

Public Perception of COVID-19 Vaccines through Analysis of Twitter Content and Users

Authors:

Sameh N. Saleh, MD*^{1,2}, Samuel A. McDonald, MD^{2,3}, Mujeeb A. Basit, MD, MMSc^{1,2},
Sanat Kumar^{2,4}, Reuben J. Arasaratnam, MD, MPH¹, Trish M. Perl, MD¹, Christoph U.
Lehmann, MD^{2,5}, Richard J. Medford, MD^{1,2}

¹ Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323
Harry Hines Boulevard, Dallas, TX 75390

² Clinical Informatics Center, University of Texas Southwestern Medical Center, 5323 Harry
Hines Boulevard, Dallas, TX 75390

³ Department of Emergency Medicine, University of Texas Southwestern Medical Center, 5323
Harry Hines Boulevard, Dallas, TX 75390

⁴ Lebanon Trail High School, 5151 Ohio Dr, Frisco, TX 75035

⁵ Departments of Pediatrics, Bioinformatics, Population & Data Sciences, University of Texas
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390

Corresponding Author:

Sameh N. Saleh, MD
5323 Harry Hines Blvd, Dallas, TX 75219
M: 571-338-3680
F: 214-648-9478
E: sameh.n.saleh@gmail.com

Keywords:

COVID-19, Vaccine, Twitter, Social Media, Public Opinion, COVID-19 Vaccines, SARS-CoV-
2, Vaccination, Vaccination Refusal, Vaccine Hesitancy, Natural Language Processing,
Sentiment Analysis, Topic Modeling, Demographic Inference

35 Abstract

36 Twitter is a robust medium to understand wide-scale, organic public perception about the
37 COVID-19 vaccine. In this cross-sectional observational study, we evaluated 2.4 million English
38 tweets from nearly 1 million user accounts matching keywords (('covid*' OR 'coronavirus')
39 AND 'vaccine') during vaccine development from Feb 1st through Dec 11th, 2020. We applied
40 topic modeling, sentiment and emotion analysis, and demographic inference of users on the
41 COVID-19 vaccine related tweets to provide insight into the evolution of public attitudes.
42 Individuals generated 87.9% (n=834,224) of tweets. Of individuals, men (n=560,824)
43 outnumbered women (n=273,400) by 2:1 and 39.5% (n=329,776) of individuals were ≥ 40 years
44 old. Daily mean sentiment fluctuated congruent with news events, but overall trended positively.
45 Trust, anticipation, and fear were the three most predominant emotions; while fear was the most
46 predominant emotion early in the study period, trust outpaced fear from April 2020 onward. Fear
47 was more prevalent in tweets by individuals (26.3% vs. organizations 19.4%; $p < 0.001$),
48 specifically among women (28.4% vs. males 25.4%; $p < 0.001$). Multiple topics had a monthly
49 trend towards more positive sentiment. Tweets comparing COVID-19 to the influenza vaccine
50 had strongly negative early sentiment but improved over time. Our findings are concerning for
51 COVID-19 vaccine hesitancy, but also identify targets for educational interventions.

52

53 Introduction

54 With the global continuation of the COVID-19 pandemic, the large-scale administration of a
55 SARS-CoV-2 vaccine (referred from here on as the COVID-19 vaccine) is crucial to achieve
56 herd immunity and curtail further spread of the virus¹. As governments work to approve and
57 distribute safe and effective vaccines,² important questions regarding vaccination willingness
58 persist: What are the attitudes and perceptions of the public³ to these vaccines and how can they
59 affect vaccine uptake⁴? These questions are important to develop an education and outreach
60 approach to achieve the desired vaccine penetration to achieve herd immunity⁵. In 2019, prior to
61 the COVID-19 pandemic, the World Health Organization (WHO) had identified vaccine
62 hesitancy as one of the top 10 greatest global health threats⁶. While surveys on attitudes and
63 perception of a COVID-19 vaccine show significant vaccine hesitancy among the general
64 population⁷⁻⁹ and health care providers^{10,11}, studies remain small in size, tend to focus on local
65 participants, are prone to sampling error from non-probability sampling and reporting bias, and
66 perhaps most poignantly, cannot capture real-time changes in vaccine willingness. Crowdfunding
67 platforms may provide an indication of emerging community needs related to COVID-19 but fail
68 to provide a continuous assessment of community sentiment¹².

69
70 Twitter, the microblogging platform, with over 187 million daily monetizable active users,¹³
71 serves as a robust medium to better understand wide-scale, organic public perception about the
72 COVID-19 vaccine. With nearly 400 million mentions, #COVID19 was the most used hashtag
73 on Twitter in 2020¹⁴. Social media has become increasingly recognized for its rapid information
74 dissemination (whether accurate or not) and dispersion of sentiment that quickly crosses
75 geographic and social boundaries¹⁵. Analysis of social media text can inform real-time changes

76 and evolution in population-level attitudes¹⁶. As evident with the rise of the “infodemic” during
77 the COVID-19 pandemic, Twitter has become a particularly useful data source in public health
78 and healthcare-related research¹⁷ and has been repeatedly used to study public sentiment and
79 understand trends throughout the COVID-19 pandemic¹⁸⁻²¹. Earlier in the COVID-19 pandemic,
80 we were able to demonstrate initial public sentiment regarding the virus, its origin and spread,
81 and measures to limit its spread²² as well as early support for social distancing²³ on Twitter.

82
83 Social media, and specifically Twitter, has been shown to be a major factor in vaccine uptake
84 and should be monitored and potentially used for interventions to address vaccine hesitancy²⁴.
85 Examining sentiments towards the influenza A H1N1 vaccine in 2009 showed that projected
86 vaccination rates based on Twitter sentiment were similar to vaccination rates estimated by
87 traditional phone surveys used by the Centers for Disease Control and Prevention (CDC)²⁵. A
88 previous study noted information exposure on Twitter may account for differences in human
89 papillomavirus (HPV) vaccine uptake that are not accounted for by socioeconomic factors like
90 education, insurance, or poverty²⁶. Another study noted that there is a significant relationship
91 between social media use by the public and organized action and public doubts of vaccine
92 safety²⁷.

93
94 We aimed to apply content and sentiment analysis on COVID-19 vaccine related tweets as well
95 as analysis of the responsible, originating user accounts to provide insight into the evolution of
96 public attitudes about the COVID-19 vaccines over time. We hypothesized that content analysis
97 from the start of the pandemic will identify important themes of discussion (especially those with
98 negative sentiment or evidence of misinformation) throughout the vaccine development process

99 that would inform health care officials, public health agencies, and policy makers and could be
100 used to aid in the outreach and educational interventions for the COVID-19 vaccine to the
101 general public.

102 Methods

103 Data Source

104 We performed a cross-sectional observational study of English-language tweets obtained by
105 matching the keywords (('covid*' OR 'coronavirus') AND 'vaccine') from February 1, 2020 to
106 December 11th, 2020. December 11th was chosen as an end date to mark the United States Food
107 and Drug Administration's first emergency use authorization of a COVID-19 vaccine²⁸. We used
108 the *snsrape* library²⁹ to obtain ("scrape") tweets identified through Twitter's advanced search
109 tool, which returns a relevant sample of tweets. We manually reviewed a random subsample of
110 1,000 tweets and verified the tweets' relevance to the topic of COVID-19 vaccination. We
111 extracted 21 and 20 variables related to the tweets and to the posting user accounts, respectively
112 (Supplemental Table S1 and S2).

113 Data Processing

114 We measured total daily tweets and completed descriptive statistics for collected variables. We
115 applied natural language processing techniques to process, analyze, and visualize the text from
116 tweets. To preprocess the tweet text for analysis ("cleaning"), we removed hyperlinks, user tags,
117 and words of little analytical value. We also returned words to their root form and segmented
118 text into one- and two-word terms. Further details are discussed in Supplemental Appendix A.
119 We visualized the top 300 processed terms as a word cloud with larger font size representing
120 greater term frequency. All analyses were conducted using Python, version 3.8.2 (Python

121 Software foundation). Institutional review board approval was not required because this study
122 used only publicly available data.

123 Sentiment Analysis

124 Sentiment analysis describes the affect of a piece of text — the intrinsic attractiveness or
125 aversiveness of a subject such as events, objects, or situations³⁰. We used the Valence Aware
126 Dictionary and sEntiment Reasoner (VADER)³¹ to analyze the sentiment polarity of a tweet.
127 VADER, a lexicon and rule-based tool, was particularly designed for sentiment analysis for
128 social media text. In addition to regular words, VADER leverages punctuation, emoticons,
129 emojis, sentiment-laden slang words and acronyms, as well as syntax and capitalization schemas
130 to inform labeling of a positive, neutral, and negative score for each document. These three
131 scores were combined to form a normalized, weighted composite score. Overall positive (≥ 0.05),
132 neutral (-0.05 to 0.05), and negative (≤ -0.05) sentiments are defined at standardized composite
133 score thresholds. When sentiment has been aggregated, we refer to an average sentiment of
134 ≥ 0.05 as positive and ≤ -0.05 as negative. Trends in sentiment over time were determined using
135 the Mann-Kendall trend test. We used the *TextBlob* library³² to label each tweet from a range of
136 0 (objective) to 1 (subjective) where objective tweets relay factual information and subjective
137 tweets typically communicate an opinion or belief. Finally, we used the *NRCLex* library to label
138 words within each tweet with corresponding emotional affects (i.e., Plutchik’s wheel of emotions
139 which include anger, anticipation, fear, disgust, joy, sadness, surprise, and trust) based on the
140 National Research Council Canada (NRC) affect lexicon³³. Based on these labels, we identified
141 tweets with their primary emotion and visualized how the proportion of eligible tweets (i.e.,
142 those with an identified primary emotion) with a particular primary emotion changed over time.

143 Topic Modeling

144 After cleaning the tweets to distill analyzable text as described in the methods, we applied a
145 machine learning algorithm called Correlation Explanation (CorEx)³⁴ to identify clusters of
146 topics for all tweets. CorEx identifies the most informative topics based on a set of latent factors
147 that best explain the correlations in the data in turn maximizing the total correlation or the
148 multivariate mutual information³⁵. Each document (in our case, each tweet) may include multiple
149 topics. We iterated through a range from 2 to 20 topics and trained a separate model for each
150 number of topics with the goal of identifying the model with the maximum total correlation,
151 ultimately choosing 15 topics for the topic model. We presented the top 20 words for each topic
152 cluster to author CUL without prior access to individual tweets from the dataset to manually
153 label a theme for each topic. The manually labeled topic labels were reviewed by two other
154 authors SNS and RJM with unanimous agreement. We visualized the monthly distribution of
155 topics over time and utilized a heat map to visualize how the mean sentiment of each topic has
156 changed per month.

157 User Exploration and Demographic Inference

158 Given that each tweet has one authoring account, we identified all unique user accounts in our
159 dataset and provided descriptive statistics with metadata available for the users, including the
160 launch date of the account, followers (accounts following them), follows (accounts they follow),
161 lifetime posts, likes, and media shared, as well as profile pictures, description information, and
162 verified status (badge to indicate an account of public interest that has been verified to be
163 authentic). To better understand demographic differences, we applied a previously validated
164 deep learning system through the *m3inference* library³⁶ to infer the account user as an individual
165 or an organization and if labeled as an individual, their gender (female or male) and age group

166 (≤ 18 years old, 19-29 years old, 30-39 years old, and ≥ 40 years old) based on multimodal input
167 that includes username, display name, description, and profile picture image. Each label using
168 the algorithm has an accompanying probability. The automatic demographic detection was
169 particularly designed for Twitter profiles for health-related cohort studies³⁷. We provided
170 summary statistics for the demographics identified and stratified sentiment and subjectivity
171 analyses by the different demographic groups to evaluate for differences. We used Mann-
172 Whitney U and χ^2 where appropriate to determine significance. Alpha level of significance was
173 set a priori at 0.05 and all hypothesis testing was two-sided. We did not adjust for multiple
174 comparisons as this was an exploratory study and should be interpreted as hypothesis-generating.

175 Results

176 A total of 2,356,285 tweets were extracted for the study period, of which 2,287,344 tweets were
177 English-only and included for evaluation. The tweets were generated by 948,666 accounts which
178 had been active for an average of 6.9 years (interquartile range [IQR], 2.6 - 10.0) with a median
179 of 267 (IQR, 55 - 1,100) followers and 3,600 (519 - 15,572) lifetime likes. Only 2.9%
180 (n=27,443) of accounts were verified (Table 1). Of the tweets analyzed, 54% (n=1,235,575) had
181 a link, 40.1% (n=916,585) mentioned other twitter accounts, 18.1% (n=414,173) used hashtags,
182 and 11.9% (n=273,278) contained media like an image or video. In terms of engagement, 41.3%
183 (n=943,639), 24.0% (n=548,863), and 20.7% (n=473,204) of tweets received likes, replies, and
184 retweets, respectively (Table 1). Individuals (vs. organizations) generated 87.9% (n=834,224) of
185 tweets. Of individuals, men (n=560,824) outnumbered women (n=273,400) by 2:1 and 39.5% (n
186 = 329,776) of individuals were ≥ 40 years old (Table 1).

187

188 Daily tweets abruptly spiked to 51,176 tweets on November 9th, the day Pfizer and BioNTech
189 announced their vaccine's effectiveness³⁸ (up from 4,052 tweets on November 8th) and peaked
190 on December 8th with 55,779 tweets. Tweets from November 1st to the end of the study period on
191 December 11th accounted for 39.8% (n = 910,593) of all tweets (Figure S1). The corpus of
192 tweets contained over 62.5 million words and 416 million characters. The ten most commonly
193 tweeted terms and their frequencies were as follows: "people" (228,482), "trial" (206,310),
194 "take" (181,598), "flu" (159,043), "trump" (149,042), "first" (147,103), "make" (142,242), "test"
195 (131,719), "need" (126,846), and "one" (122,966). Figure 1 displays a word cloud of the top 300
196 words with larger font size concordant with frequency.

197

198 Daily mean sentiment of tweets fluctuated congruent with news events, but overall trended
199 positively throughout the study period (Mann-Kendall statistic=10,122; tau=0.218; p<0.001)
200 (Figure 2a). Several days in early to mid-March and on October 13th saw particularly negative
201 sentiments, coinciding with news of the declaration of a pandemic by the WHO and Johnson &
202 Johnson's halting of their vaccine trial on October 12th³⁹, respectively. Highest daily mean
203 positive sentiment revolved around Moderna's July 14th announcement of a safe vaccine with
204 "robust immune response" in an early trial⁴⁰ and Pfizer's November 9th announcement of over
205 90% effectiveness of its vaccine³⁸. Twitter accounts representing organizations had more positive
206 sentiments than tweets from individuals (median weekly difference, 0.118; IQR, 0.091 to 0.144),
207 but there was no significant difference in polarity for age (median weekly difference, 0.006;
208 IQR, -0.011 to 0.019) and only minimal positive difference for males (median weekly difference,
209 0.030; IQR, 0.012 to 0.044) (Figure 2b-d).

210

211 The sentiment trends were reflected by the primary emotions identified in the COVID-19
212 vaccine tweets by month (Figure 3a). Fear started as the most prevalent primary emotion in
213 nearly 40% of eligible tweets early on but decreased to under 20% by the end of the study
214 period. Conversely, trust increased from below 20% to around 40% and outpaced fear in April
215 2020, maintaining as the most prevalent primary emotion thereafter. Anticipation was the second
216 most prevalent primary emotion for most of the study period, steadily ranging from 25% to 30%.
217 All other emotions were consistently expressed as the predominant emotion in less than 10% of
218 eligible tweets. Individuals had an increased predominance of fear (26.3% vs. 19.4%; $p < 0.001$)
219 and decreased predominance of anticipation (25.9% vs. 33.6%; $p < 0.001$) and trust (32.5% vs.
220 35.2%; $p < 0.001$). For individual accounts, women had more fear (28.4% vs. males 25.4%;
221 $p < 0.001$) with less anticipation (23.8% vs. 26.8%; $p < 0.001$) than men, but no significant
222 difference in trust (32.3% vs. 32.5%, $p = 0.11$). Those less than 40 years old had more fear (26.6%
223 vs. 26.0%; $p < 0.001$) and less trust (32.0% vs. 33.0%; $p < 0.001$) (Figure 3b-d). Tweets throughout
224 the year tended to be more objective (where 0 is fully objective and 1 as fully subjective) with
225 limited daily variation (overall mean 0.359; std 0.028) (Figure S2).

226
227 Table 2 shows each topic label with their key words and sample tweets. Figure S3 shows the 15
228 topics obtained from topic modeling with the proportion of tweets per month that contained each
229 topic. The dominant topic (topic 15) focused on mask use and public reactions. Discussions
230 about misinformation and conspiracy theories comprised the next most common topic, peaking
231 in May and staying relatively consistent from July through December. Tweets related to the
232 Indian and Russian governments' decision on producing and using the Sputnik V vaccine (topic
233 2) spiked in August. Discussion of Emergency Use Authorizations (EUA) and vaccine approvals

234 (topic 12) did not spike until November 2020 with the approval of the Pfizer and Moderna
235 vaccines. Several topics had strong mean positive sentiments throughout the study period,
236 including discussions of biotechnology companies and the stock market (topic 3), vaccination
237 firsts (topic 4), vaccine development (topic 6), and EUAs (topic 12). Other topics showed a
238 progressive trend from positive to negative throughout the study period including discussion of
239 US politics and the election (topic 1), the FDA and CDC (topic 14), and mask use and public
240 reactions (topic 15). Tweets comparing COVID-19 to influenza (topic 5) and its vaccine had
241 strongly negative early sentiment but improved over time (Figure 4). Compared to the rest of
242 individual users (n=810,318), those exhibiting negative sentiment posting about topic 5
243 (n=51,686) were proportionally more likely to be ≥ 40 years old (45.1% vs. 39.6%; $p < 0.001$) and
244 female (34.0% vs. 32.7%; $p < 0.001$). The only other topic with persistently negative sentiment
245 was discussion of misinformation and conspiracy theories (topic 13). Those exhibiting negative
246 sentiment posting about topic 13 (n=166,819) were more likely than other user accounts
247 (n=741,388) to be individuals (90.9% vs. 87.3%; $p < 0.001$) and of those individual accounts,
248 more likely to be female (34.4% vs. 32.4%; $p < 0.001$).

249 Discussion

250 Twitter is a rich medium that can serve as both thermometer and thermostat for the COVID-19
251 vaccine, which is a crucial public health strategy to combat the pandemic. It can provide insight
252 into public perception of a COVID-19 vaccine, but can also be used to understand and combat
253 knowledge deficits and vaccine hesitancy through information and education^{41,42}. A majority
254 (59%) of US Twitter users regularly obtain news on Twitter, proportionally more than any other
255 social media platform⁴³. We analyzed nearly 2.4 million COVID-19 vaccine related tweets in
256 2020, creating a dataset that exceeded the scope of related studies^{44,45}. Generally, we believe that

257 Twitter users favored the vaccine during its development phase. Tweets with positive sentiment
258 were more prominent than tweets with negative sentiment and trust emerged as the predominant
259 emotion. However, there were periods of time (usually linked to events in the public news cycle),
260 demographic subgroups, and topic clusters that had more prominent negative sentiment and
261 emotion.

262
263 Organizational accounts were significantly more positive, exhibiting more anticipation and trust
264 and less fear. For individuals, the gender and age distribution in our dataset parallels the reported
265 proportional share of Twitter's global advertising audience⁴⁶. Women expressed more fear and
266 less anticipation, but by the end of the study period, that gap had narrowed. Those less than 40
267 years old tended to express less trust and more fear, but the margin was small.

268
269 The topic most strongly associated with negative, albeit improving, sentiment was the discussion
270 of the influenza vaccine in combination with the COVID-19 vaccine. These tweets often
271 compared deaths and illness from both diseases or expressed general vaccine mistrust to both
272 vaccines. Examples include: “@[user] *Only time I've ever had the flu is the 2 times I got flu*
273 *shots. It was not a minor case either it was the full blown flu. I refuse to get another flu shot and*
274 *I also will refuse the covid vaccine*” and “*Flu Virus equals Flu Vaccine. Coronavirus Equals*
275 *Covid-19 Vaccine...Now if the Flu shot gives you the flu, the Covid-19 Shot will give you*
276 *Coronavirus....am I in the general area of Right??*”. Notably, these users exhibiting negative
277 sentiment about this topic were more likely to be ≥ 40 years old and female. This focused topic-
278 demographic cluster, for example, exposes a direct opportunity for intervention to correct

279 misinformation and mitigate vaccine hesitancy. Conversely, the emergency use authorizations of
280 the vaccine and reports of the first vaccine recipients, which arose later in the study period, were
281 celebrated with positive sentiment and mirror the overall increasing trend in positive sentiment
282 and trust.

283
284 While the percent population immunity needed to achieve herd immunity (either through innate
285 or acquired immunity) for COVID-19 is not yet known, estimates have increased from 60-70%
286 to possibly closer to 75-85%^{47,48}. Achieving herd immunity through infection would come at an
287 untenable cost⁴⁹, making the immunization effort critical to protect lives. Therefore, it was
288 concerning to us that fear was a common and persistent predominant emotion in COVID-19
289 vaccine tweets. While the proportion of ‘trust’ tweets outpaced ‘fear’ tweets relatively early in
290 the study period, approximately 20% of eligible tweets still expressed fear in association with the
291 vaccine. If this fear translates into refusal to become immunized, we are not only likely to see a
292 prolonged pandemic, but also further increases in COVID-19 related deaths as concerning virus
293 variants take hold. As more people receive the vaccine in the future, we anticipate that
294 sentiments will become more positive over time with increased trust and vaccine uptake, but this
295 will need to be consistently studied especially in the context of newly approved vaccines and
296 news events.

297 Limitations

298 Our study was limited by several factors. First, we recognize that our dataset is not all inclusive
299 of tweets discussing the COVID-19 vaccine. Our tweet search criterion was narrow to ensure
300 accuracy of captured tweets for this initial work and did not include terms such as “shot(s)”,
301 “immunization” and “inoculation.” Moreover, despite the volume of tweets analyzed, we are

302 limited to only a relevant sample of all tweets per Twitter’s advanced search tool. Second, we
303 used existing tools to analyze sentiments and emotion of tweets that are not specific to health
304 care topics, which could have skewed our analysis. Third, tweets related to COVID-19
305 vaccination could have been flagged or removed by Twitter for containing misinformation, but
306 we were not privy to that context to determine how that could have affected our sample. Finally,
307 since we targeted only tweets in English and are unable to determine geographic location for
308 users, we are limited in making conclusions about specific countries or countries where English
309 is the not the predominant language.

310 Conclusions

311 Leveraging 2.4 million COVID-19 vaccine related tweets in 2020, we were able to successfully
312 explore sentiment, emotion, topics, and user demographics to elucidate important trends in
313 public perception about the COVID-19 vaccine. Tweets were overall positive in sentiment and
314 with growing trust. However, fear maintained as a dominant emotion raising concern regarding
315 the willingness to receive the COVID-19 vaccine and subsets of negative sentiment emerged.
316 Comparison to influenza and the influenza vaccine as well as discussion about conspiracy
317 theories were important topics with negative sentiment and showed some demographic
318 differences that could allow for informed intervention. Future work will leverage these natural
319 language processing tools to engage in targeted messaging based on user interests and emotions.

320
321
322
323
324
325
326
327
328

329
330
331
332
333
334

335
336
337
338
339
340
341
342
343
344
345
346
347
348
349

Declarations

Ethics approval and consent to participate: The University of Texas Southwestern Human Research Protection Program Policies, Procedures, and Guidance did not require institutional review board approval as all data were publicly available.

Data Availability, Materials, & Correspondence: The data that support the findings of this study are available upon request. Please request from the corresponding author.

Code Availability: The code that support the findings of this study is available upon request. Please request from the corresponding author.

Competing interests: Dr. Lehmann reports stock ownership in Celanese Corporation and Colfax Corporation. There are no other competing interests.

Funding: None.

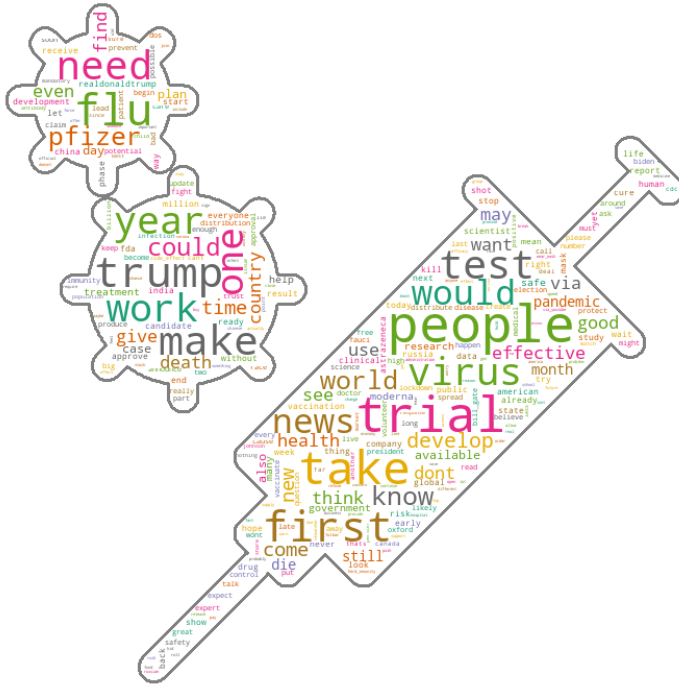
Authors' contributions: Study concept and design: SNS, CUL, RJM; Data acquisition: SNS, SK; Analysis: SNS, SK, RJM; Interpretation of data: SNS, SK, CUL, RJM; Manuscript preparation: all authors. All authors read and approved the final manuscript.

Acknowledgements: Not applicable

350 Figures and Tables

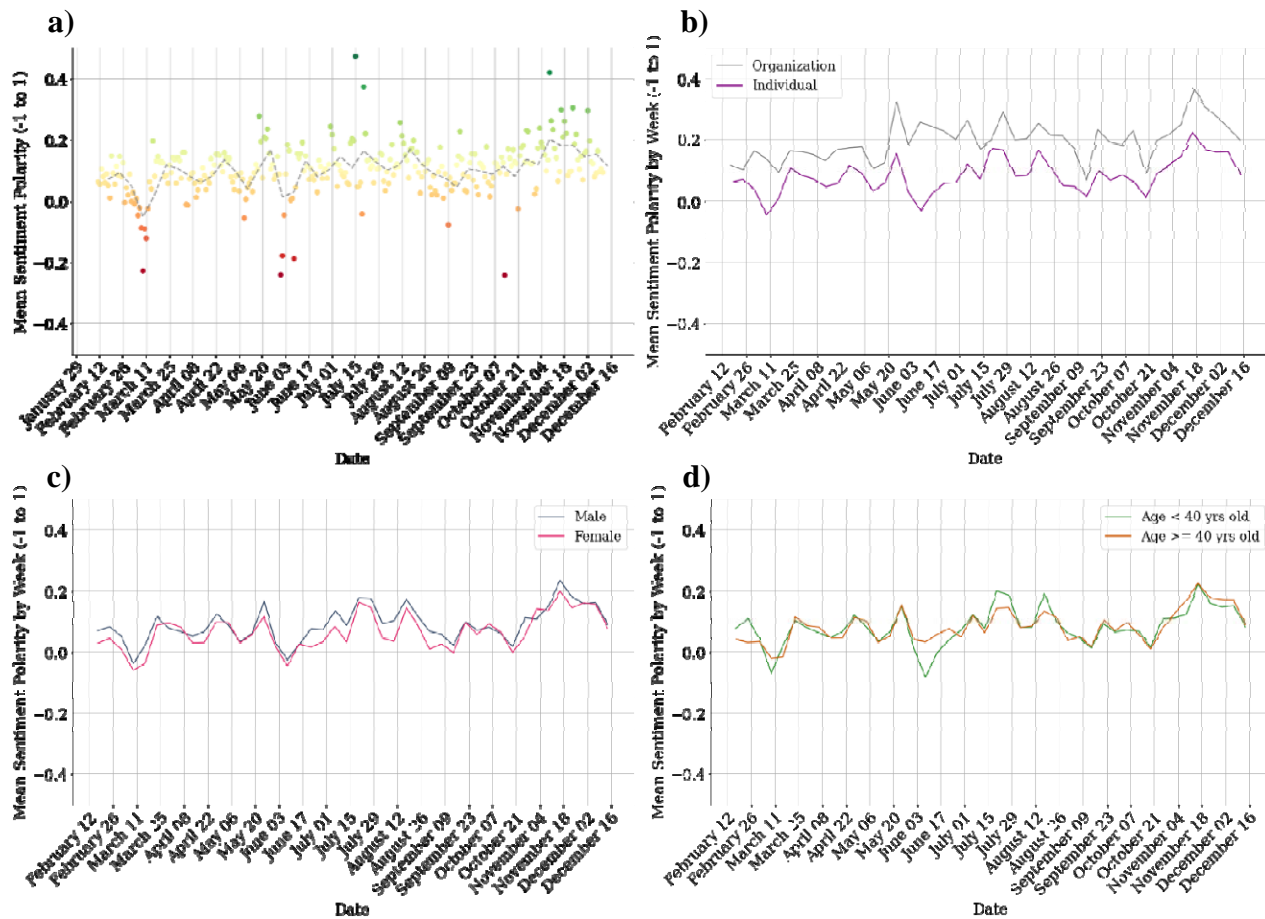
351 **Figure 1.** Word cloud of top 300 words related to COVID-19 and vaccine. Larger fonts
352 represent higher frequency in the corpus after preprocessing text.

353



354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374

375 **Figure 2a-d.** **a)** Mean sentiment polarity shown by day (as points) and by week (as a dashed
376 line). Each tweet was labeled as primarily negative (-1), neutral (0), or positive (1). **b)** Mean
377 weekly polarity stratified by individual versus organization. **c)** Mean weekly polarity stratified
378 by gender for individual accounts. **d)** Mean weekly polarity stratified by age more or less than 40
379 years than for individual accounts.



380

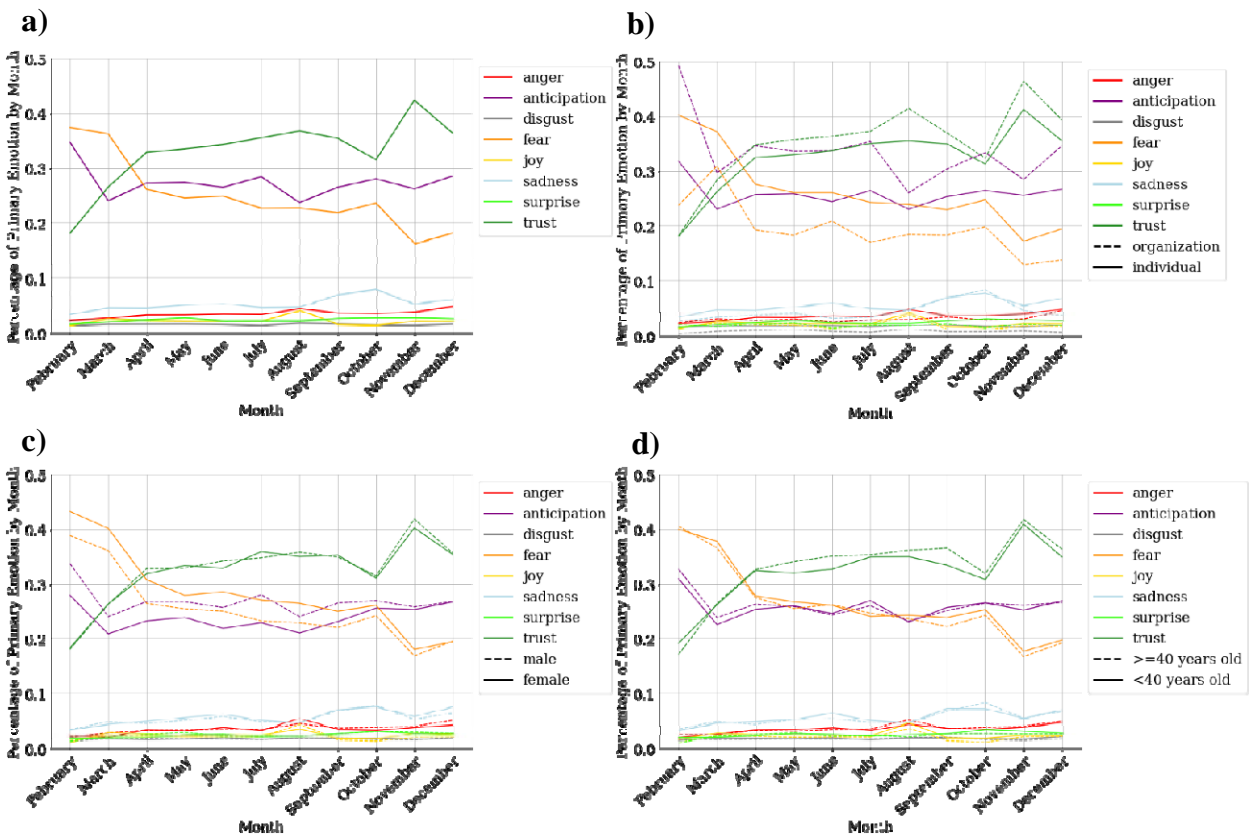
381

382

383

384

385 **Figure 3a-d.** Percent of tweets with primary emotion per month **a)** overall, **b)** stratified by
386 individual versus organization, **c)** stratified by gender for individual accounts, and **d)** stratified
387 by age more or less than 40 years than for individual accounts. Only tweets with a predominant



388 primary emotion (n = 1,489,027) are included.

389

390

391

392

393

394

395 **Figure 4.** Heat map showing mean sentiment by month for each topic. Note that a tweet can
396 include multiple topics.



397

398

399 Tables

400 **Table 1.** Tweet and user account characteristics are shown on top and inferred user
 401 demographics are shown on bottom.

Tweets n = 2,287,344		User Accounts N = 948,666	
Has link(s)	1,235,575 (54.0)	Years account active	6.9 (2.6 – 10.0)
Mentions user(s)	916,585 (40.1)	Followers	267 (55 – 1,100)
Has hashtag(s)	414,173 (18.1)	Following	407 (137 – 1,069)
Has media	273,278 (11.9)	Lifetime statuses	4,605 (1,027 – 16,365)
Is quoted tweet	133,404 (5.8)	Lifetime likes	3,600 (519 – 15,572)
Has like	943,639 (41.3)	Media shared	205 (36 – 875)
Has reply	548,863 (24.0)	Public lists, member	2 (0 – 12)
Has retweet	473,204 (20.7)	Contains description	800,619 (84.4)
Has quoted tweet	183,982 (8.0)	Location listed	659,720 (69.5)
<i>Twitter source</i>		Contains profile picture	902,666 (95.2)
Web App	686,296 (30.0)	Contains banner picture	721,542 (76.1)
iPhone/iPad	660,382 (28.9)	Contains link in profile	303,761 (32.0)
Android	432,862 (18.9)	Verified account	27,443 (2.9)
TweetDeck	60,845 (2.7)		
User Demographics N = 948,666			
		N (%)	Probability <i>median (IQR)</i>
<i>Entity</i>	Individual	834,224 (87.9)	0.999 (0.997 – 0.999)
	Organization	114,443 (12.1)	0.867 (0.727 – 0.999)
<i>Sex (of individuals)</i>	Female	273,400 (32.8)	0.992 (0.949 – 0.998)
	Male	560,824 (67.2)	0.996 (0.980 – 0.999)
<i>Age (of individuals)</i>	<40 years old	504,448 (60.5)	0.972 (0.896 – 0.994)

≤18 years old	109,327 (13.1)	0.660 (0.528 – 0.821)
19 to 29 years old	225,360 (27.0)	0.611 (50.4 – 0.754)
30 to 39 years old	169,761 (20.4)	0.765 (0.568 – 0.936)
≥40 years old	329,776 (39.5)	0.950 (0.734 – 0.996)

402

403 **Table 2.** Topic clusters identified by topic modeling. Words contributing to the model are shown

404 in decreasing order of weighting. The topics are labeled manually based on these words.

Possible Topic Label	Topic #	Tweets/Topic	Words contributing to topic model (in ↓ order of weighting)	Representative Tweet
Mask use and general reactions	15	811,844	people, mask, even, dont, would, take, know, die, death, need, one, still, many, kill, risk, never, work, way, yet, wear_mask	““Pretty much what it boils down to, at this point. Ignorance, arrogance, and stupidity will end up killing LOTS of people this year, I’m afraid! Be SMART. WEAR your mask. Wash your hands. Hold off on large gatherings until a safe, effective Covid-19 vaccine arrives. [link]”
Conspiracies and misinformation	13	557,301	want, think, fake, make, try, believe, conspiracy, bill_gate, money, really, gonna, real, force, shit, god, anything, anyone, hoax, black, put	“@WhiteHouse Also, isnt this a RNA vaccine? Super experimental albeit dangerous, could mean with DNA as well. Human Guinea pigs. Wouldn’t be surprised if the vaccine harms more then the COVID did.”
Impact on lockdowns on school, work, and economy	7	406,227	wait, year, lockdown, open, month, next, life, time, long, end, last, come, back, next year, week, school, ago, day, economy, away	“@[tag] @[tag] @[tag] @[tag] And even with a vaccine they will continue with the lockdowns, the social distance and the fear mongering... If not for the Covid, they will find something...”
Vaccine mechanisms and immunity	10	235,625	virus, mrna, immunity, antibody, prevent, infection, disease, spread, strain, protein, herd_immunity, mutate, symptom, immune_system, sarscov, prevent infection, mutation, cell, cause, infect	“If tests show one already had COVID-19 so one has antibodies and is now immune, CDC currently counts that as one infected and positive for COVID-19. After a vaccine, will every person vaccinated who therefore grows antibodies, be considered positive & infected? @realDonaldTrump”
First vaccinations and recipients	4	232,613	first, world, around, world first, become, first test, receive, country, first line, first person, world news, first country, person receive, world leader, government vote, make	“Thank the lord this is the beginning of the end: First patient receives Pfizer Covid-19 vaccine [link]”

			sure_pass, yearold_woman, day government, first dos, first world	
Emergency use authorizations and approvals	12	213,036	pfizer, moderna, approval, effective, pfizer_biontech, emergency_use, data, biontech, authorization, receive, regulator, next_week, approve pfizer, effective prevent, pfizer ceo, show effective, moderna effective, approve, data show, early data	“Pfizer’s Covid vaccine is days away from approval after data reveals it is 95% effective [link]”
US politics and election	1	212,597	trump, biden, realdonaldtrump, president, election, operation_warp, american, credit, speed, lie, take credit, joe Biden, gop, democrat, potus, win, vote, joe_biden, admin, america	“@realDonaldTrump If you want to take partial credit for the Covid-19 vaccine fine. You still LOST the election. In Georgia for example you are behind there by 12k votes. The recount wont change the outcome. I look forward to your predictable reply and the end of your regime.”
Stock market and pharma/biotech companies	3	205,557	market, stock, news, company, good news, biotech, drug company, pharma, price, billion, drug, surge, late, update, rise, break news, positive news, investor, pharma company, announce	“Markets are supported by both the cumulative upside surprises to the economy since the end of the recession and the apparently faster-than-expected progress toward a COVID-19 vaccine. [link]”
Clinical trials and participants	8	205,468	trial, clinical, human trial, phase, human, volunteer, participant, oxford, begin, phase clinical, show, trial participant, result, volunteer trial, ahead_large trial, ahead_large, show_promise, immune_response, test, number	“Coronavirus Vaccine Update Oxford’s COVID-19 vaccine trial in Brazil begins: Scientists say coronavirus jab may not work for older adults [link]”
Vaccine development and supporters	6	204,083	research, development, global, effort, develop, fund, researcher, global effort, join, effort develop, access, help, target, hacker, treatment, support, dolly_parton, research development, accelerate, collaboration	“As the world continues to feel the impact of COVID-19, the biopharmaceutical industry is working around the clock to identify and develop safe and effective vaccines to prevent infection, while also researching and developing new therapies to treat those infected with the virus.”
Russian response and global partners	2	187,369	russia, india, via, sputnik, russian, china, putin, serum_institute, indian, covaxin, chinese, bharat_biotech, icmr, hacker_target, time india, via nbcnews, russia sputnik, indias_serum, narendramodi, possible	“A Sputnik moment, president #Putin has announced that #Russia is the first country in the world to register a #Covid_19 vaccine. 10s of countries already requested it [link]”
FDA and CDC	14	179,396	trump, fda, failure, administration, trump administration, fda approval, food_drug, food_drug administration, fda approve, cdc,	“@CDCgov if you try and push through an unproven vaccine because of Trump’s desperation to recover from his abysmal handling of Covid-

Philanthropy and public health

Comparison to influenza

Safety and side effects from trials

		trump admin, white_house, trump claim, president_donald trump, president_donald, cuomo, trump supporter, want, take, exist_sustainable	19...good luck. No one I know myself included will be getting vaccinated.”
9	156,183	health, public, mandatory, public health, health official, health care, public trust, official, gavi_sdg, cdc_gatesfoundation, read_billgates cdc_gatesfoundation, gavi_sdg vaccination, make mandatory, read_billgates, health expert, care_worker, health minister, health care_worker, clinton, obama_bush	“@[tag] This presents a problem and crashes into the argument, should covid vaccines be mandated. I initially thought that it will need more then encouragement and common sense from the public but these vaccine deniers are going to deprive people of protection through fear. Arrest them.”
5	151,177	flu, shot, influenza, flu shot, every year, season, seasonal, die flu, every, flu death, year flu, kill, jab, take flu, first shot, virus, people, side_effect shot, compare, via	“Ok so I’m usually not super crunchy about everything but I’ve been hospitalized 2x after getting the flu shot bc of how badly I got the flu within months so I was told not to get the shot by my drs. what does that mean for COVID’s vaccine? Like what if I react the same?”
11	137,445	astrazeneca, safety, effect, johnson johnson, pause, study, unexplained_illness, astrazeneca trial, johnson_pause, long_term effect, oxford_university, pause trial, safety efficacy, efficacy, safety concern, resume, put hold, astrazeneca study, illness, side	‘AstraZeneca COVID-19 vaccine study put on hold due to suspected adverse reaction in UK participant [link]’

405

406

407 References

- 408 1. Sridhar, D. & Gurdasani, D. Herd immunity by infection is not an option. *Science* **371**, 230–
409 231 (2021).
- 410 2. Forni, G., Mantovani, A., & COVID-19 Commission of Accademia Nazionale dei Lincei,
411 Rome. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Differ* **28**,
412 626–639 (2021).
- 413 3. Caserotti, M. *et al.* Associations of COVID-19 risk perception with vaccine hesitancy over
414 time for Italian residents. *Soc Sci Med* **272**, 113688 (2021).
- 415 4. Wong, M. C. S. *et al.* Acceptance of the COVID-19 vaccine based on the health belief
416 model: A population-based survey in Hong Kong. *Vaccine* **39**, 1148–1156 (2021).
- 417 5. Feleszko, W., Lewulis, P., Czarniecki, A. & Waszkiewicz, P. Flattening the Curve of
418 COVID-19 Vaccine Rejection-An International Overview. *Vaccines (Basel)* **9**, (2021).
- 419 6. *Ten threats to global health in 2019*. [https://www.who.int/news-room/spotlight/ten-threats-](https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019)
420 [to-global-health-in-2019](https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019).
- 421 7. Alley, S. J. *et al.* As the Pandemic Progresses, How Does Willingness to Vaccinate against
422 COVID-19 Evolve? *Int J Environ Res Public Health* **18**, (2021).
- 423 8. Reiter, P. L., Pennell, M. L. & Katz, M. L. Acceptability of a COVID-19 vaccine among
424 adults in the United States: How many people would get vaccinated? *Vaccine* **38**, 6500–6507
425 (2020).
- 426 9. Malik, A. A., McFadden, S. M., Elharake, J. & Omer, S. B. Determinants of COVID-19
427 vaccine acceptance in the US. *EClinicalMedicine* **26**, 100495 (2020).

- 428 10. Shaw, J. *et al.* Assessment of U.S. health care personnel (HCP) attitudes towards COVID-19
429 vaccination in a large university health care system. *Clin Infect Dis* (2021)
430 doi:10.1093/cid/ciab054.
- 431 11. Verger, P. *et al.* Attitudes of healthcare workers towards COVID-19 vaccination: a survey in
432 France and French-speaking parts of Belgium and Canada, 2020. *Euro Surveill* **26**, (2021).
- 433 12. Saleh, S. N., Lehmann, C. U. & Medford, R. J. Early Crowdfunding Response to the
434 COVID-19 Pandemic: Cross-sectional Study. *J Med Internet Res* **23**, e25429 (2021).
- 435 13. *Q3 2020 Letter to Shareholders*.
436 [https://s22.q4cdn.com/826641620/files/doc_financials/2020/q3/Q3-2020-Shareholder-](https://s22.q4cdn.com/826641620/files/doc_financials/2020/q3/Q3-2020-Shareholder-Letter.pdf)
437 Letter.pdf.
- 438 14. McGraw, T. *Spending 2020 Together on Twitter*.
439 https://blog.twitter.com/en_us/topics/insights/2020/spending-2020-together-on-twitter.html.
- 440 15. Stieglitz, S. & Dang-Xuan, L. Emotions and Information Diffusion in Social Media—
441 Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information*
442 *Systems* **29**, 217–248 (2013).
- 443 16. Depoux, A. *et al.* The pandemic of social media panic travels faster than the COVID-19
444 outbreak. *Journal of Travel Medicine* **27**, taaa031 (2020).
- 445 17. Sinnenberg, L. *et al.* Twitter as a Tool for Health Research: A Systematic Review. *Am J*
446 *Public Health* **107**, e1–e8 (2017).
- 447 18. Chang, C.-H., Monselise, M. & Yang, C. C. What Are People Concerned About During the
448 Pandemic? Detecting Evolving Topics about COVID-19 from Twitter. *J Healthc Inform Res*
449 1–28 (2021) doi:10.1007/s41666-020-00083-3.

- 450 19. Chandrasekaran, R., Mehta, V., Valkunde, T. & Moustakas, E. Topics, Trends, and
451 Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study.
452 *Journal of Medical Internet Research* **22**, e22624 (2020).
- 453 20. Xue, J. *et al.* Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine
454 Learning Approach. *Journal of Medical Internet Research* **22**, e20550 (2020).
- 455 21. Gallotti, R., Valle, F., Castaldo, N., Sacco, P. & De Domenico, M. Assessing the risks of
456 ‘infodemics’ in response to COVID-19 epidemics. *Nature Human Behaviour* **4**, 1285–1293
457 (2020).
- 458 22. Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M. & Lehmann, C. U. An ‘Infodemic’:
459 Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the
460 Coronavirus Disease 2019 Outbreak. *Open Forum Infect Dis* **7**, ofaa258 (2020).
- 461 23. Saleh, S. N., Lehmann, C. U., McDonald, S. A., Basit, M. A. & Medford, R. J.
462 Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing
463 on Twitter. *Infect Control Hosp Epidemiol* **42**, 131–138 (2021).
- 464 24. *Systematic scoping review on social media monitoring methods and interventions relating to*
465 *vaccine hesitancy*. [https://www.ecdc.europa.eu/en/publications-data/systematic-scoping-](https://www.ecdc.europa.eu/en/publications-data/systematic-scoping-review-social-media-monitoring-methods-and-interventions)
466 [review-social-media-monitoring-methods-and-interventions](https://www.ecdc.europa.eu/en/publications-data/systematic-scoping-review-social-media-monitoring-methods-and-interventions) (2020).
- 467 25. Salathé, M. & Khandelwal, S. Assessing Vaccination Sentiments with Online Social Media:
468 Implications for Infectious Disease Dynamics and Control. *PLOS Computational Biology* **7**,
469 e1002199 (2011).
- 470 26. Dunn, A. G. *et al.* Mapping information exposure on social media to explain differences in
471 HPV vaccine coverage in the United States. *Vaccine* **35**, 3033–3040 (2017).

- 472 27. Wilson, S. L. & Wiysonge, C. Social media and vaccine hesitancy. *BMJ Global Health* **5**,
473 e004206 (2020).
- 474 28. *FDA Takes Key Action in Fight Against COVID-19 By Issuing Emergency Use Authorization*
475 *for First COVID-19 Vaccine*. [https://www.fda.gov/news-events/press-announcements/fda-](https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19)
476 [takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19.](https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19)
- 477 29. *Snsrape*. (Github).
- 478 30. Frijda, N. H. *The emotions*. (Cambridge University Press; Editions de la Maison des
479 sciences de l’homme, 1986).
- 480 31. Hutto, C. & Gilbert, E.). VADER: A Parsimonious Rule-based Model for Sentiment
481 Analysis of Social Media Text. in (2014).
- 482 32. Loria, S. *Textblob: simplified text processing*. (Textblob).
- 483 33. Mohammad, S. M. & Turney, P. D. Emotions evoked by common words and phrases: using
484 mechanical turk to create an emotion lexicon. in *Proceedings of the NAACL HLT 2010*
485 *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* 26–
486 34 (Association for Computational Linguistics, 2010).
- 487 34. Gallagher, R. J., Reing, K., Kale, D. & Steeg, G. V. Anchored Correlation Explanation:
488 Topic Modeling with Minimal Domain Knowledge. *arXiv:1611.10277 [cs, math, stat]*
489 (2018).
- 490 35. Steeg, G. V. & Galstyan, A. Discovering Structure in High-Dimensional Data Through
491 Correlation Explanation. *arXiv:1406.1222 [cs, stat]* (2014).
- 492 36. Wang, Z. *et al.* Demographic Inference and Representative Population Estimates from
493 Multilingual Social Media Data. in *The World Wide Web Conference 2019* 2056–2067 (ACM,
494 2019). doi:10.1145/3308558.3313684.

- 495 37. Yang, Y.-C., Al-Garadi, M. A., Love, J. S., Perrone, J. & Sarker, A. *Automatic Gender*
496 *Detection in Twitter Profiles for Health-related Cohort Studies*.
497 <http://medrxiv.org/lookup/doi/10.1101/2021.01.06.21249350> (2021)
498 doi:10.1101/2021.01.06.21249350.
- 499 38. *Pfizer and Biontech announce Vaccine Candidate against Covid-19 Achieved Success in*
500 *First Interim Analysis From Phase 3 Study*. [https://www.pfizer.com/news/press-](https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against)
501 [release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against](https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against).
- 502 39. AJMC Staff. A Timeline of COVID-19 Developments in 2020. *AJMC* (2021).
- 503 40. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl*
504 *J Med* **383**, 1920–1931 (2020).
- 505 41. Subbaraman, N. This COVID-vaccine designer is tackling vaccine hesitancy - in churches
506 and on Twitter. *Nature* **590**, 377 (2021).
- 507 42. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K. & Larson, H. J. Measuring the
508 impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat*
509 *Hum Behav* **5**, 337–348 (2021).
- 510 43. Shearer, E. & Mitchell, A. *News Use Across Social Media Platforms in 2020*.
511 <https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/>
512 (2021).
- 513 44. Damiano, A. D. & Allen Catellier, J. R. A Content Analysis of Coronavirus Tweets in the
514 United States Just Prior to the Pandemic Declaration. *Cyberpsychol Behav Soc Netw* **23**,
515 889–893 (2020).

- 516 45. Jang, H., Rempel, E., Roth, D., Carenini, G. & Janjua, N. Z. Tracking COVID-19 Discourse
517 on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-
518 Based Sentiment Analysis. *J Med Internet Res* **23**, e25431 (2021).
- 519 46. Kemp, S. *Digital 2020 Global Digital Overview*.
520 https://p.widencdn.net/1zybur/Digital2020Global_Report_en (2020).
- 521 47. Fontanet, A. & Cauchemez, S. COVID-19 herd immunity: where are we? *Nat Rev Immunol*
522 **20**, 583–584 (2020).
- 523 48. McNeil Jr., D. G. [https://www.nytimes.com/2020/12/24/health/herd-immunity-covid-](https://www.nytimes.com/2020/12/24/health/herd-immunity-covid-coronavirus.html)
524 [coronavirus.html](https://www.nytimes.com/2020/12/24/health/herd-immunity-covid-coronavirus.html). *The New York Times* (2021).
- 525 49. Randolph, H. E. & Barreiro, L. B. Herd Immunity: Understanding COVID-19. *Immunity* **52**,
526 737–741 (2020).
- 527