However, we did not find any evidence for hypothesis S2. Evidence for hypothesis O1 can be
330 found in the history of vaccine supply on the official website of the Prime Minister’s Office of
331 Japan (Prime Minister of Japan and his Cabinet) and is not discussed in this paper.

332
333 Several results support hypothesis S2. In Fig. 1 of the unigram token analysis, the keyword “infect”
334 ranked the top three, except in February, and ranked first from May to August, during Japan’s
335 fourth and fifth wave infections. The keywords “venue” and “reserve” also ranked up from May.
336 No keywords related to increased vaccine confidence were found. Fig. 4 in sentiment analysis
337 showed that negative sentiment overwhelmed positive, consistent with the results of Chen et al.
338 that Japan showed dissatisfaction compared with neighboring countries [38]. Combined with our
339 result that “infect” was the top-1 keyword and “side effect” ranked eighth in the unigram token
340 analysis, our results support that the Japanese public was more concerned about infection than the
341 side effects of COVID-19 vaccines.

342
343

344 More evidence for hypothesis S2 was obtained from the topic modeling results. From the keywords
345 of topic-1 (“infect”), we can see that the public was concerned about the infection and death rate.
346 The mutated virus and empowered cases also led to fear. Willis et al. found that less fear of
347 infection may lead to a lower willingness to be vaccinated [40], which is complementary to our
348 results. From the keywords of topic-2 (“vaccine confidence”), we can see that the side effects of
349 the vaccines were the most concerning, but the following keywords were about the effectiveness
350 of vaccines on the mutated virus, reservation of vaccines, and medical care conditions. Previous
351 surveys in different countries have indicated that fear of vaccine safety is the key factor for low
352 vaccine acceptance [41,42]. Also, the “side effect” weight in topic-2 was much less than “infect”
353 in topic-1. The top keywords in the two topics indicated that people were more concerned about
354 COVID-19 than the side effects of vaccines. Bendau et al. reported a significantly positive
355 correlation between fears of infection and vaccine acceptance and a significant negative correlation
356 between fear of vaccine safety and vaccine acceptance [43]. Therefore, it is important to distinguish
357 the mainstream fear emotion to determine the reason for the high vaccination rate. Fig. 7 provided
358 details about the ratio between the expected number of tweets related to “infect” and “vaccine
359 confidence.” In most cases, the ratio was larger than 1, indicating that the public was more
360 concerned about infection than the safety and effectiveness of vaccines. Higher ratios were
361 observed in April and from July to the end of September, periods of Japan’s fourth and fifth waves
362 of infection. There was also a relatively long period of less than one ratio from mid-February to
363 mid-March, which was the period when the vaccines were less effective against the mutated virus
364 (February 10), severe side effects of the AstraZeneca vaccine (March 12), and several side effects
365 in Japan (February 21, March 7, March 10). However, the ratio soon increased because of the
366 fourth wave infection. This example also proved that fear of infection overcame the vaccine safety
367 concern.

368
369 We also provide evidence of a strong relationship between vaccination and hypothesis O2. Bigram
370 analysis in Fig 2 showed that “reservation + possible” ranked first since June, shortly after the
371 large-scale vaccination started, which might reflect the strong concerns about the vaccine
372 reservation by the public. Unigram token analysis in Fig. 3 showed that tweets including the
373 keywords “reserve” and “venue” were significantly highly correlated ($r > 0.9$, $P < 0.01$) with the
374 daily number of vaccination cases in Japan, and most of them were from government official
375 accounts. The bigram “reservation + possible” also showed high correlation ($r > 0.9$, $P < 0.01$)
376 with the daily vaccination cases. Because reservation information should always lead to the actual
377 vaccination, this result indicated that in addition to sufficient vaccine supply, reservation
378 information delivery might also be important in large-scale vaccination. Also, the time lag for the
379 maximum cross correlation was 0, which may indicate the efficiency of the reservation information
380 posted on Twitter. Our results were consistent with Fu’s work that inflexible information systems
381 for vaccine reservation can impair immunization services in the community [44].

382
383 We did not find any evidence for hypothesis S1. Macaraan reported a shift from hesitancy to
384 confidence towards the COVID-19 vaccination program among Filipinos [39]. Okubo [45]
385 reported a shift from hesitancy to confidence, but they also admitted that the shift might come from
386 the differences in metrics of the survey with previous work [46]. Following these works, we looked
387 for a similar shift in sentiment or emotions from negative to positive, but negative sentiments
388 overwhelm positive sentiments in Fig. 4, and fear dominates all the emotions in Fig. 5. The positive

389 emotions, “anticipation,” “trust,” and “joy” did not increase during the entire period. These two
390 results made it difficult to conclude increased vaccine confidence.

391
392 Our results were partially related to the 5 C model (confidence, competence, convenience,
393 calculation, and collective responsibility) measuring vaccine hesitancy [39,47]. Confidence and
394 complacency are two subjective measures that are directly related to individuals. In our work, the
395 LDA theme “vaccine confidence” belonged to “confidence,” and “fear of infection” belonged to
396 “complacency.” In Japan, fear of infection may drive a high vaccination rate. The delivery of
397 reservation information may be an extension to “convenience,” which was previously defined as
398 “physical availability, affordability and willingness-to-pay, geographical accessibility, ability to
399 understand (language and health literacy), and appeal of immunization service affect uptake [48].”
400 Our work indicates that information about vaccination reservations should also be considered for
401 the convenience of vaccination.

402
403 We admit that our research might have some potential limitations: (1) the imbalance of the
404 demographics of Twitter users in Japan [49] may cause bias in the results; (2) status of the user on
405 a certain day (at home or not, other events on that day, etc.) may also bias the dataset [50]; (3) due
406 to the lack of a reliable public model for sentiment analysis in the Japanese language, the cloud
407 service AWS was used for sentiment analysis; (4) filtering keywords may include irrelevant or
408 missing related tweets; (5) anti-vaccine tweets, especially rumors, were not distinguished or
409 analyzed separately in this study. However, feature works can be combined with classical surveys
410 to train the sentiment analysis model and model to distinguish rumors from tweets to overcome
411 these limitations.

412
413 This retrospective study aimed to determine the reasons for the fast vaccination process in Japan,
414 which might be instructive for propelling worldwide vaccination towards infectious diseases. In
415 conclusion, our work indicated that awareness of the danger of COVID-19 increased the
416 willingness to be vaccinated; with a sufficient supply of vaccines, effective reservation information
417 delivery might provide more opportunities for people to be vaccinated. Models measuring vaccine
418 hesitancy might also need to add efficiency in delivering reservation information as a metric. Based
419 on our findings, we recommend public health policymakers and the government to share fair and
420 prompt information about the infectious diseases and vaccination. Also, efforts on tied cooperation
421 among multi-level relevant organizations and new media operation may lead to achieve smoother
422 delivery of vaccine-reservation information.

423 **Authors’ contribution**

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

427 **Funding**

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

429 **Conflicts of Interest**

430 None declared.

431 **Ethical Approval**

432 This study used publicly available and accessible tweets collected by Georgia State University's
433 Panacea Lab allowing free download. We assert that our analysis is compliant with Twitter's usage
434 policy in aggregate form without identifying specific individuals who made the Twitter posts. Also,
435 the numbers of infections, deaths, and vaccinated cases downloaded from the PMOJ are open
436 government data. Therefore, the activities described do not meet the requirements of human
437 subject research and did not require review by an institutional review board.

438 **Abbreviations**

439 COVID-19: Coronavirus disease 2019
440 IDs: identifications
441 PMOJ: Prime Minister's Office of Japan
442 amp: ampersands
443 NLP: natural language processing
444 PELT: pruned exact linear time
445 AWS: Amazon Web Service
446 DOV: degree of valence
447 LDA: Latent Dirichlet allocation

448 **References**

- 449 1. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of
450 COVID-19 in the U.S. *Soc Sci Med.* 2020;263: 113356.
451 doi:10.1016/j.socscimed.2020.113356
- 452 2. Experts warn of 6th wave as COVID cases decrease in Japan but rise overseas. In: *The*
453 *Mainichi* [Internet]. [cited 22 Feb 2022]. Available:
454 <https://mainichi.jp/english/articles/20211030/p2a/00m/0na/020000c>
- 455 3. Fifth wave had fewer big clusters possibly due to vaccinations. In: *The Asahi Shimbun*
456 [Internet]. [cited 22 Feb 2022]. Available: <https://www.asahi.com/ajw/articles/14461519>
- 457 4. Larson HJ. Vaccine confidence and public trust as drivers of vaccine failure. *Int J Infect Dis.*
458 2014;21: 50. doi:10.1016/j.ijid.2014.03.522
- 459 5. CDC. What is vaccine confidence? In: Centers for Disease Control and Prevention
460 [Internet]. 7 Feb 2022 [cited 22 Feb 2022]. Available: [https://www.cdc.gov/vaccines/covid-](https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html)
461 [19/vaccinate-with-confidence/building-trust.html](https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html)
- 462 6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey
463 of potential acceptance of a COVID-19 vaccine. *Nat Med.* 2021;27: 225–228.
464 doi:10.1038/s41591-020-1124-9

- 465 7. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in
466 vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective
467 temporal modelling study. *Lancet*. 2020;396: 898–908. doi:10.1016/S0140-6736(20)31558-
468 0
- 469 8. Mahase E. Covid-19: UK has highest vaccine confidence and Japan and South Korea the
470 lowest, survey finds. *BMJ*. 2021;373: n1439. doi:10.1136/bmj.n1439
- 471 9. Gordon A, Reich MR. The Puzzle of Vaccine Hesitancy in Japan. *J Jpn Stud*. 2021;47: 411–
472 436. doi:10.1353/jjs.2021.0047
- 473 10. Kunitoki K, Funato M, Mitsunami M, Kinoshita T, Reich MR. Access to HPV vaccination
474 in Japan: Increasing social trust to regain vaccine confidence. *Vaccine*. 2021.
475 doi:10.1016/j.vaccine.2021.08.085
- 476 11. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al.
477 Coronavirus Pandemic (COVID-19). *Our World in Data*. 2020 [cited 22 Jan 2022].
478 Available: <https://ourworldindata.org/covid-vaccinations>
- 479 12. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-
480 19 vaccine sentiments and opinions on Twitter. *Int J Infect Dis*. 2021;108: 256–262.
481 doi:10.1016/j.ijid.2021.05.059
- 482 13. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as
483 a Tool for Health Research: A Systematic Review. *Am J Public Health*. 2017;107: e1–e8.
484 doi:10.2105/AJPH.2016.303512
- 485 14. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: Topic
486 modeling and sentiment analysis. *J Med Internet Res*. 2021;23: e24435. doi:10.2196/24435
- 487 15. Huangfu L, Mo Y, Zhang P, Zeng D, He S. Analyzing COVID-19 vaccine tweets following
488 vaccine rollout: A sentiment-based topic modeling approach. *J Med Internet Res*. 2021.
489 doi:10.2196/31726
- 490 16. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People’s
491 Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll
492 Analysis. *J Med Internet Res*. 2021;23: e28973. doi:10.2196/28973
- 493 17. Social media stats Japan. In: *StatCounter Global Stats* [Internet]. [cited 22 Jan 2022].
494 Available: <https://gs.statcounter.com/social-media-stats/all/japan>
- 495 18. Twitter: most users by country. In: *Statista* [Internet]. [cited 22 Jan 2022]. Available:
496 [https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-](https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/)
497 [countries/](https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/)
- 498 19. Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T, et al. Public opinion and
499 sentiment before and at the beginning of COVID-19 vaccinations in Japan: Twitter analysis.
500 *JMIR Infodemiology* 2022;2(1):e32335 doi: 10.2196/32335
501

- 502 20. Kwok SWH, Vadde SK, Wang G. Tweet Topics and Sentiments Relating to COVID-19
503 Vaccination Among Australian Twitter Users: Machine Learning Analysis. *J Med Internet*
504 *Res.* 2021;23: e26953. doi:10.2196/26953
- 505 21. Yang X, Sornlertlamvanich V. Public Perception of COVID-19 Vaccine by Tweet
506 Sentiment Analysis. 2021 International Electronics Symposium (IES). 2021. pp. 151–155.
507 doi:10.1109/IES53407.2021.9594036
- 508 22. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, et al. Public perception of COVID-19
509 vaccines on Twitter in the United States. *medRxiv.* 2021. doi:10.1101/2021.10.16.21265097
- 510 23. Monselise M, Chang C-H, Ferreira G, Yang R, Yang CC. Topics and Sentiments of Public
511 Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. *J Med Internet*
512 *Res.* 2021;23: e30765. doi:10.2196/30765
- 513 24. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A
514 sentiment analysis. *Vaccine.* 2021. doi:10.1016/j.vaccine.2021.08.058
- 515 25. Shim J-G, Ryu K-H, Lee SH, Cho E-A, Lee YJ, Ahn JH. Text Mining Approaches to
516 Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in
517 Korea. *Int J Environ Res Public Health.* 2021;18. doi:10.3390/ijerph18126549
- 518 26. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19
519 Twitter chatter dataset for open scientific research -- an international collaboration. *ArXiv.*
520 2020. Available: <https://www.ncbi.nlm.nih.gov/pubmed/32550247>
- 521 27. Prime Minister of Japan and his cabinet. [cited 22 Feb 2022]. Available:
522 <https://japan.kantei.go.jp/>
- 523 28. Bird S, Klein E, Loper E. Natural language processing with python. O'Reilly Media; 2009.
524 Available: <https://books.google.at/books?id=KG1bfiiP1i4C>
- 525 29. Montani I, Honnibal M, Honnibal M, Van Landeghem S, Boyd A, Peters H, et al.
526 explosion/spaCy: v3.2.1: doc_cleaner component, new Matcher attributes, bug fixes and
527 more. 2021. doi:10.5281/zenodo.5764736
- 528 30. Wambui D. The power of the pruned exact linear time(PELT) test in multiple changepoint
529 detection. *Am j theor appl stat.* 2015;4. doi:10.11648/j.ajtas.20150406.30
- 530 31. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput*
531 *Intell.* 2013;29: 436–465. doi:10.1111/j.1467-8640.2012.00460.x
- 532 32. Nikita M. Tuning of the Latent Dirichlet Allocation Models Parameters [R package
533 ldatuning version 1.0.2]. 2020 [cited 15 Dec 2021]. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=ldatuning)
534 [project.org/package=ldatuning](https://CRAN.R-project.org/package=ldatuning)
- 535 33. Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN. On Finding the Natural
536 Number of Topics with Latent Dirichlet Allocation: Some Observations. *Advances in*
537 *Knowledge Discovery and Data Mining.* Springer Berlin Heidelberg; 2010. pp. 391–402.

- 538 doi:10.1007/978-3-642-13657-3_43
- 539 34. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model
540 selection. *Neurocomputing*. 2009;72: 1775–1781. doi:10.1016/j.neucom.2008.06.011
- 541 35. Griffiths TL, Steyvers M. Finding scientific topics. *Proc Natl Acad Sci U S A*. 2004;101
542 Suppl 1: 5228–5235. doi:10.1073/pnas.0307752101
- 543 36. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc
544 information retrieval. *Doc numér*. 2014;17: 61–84. doi:10.3166/dn.17.1.61-84
- 545 37. Mori H, Naito T. A rapid increase in the COVID-19 vaccination rate during the Olympic
546 and Paralympic Games 2021 in Japan. *Hum Vaccin Immunother*. 2021; 1–2.
547 doi:10.1080/21645515.2021.2010440
- 548 38. Chen CWS, Lee S, Dong MC, Taniguchi M. What factors drive the satisfaction of citizens
549 with governments’ responses to COVID-19? *Int J Infect Dis*. 2021;102: 327–331.
550 doi:10.1016/j.ijid.2020.10.050
- 551 39. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence:
552 Development of a measure assessing the 5C psychological antecedents of vaccination. *PLoS*
553 *One*. 2018;13: e0208601. doi:10.1371/journal.pone.0208601
- 554 40. Willis DE, Andersen JA, Bryant-Moore K, Selig JP, Long CR, Felix HC, et al. COVID-19
555 vaccine hesitancy: Race/ethnicity, trust, and fear. *Clin Transl Sci*. 2021;14: 2200–2207.
556 doi:10.1111/cts.13077
- 557 41. Pugliese-Garcia M, Heyerdahl LW, Mwamba C, Nkwemu S, Chilengi R, Demolis R, et al.
558 Factors influencing vaccine acceptance and hesitancy in three informal settlements in
559 Lusaka, Zambia. *Vaccine*. 2018;36: 5617–5624. doi:10.1016/j.vaccine.2018.07.042
- 560 42. Robertson E, Reeve KS, Niedzwiedz CL, Moore J, Blake M, Green M, et al. Predictors of
561 COVID-19 vaccine hesitancy in the UK household longitudinal study. *Brain Behav Immun*.
562 2021;94: 41–50. doi:10.1016/j.bbi.2021.03.008
- 563 43. Bendau A, Plag J, Petzold MB, Ströhle A. COVID-19 vaccine hesitancy and related fears
564 and anxiety. *Int Immunopharmacol*. 2021;97: 107724. doi:10.1016/j.intimp.2021.107724
- 565 44. Fu C. Milestone and challenges: lessons from defective vaccine incidents in China. *Hum*
566 *Vaccin Immunother*. 2020;16: 80. doi:10.1080/21645515.2019.1646579
- 567 45. Okubo R, Yoshioka T, Ohfuji S, Matsuo T, Tabuchi T. COVID-19 Vaccine Hesitancy and
568 Its Associated Factors in Japan. *Vaccines (Basel)*. 2021. doi:10.3390/vaccines9060662
- 569 46. Machida M, Nakamura I, Kojima T, Saito R, Nakaya T, Hanibuchi T, et al. Acceptance of a
570 COVID-19 Vaccine in Japan during the COVID-19 Pandemic. *Vaccines*. 2021;9: 210.
571 doi:10.3390/vaccines9030210
- 572 47. Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batouré O, Anya B-PM, et al. Vaccine

- 573 hesitancy in the era of COVID-19: could lessons from the past help in divining the future?
574 Hum Vaccin Immunother. 2021; 1–3. doi:10.1080/21645515.2021.1893062
- 575 48. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy:
576 Definition, scope and determinants. Vaccine. 2015;33: 4161–4164.
577 doi:10.1016/j.vaccine.2015.04.036
- 578 49. Social Media & User-Generated Content. In: Statista [Internet]. [cited 30 Apr 2022].
579 Available: [https://www.statista.com/markets/424/topic/540/social-media-user-generated-](https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/)
580 [content/](https://www.statista.com/markets/424/topic/540/social-media-user-generated-content/)
- 581 50. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal
582 investigation of tourist attraction visit sentiment on Twitter. PLoS One. 2018;13: e0198857.
583 doi:10.1371/journal.pone.0198857

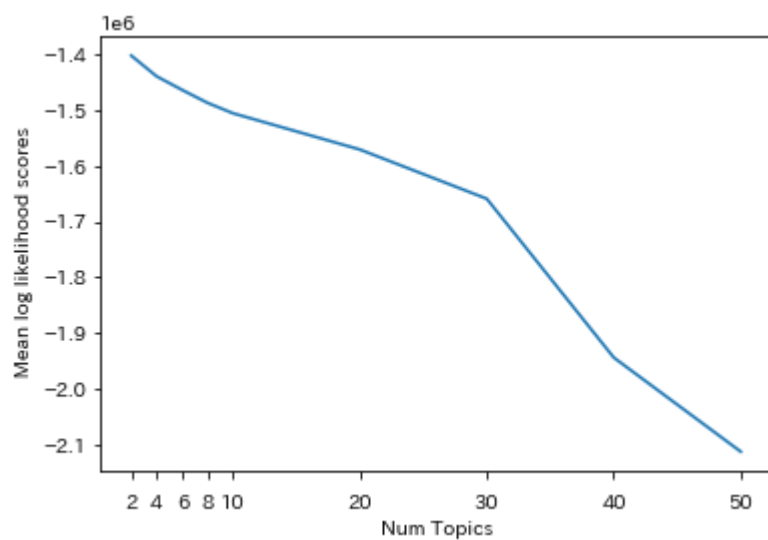


Fig. 1. Mean Log likelihood scores for different LDA topic numbers using five-fold cross validation.