



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## An analysis of COVID-19 vaccine sentiments and opinions on Twitter



Samira Yousefinaghani<sup>a</sup>, Rozita Dara<sup>a,\*</sup>, Samira Mubareka<sup>b</sup>, Andrew Papadopoulos<sup>c</sup>, Shayan Sharif<sup>d</sup>

<sup>a</sup> School of Computer Science, University of Guelph, Guelph, Ontario, Canada

<sup>b</sup> Sunnybrook Health Sciences Center, Toronto, Ontario, Canada

<sup>c</sup> Department of Population Medicine, University of Guelph, Guelph, Ontario, Canada

<sup>d</sup> Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

### ARTICLE INFO

#### Article history:

Received 15 April 2021

Received in revised form 12 May 2021

Accepted 22 May 2021

#### Keywords:

Communicable diseases

Social media

Text mining

Vaccine

### ABSTRACT

**Objective:** We identified public sentiments and opinions toward the COVID-19 vaccines based on the content of Twitter.

**Materials and methods:** We retrieved 4,552,652 publicly available tweets posted within the timeline of January 2020 to January 2021. Following extraction, we identified vaccine sentiments and opinions of tweets and compared their progression by time, geographical distribution, main themes, keywords, posts engagement metrics and accounts characteristics.

**Results:** We found a slight difference in the prevalence of positive and negative sentiments, with positive being the dominant polarity and having higher engagements. The amount of discussion on vaccine rejection and hesitancy was more than interest in vaccines during the course of the study, but the pattern was different in various countries. We found the accounts producing vaccine opposition content were partly Twitter bots or political activists while well-known individuals and organizations generated the content in favour of vaccination.

**Conclusion:** Understanding sentiments and opinions toward vaccination using Twitter may help public health agencies to increase positive messaging and eliminate opposing messages in order to enhance vaccine uptake.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Introduction

There has been a concerted global effort to develop and test COVID-19 vaccines since the pandemic was declared in March 2020. Although public health prevention measures have proven to be somewhat effective in limiting the spread of COVID-19, protective and sustained immunity through vaccination will be of great importance in ending the pandemic. It is estimated that at least over 70% of the population will need to be vaccinated (Orenstein and Ahmed, 2017; Aguas et al., 2020) to reach some level of herd immunity. To achieve this, public support for vaccination is essential. Hence, it is extremely important to understand the public's opinion toward vaccination, and thereby their willingness to become vaccinated.

Although classical surveys are useful to investigate public health viewpoints (Peretti-Watel et al., 2020), social media is increasingly used for discussing and sharing viewpoints about infectious disease outbreaks health topics (Velasco et al., 2014; Yousefinaghani et al., 2019; Guess et al., 2020). The current COVID-19 pandemic has resulted in a surge of social media use as a forum for discussing an array of topics about the pandemic, including vaccines. Indeed, social media users can be exposed to negative sentiments and misinformation, which may influence individual views and lead to vaccine hesitancy or refusal (Piedrahita-Valdés et al., 2021). Vaccine hesitancy is considered one of the 10 major threats to global health according to the World Health Organization (WHO) (Puri et al., 2020; Kunneman et al., 2020; Piedrahita-Valdés et al., 2021). Additionally, misinformation can destroy trust in science and public health authorities and lead to a drop in vaccine uptake (Steffens et al., 2019; Bonnevie et al., 2020b; Hotez et al., 2020).

Several studies have performed simple descriptive analyses of vaccine related Twitter data to assess people's attitude towards

\* Corresponding author at: School of Computer Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada.  
E-mail address: [drozita@uoguelph.ca](mailto:drozita@uoguelph.ca) (R. Dara).

vaccination (DeVerna et al., 2021), and to identify sentiments (Piedrahita-Valdés et al., 2021), dominant opinions (Surian et al., 2016), themes (Nuzhath et al., 2020), information flow between users and most influential users for a particular sentiment (Kang et al., 2017), patterns (Huang et al., 2017; Yin et al., 2021) and discussion communities on vaccination (Bello-Orgaz et al., 2017).

Previous studies have correlated activities on social media to vaccine hesitancy and anti-vaccine movements (Tangherlini et al., 2016; Broniatowski et al., 2018; Burki, 2019; Johnson et al., 2020). Tangherlini and colleagues (2016) sought the drivers of vaccine hesitancy from blogs and they found these platforms may be used by parents to promote vaccine opposition sentiments to other parents. Bonnevie et al. (2020b) examined the progression of vaccine opposition by assessing Twitter conversation themes within the United States and concluded that influential Twitter accounts contributed to a great portion of vaccine-opposition messages. Other research (Cossard et al., 2020) examined the extent to which vaccination conversations can lead to vaccine hesitancy as well as how the structure of communities is different in skeptics and advocates. Cossard et al. (2020) showed that the clusters in anti-vaccine communities are more connected than those in pro-vaccine communities.

Addressing vaccine opposition and fostering vaccine confidence by studying the sentiments and opinions toward vaccination can aid in vaccine uptake among the population (Ferrer and Ellis, 2019). This can be accomplished by designing effective vaccine-promoting communication by tailoring messages using acquired knowledge about vaccine sentiments and opinions. Opinions for and against COVID-19 vaccination change by community features such as demographics, income, and religious or family status (Lyu et al., 2020).

The present study was designed to gain insight into sentiments and opinions toward COVID-19 vaccination through 4.5 million publicly available Twitter posts collected over one year from January 2020 to January 2021. Twitter was employed as the main source due to its popularity as a forum for discussions related to health information (Love et al., 2013). The main objectives were as follows: (1) to track frequent hashtags, frequent mentions, main keywords, and main themes in tweets with positive and negative sentiments; (2) to compare tweets with negative sentiments toward vaccine manufacturers; (3) to identify anti-vaccine, vaccine hesitant and pro-vaccine opinions in tweets; (4) to compare the evolution of tweets with different opinions posted from several locations; and (5) to identify top users' characteristics and tweets engagement in each vaccine opinion.

Published research on opinions toward COVID-19 vaccination on social media usually focused on a short-term or a single location (Bonnevie et al., 2020a; Abd Rahim and Rafie, 2020; Kwok et al., 2021). To the best of our knowledge, this is the first large-scale study of Twitter to identify opinions toward COVID-19 vaccination and their progression over time and location. Importantly, better understanding public opinion by categorizing Twitter content into anti-vaccine, vaccine hesitant and pro-vaccine opinions is critical information required by public health authorities and can support their decision-making efforts.

Obtaining data on public opinion has been challenging and governments have usually relied on surveys with limitations such as unrepresentative samples. Employing data from Twitter as a popular social media outlet might be more representative of actual opinions than survey data and provide opportunities for real-time analyses of public sentiments. Currently, little is known about public opinions regarding COVID-19 vaccines. In particular, identifying vaccine hesitancy opinions is of great importance as the refusal to take COVID-19 vaccines is concerning from a public health point of view.

## Materials and methods

### Data acquisition

We used sncrape (Sncrape, 2021) to collect historical tweets regarding the COVID-19 vaccination. A combination of “vaccine” and COVID-19 related terms (“covid”, “coronavirus”, “ncov2019” and “SARS-CoV-2”) were given as input to retrieve tweets published between January 7, 2020 and January 3, 2021. The ‘tweets’ data table include following fields: ‘id’, ‘date’, ‘tweet’, ‘url’, ‘username’, ‘outlinks’, ‘retweetCount’, ‘replyCount’, ‘likeCount’ and ‘quoteCount’.

In the current study, we utilized a user lookup Twitter API to access location and engagement information of users since sncrape does not provide any geographical information. The geographical information was identified from self-reported profile locations in Twitter which was available for approximately 70% of the included users. Furthermore, a text-matching query was implemented to detect the country that location information referred to.

In total, 4,552,652 posts were pulled from Twitter. These tweets were generated by 1,566,590 users and contained 1,012,419 hashtag and 2,258,307 mention terms.

### Vaccine sentiment analysis

To assign a polarity of ‘positive’, ‘negative’ or ‘neutral’ to each tweet, we utilized Valence Aware Dictionary and sEntiment Reasoner (VADER), a Python lexicon and rule-based sentiment analysis tool (Hutto and Gilbert, 2014). VADER is designed to determine sentiments of social media posts based on individual words and sentences (Elbagir and Yang, 2019). Preprocessing was performed on the text of tweets to discard unwanted characters and words such as punctuations, unicode errors, links, emails, currency symbols and numbers.

Further, sentiments were assigned based on scores given by VADER. Any tweet with a score of 0.25 or greater was categorized as a positive sentiment, a score of  $-0.25$  or less was categorized as a negative sentiment, and any score between those values was categorized as a neutral sentiment. Subsequently, the following assessments were conducted:

**Evolution of sentiments.** A graph showing the weekly progression of positive and negative sentiments was generated to estimate the public's sentiment toward vaccination. The temporal trends in the graph can help observe how policies, decisions and key events on vaccinations have changed people's feelings.

**Average engagement metrics.** Metadata accompanying posts included measures of engagements such as how many times each tweet was shared (retweets), was liked (favourites), people responded to (replies) and retweeted with comment (quotes). A summary of engagement metrics was calculated by averaging the metrics including retweets, favourites, replies and quotes over posts with positive or negative sentiments.

**Frequent hashtags and mentions.** Regular expressions were used to identify all hashtags and mentions in the text from all tweets. Hashtags are usually used with a prefixed # symbol to indicate the topic associated with the tweet. A mention in a tweet is recognized by a @ sign followed by a username and is used to refer to or communicate with a particular user. We stored mentions and hashtags in their specified data tables and indicated whether each hashtag or mention belongs to a tweet with positive or negative sentiment.

**Keywords and themes.** Keyword extraction algorithms can automatically identify a set of terms that are important and best represent the subject of a document. Keyword extraction starts

with detecting possible candidate keywords from the text. Subsequently, a rank is calculated for each keywords and key-phrases sorted from high to low to select the top  $n$  candidate terms. The keywords were extracted using the keywords function from the Gensim library. The text summarization module supports keyword extraction.

We created two corpora using tweets with positive and negative sentiments. Next, keyword extraction was applied to each corpus and related keywords were obtained. Finally, a few themes were recognized for each category based on the most important keywords.

**Comparing vaccine manufacturers.** With more than 100 million confirmed COVID-19 cases worldwide and the number increasing daily, different pharma companies are distributing vaccines. We compared the relative social media opinion and engagement toward major companies, such as Pfizer, Moderna, AstraZeneca and Johnson & Johnson.

### Vaccine opinions

Vaccine-related Twitter conversations were categorized into three opinions of anti-vaccine, vaccine hesitant and pro-vaccine. Individuals with anti-vaccine opinion may hold beliefs such as COVID-19 is not a serious threat and are unlikely to accept a COVID-19 vaccine. On the other hand, vaccine hesitant individuals usually are uncertain about receiving vaccines (Owen et al., 2021).

**Opinion categorization.** Three categories of opinions were identified by querying specific phrases (see Table A1 in Appendix) with Boolean operators from a sample of 500,000 tweets randomly selected throughout the entire vaccine-related data (0.125% of the corpus). The phrases were selected manually by assessing 2% of the tweets ( $n = 10,000$ ) and then verified by the experts in the field. The research team removed any tweets that were neutral to any of defined vaccine opinions by searching specific keywords in data. For example, a tweet regarding ‘vaccines taking long time to be developed’ was deemed to be vaccine-neutral and not included in the analysis.

Subsequently, the phrases were searched to assign opinion labels (i.e. anti-vaccine, vaccine hesitant and pro-vaccine). The correctness of label assignments was reviewed for 5000 or 1% of the sample tweets manually. If the wrong labels were beyond 500 posts (10%), we tuned the keywords and re-labelled until obtaining the desirable percentage of correct labels (90%).

It is worth noting that the outcome for the ‘anti-vaccine’ category was filtered using 34 terms to exclude opposite statements. For example, the tweet “Bill Gates is not secretly plotting microchips in a coronavirus vaccine. Misinformation and conspiracy theories are dangerous for everyone.” was discarded from the category of ‘anti-vaccine’ as it is expressing the opinion against anti-vaxxers.

**Evolution of opinions.** When vaccine opinions of tweets were identified, the progression of the anti-vaccine, vaccine hesitant and pro-vaccine opinions was compared over time. Also, the advancement of vaccine opinions in the United States, Australia, India, the United Kingdom, Canada and Ireland was explored. These English-speaking countries were selected as the tweets from these countries covered a large proportion of total tweets that were collected in the study. Therefore, the selected countries were more representative of the evolution of vaccine opinions. The number of posts in each country for durations of January 2020 to April 2020, May 2020 to August 2020 and September 2020 to December 2020 was plotted.

**Top users’ characteristics and tweets engagements.** The top users of each opinion were identified and then reviewed to understand potential differences in the occupations and intentions of each opinion’s users. Top users of each opinion are defined as the

first 100 Twitter accounts that contributed to the tweets labelled with that opinion. Further, the average characteristics of each opinion’s users such as verified status, number of friends and number of followers were examined. Verified status can be considered as a badge given by Twitter to users that are authentic. Followers are defined as users following a specific user while friends are users that a specific user is following. Next, the count and engagement metrics, i.e. the number of retweets, favourites, replies and quotes, of tweets published by the top 100 users in each opinion were compared.

## Results

A total of 4,552,652 tweets were collected between the January 7, 2020 and January 3, 2021 inclusive. The distribution of the collected vaccine-related tweets over one year is shown in Figure 1. The highest number of tweets was posted during the second week of December with about 400,000 tweets. There was a slight increase in the number of vaccine-related tweets in June (weeks 13, 14, and 15) due to news regarding evidence that a tuberculosis vaccine might help fight SARS-CoV-2. The majority of Twitter users wanted to know whether this existing vaccine can prevent the spread of the COVID-19. However, the second increase in the start of November, was due to the announcement of success in the development of various vaccines.

### Vaccine sentiment analysis

VADER categorized the tweets into three categories of positive, negative and neutral. Figure 2 shows the percentage of tweets in each category. The neutral category accounted for the 41% of the tweets, followed by the positive category accounting for 34% and negative category accounting for 25%. The negative sentiments were related to a range of concerns, but the majority usually focused on vaccine development being time-consuming, doubts in vaccine safety or reaction to governments, political figures and manufacturers. On the other hand, positive tweets were usually about scientific breakthroughs, medical advice and spreading hope.

The number of tweets with positive and negative sentiments per week can be found in Figure 3. In general, positive tweets were the dominant sentiment in almost all weeks over the course of the study. In particular, there were a number of elevations in March, May, July and November.

As shown in Table 1, positive tweets gained higher average engagement metrics than tweets with negative sentiments. For example, positive tweets received 20 favourite hits on average compared with 15 hits in negative tweets.

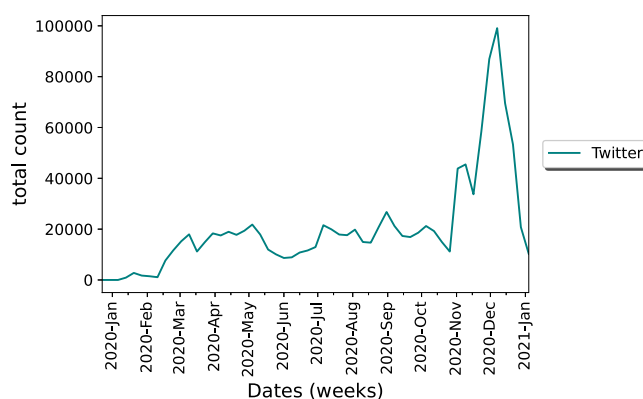


Figure 1. Distribution of COVID-19 vaccine-related tweets.

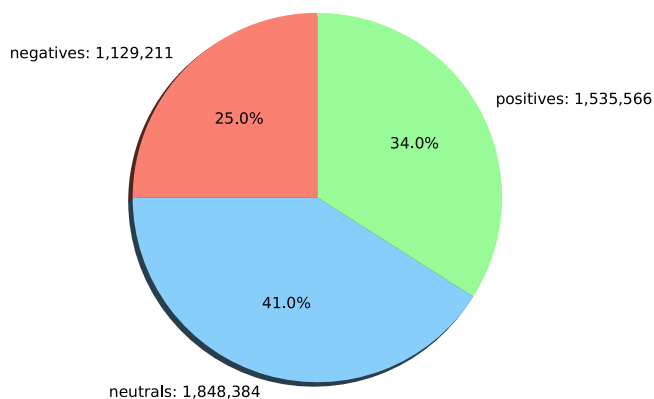


Figure 2. Distribution of sentiment polarities of tweets.

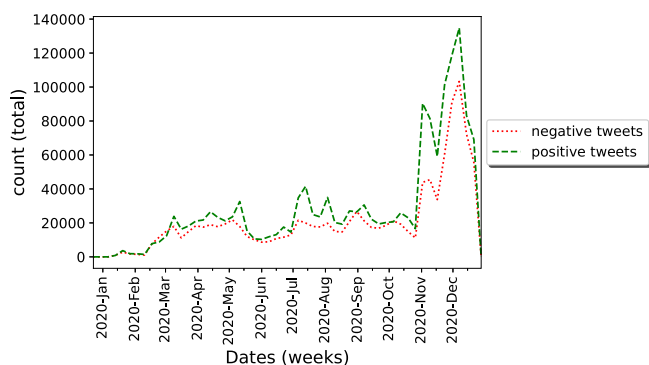


Figure 3. Evolution of public's sentiments toward vaccination in Twitter.

Table 1

Average engagement metrics for vaccine sentiments.

Sentiment polarity	Retweets	Favourites	Replies	Quotes
Positives	4.56	20.43	1.87	0.8321
Negatives	4.33	15.79	1.39	0.5899

Table A2 in the appendix lists the most tweeted hashtags and mentions in each sentiment. Positive hashtags were mainly related to medical research and health services while negative hashtags were related to commentary on political figures and conspiracy theories. The mentions in positive tweets were usually Channels and Alliances of vaccine development, while the negative mentions included popular news agencies.

Table 2

Extracted keywords and themes for positive and negative polarities.

Sentiment polarity	Themes	Keywords
Positives	Happiness and Hope	congratulates, thrill, appeal, cheerful, motivator, adorable, innovators, fitness, favouring, fortunes, adventures, confident, overcoming, efficiency, winnings, productivity, fascinating
	Support	aid, kindness, hug, pardoned, wished, truthfulness, greatness, devoting, appreciating, consulted, facilitate, propel, assistance
	Religion	blesses, faiths, forgiveness, thankful
Negatives	Fear and Frustration	distorted, adverse, torturous, hoard, wrecks, poisoned, harassment, jeopardize, untruths, neglectful, forces, plummeted, horrified, shatter, catastrophic, stupid, obfuscate, annoyed, wound, obsession, exacerbating, denounce, immune, refutes, blackness, flaw, conservatives, destruct, hurting, confounds, lament, succumbing, snubbing, inhumanity, reformed, deteriorate, wasted, questioning, resistance, weakening, impaired, bitching, frightens, collapsed, ridicule, frustrates, alcoholism, distrust, disparaged, punished, distressed, overwhelms, shocks, paradoxically, troubled, victims, depressed
	Disappointment	Trap, stealing, misguide, misleads, fool, scammed, impossible, discouraged, blackmail, poisoned, pocketed, disregarding, slanders
	Anger	disagreeing, shooting, angering, aggravated, outrage, terrifies, crazies, hates
	Politics	dominate, narcissists, strike, immigrant, corrupting, enslavement, control, implants, finishers, war, terrorizes, blinding, conspiracy, prosecutions, politicized, misrepresents, mocked, arrest, jail

In addition to frequent hashtags and mentions, we compared important keywords, themes and the most active users and in each sentiment group. As shown in Table 2, the majority of keywords in the positive polarity were related to hope and happiness, followed by support and spirituality. The keywords in the negative polarity were largely related to fear and frustration, followed by disappointment, anger and political themes. The results presented in Table A3 in the appendix reveal that the most engaged authors of positive content were news agencies such as The Reuters while personal accounts were the most active in creating contents with negative sentiment.

A comparison between various vaccine manufacturers is depicted in Figure 4. In general, the negative sentiments regarding main vaccine brands started at the end of August 2020. The negative conversations related to Pfizer saw the first peak in late October and the second peak in late November. During the same time period between October and December, a rise in negative sentiments about Moderna and AstraZeneca occurred, but about four times less than those regarding the Pfizer vaccine. Other considerable elevation in negative posts was in late August when AstraZeneca started the final stage of trials and early October when Johnson & Johnson resumed the investigation of the Janssen vaccine.

### Vaccine opinions

We categorized 500,000 randomly selected tweets into 95,584 anti-vaccine, 88,486 hesitant, 37,278 pro-vaccine and 278,652 neutral tweets. As shown in Figure 5, at the beginning of the year, pro-vaccine tweets were the dominant opinion. However, in the middle of March when COVID-19 was declared as a pandemic, anti-vaccine posts started to increase and stayed at the same level until the end of May. By the start of July, the vaccine hesitant content started to slightly increase and remained steady until early November. Moreover, we found an opposite trend between anti-vaccine and pro-vaccine plots in several points in time. When anti-vaccine discussions rose in early April and late December, there were slight decays in the users' desire to take COVID-19 vaccines.

In early November, pro-vaccine tweets jumped to approximately 4000 but dropped in the following 15 days. A possible explanation of this could be the announcement of success in the Pfizer vaccine on November 9th. At the same time, the hesitant and anti-vaccine content had started to increase, with the hesitant posts reaching a peak twice as high as it was for anti-vaccine. Interestingly, the higher the growth in anti-vaccine activities, the higher the doubt and uncertainty in taking vaccines.

Subsequently, the distribution of vaccine opinions was visualized across three periods (i.e. every four months) in the United States, Australia, India, the United Kingdom, Canada and Ireland.



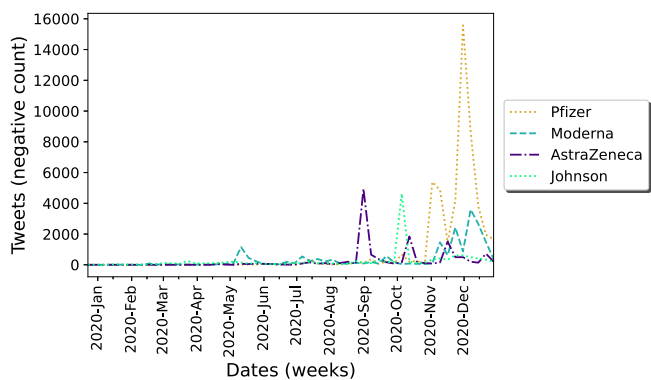


Figure 4. Negative posts regarding vaccine brands.

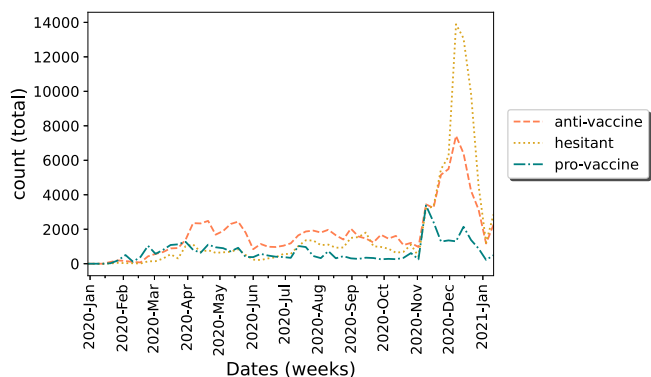


Figure 5. Evolution of vaccine opinions in Twitter.

The results in Figure 6 revealed a relatively high number of positive opinions towards vaccination for all three periods in India, Canada and Ireland. Interestingly, in India, for all three periods, the pro-vaccine activity was dominant, albeit from September, the number of tweets regarding hesitancy towards vaccination grew dramatically. On the other hand, an opposite trend was the case for the United States and the United Kingdom.

Furthermore, we identified the most active and influential users in each vaccine opinion group. The most important authors in the pro-vaccine group were well-known people and agencies, including public health institutions, physicians, television channels, newspapers, international organizations, health professors and writers.

For anti-vaccine tweets, many of the top users were already suspended by Twitter. However, among those that we could review, several Twitter bots were found. Bots are accounts that are recently created and retweet other tweets automatically. The screen names of the majority of these accounts are a name followed by a number. Other users posting anti-vaccine tweets included political activists, authors and artists. In the present study, political activists are defined as Twitter users that the main theme of their posts is politics.

In tweets with hesitant labels, we found some media pages publishing about the outbreak and breaking news regarding vaccines. We noticed a few suspended accounts in this group as well. The rest of these accounts were general Twitter users with small engagement metrics.

In further analysis, we calculated the average characteristics of users in each opinion group. The obtained results in Table 3 reveal that only 4.7% of anti-vaccine accounts were verified users while it was three and four times more for hesitant and pro-vaccine groups, respectively. A similar pattern was found for the number of followers with about 74,000 followers for users with opposite

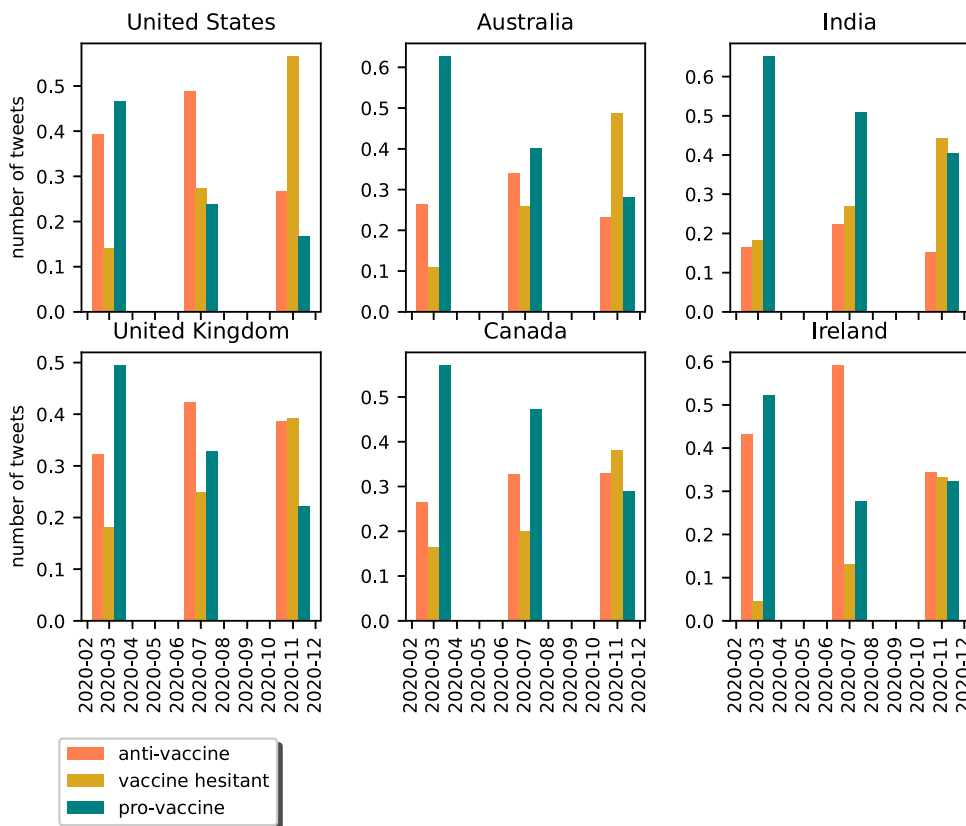


Figure 6. Evolution of vaccine opinions across countries.

**Table 3**  
Average characteristics of Twitter users across vaccine opinions.

Vaccine opinion	Verified	Friends	Followers
Anti-vaccine	4.7%	2915	73,936
Hesitant	15%	3910	257,536
Pro-vaccine	20%	3424	304,259

**Table 4**  
Average engagement metrics for tweets by top 100 users.

Vaccine opinion	Count	Retweets	Favourites	Replies	Quotes
Anti-vaccine	58.79	4.17	5.61	0.64	0.36
Hesitant	36.54	5.10	10.28	1.67	1.54
Pro-vaccine	21.56	8.32	22.20	2.83	1.77

opinions toward vaccination. However, there was no considerable difference in the average number of friends for opinion groups.

Finally, we compared the number of posts and average engagement metrics of posts published by top 100 users in opinion groups. In Table 4, it is apparent that anti-vaccine individuals published more posts, but their engagement metrics were low whereas pro-vaccine group published less posts with higher engagement metrics. One explanation for this could be the fact that posting against vaccines may have been done through automatic methods (bots) with accounts that have a limited number of followers.

## Discussion and conclusions

In the present study, sentiment and opinion analyses of approximately 4.5M tweets concerning COVID-19 vaccines. The Twitter platform that was used in the present study may be a valuable tool for public health promotion to reinforce vaccine acceptance and decrease vaccine hesitancy and opposition.

Overall, understanding sentiments and opinions toward vaccination can help public health authorities reinforce positive language and comments within the positive posts while dispelling combative language promoting misinformation within negative posts.

Moreover, public health agencies may be able to work through Twitter and other media outlets to increase positive messaging, reduce negative and opposing messages and pro-actively suspend anti-vaccination accounts such as bots in order to encourage and enhance the uptake of a vaccine.

The results here revealed that the patterns of defined sentiments and opinions have changed in response to vaccine-related events during the pandemic. In general, the positive sentiment about the COVID-19 vaccine was the dominant polarity on Twitter. This is aligned with findings of other recent studies on sentiments towards vaccination (Kwok et al., 2021; Hussain et al., 2021). Kwok and colleagues (2021) found that positive sentiments on COVID-19 vaccines formed two-thirds of sentiments. Similarly, Hussain and colleagues (2021) found that the overall sentiments of tweets and Facebook posts related to vaccination were positive in the US and UK. In addition, positive tweets showed larger engagement metrics than negative tweets in the present study, which is in line with the findings of Piedrahita-Valdés et al. (2021). Moreover, our results showed that vaccine objection and hesitancy were generally more prevalent than vaccine interest, but opinions showed different patterns for each country.

The main topics in positive tweets included hope, support and faith while negative tweets were usually related to fear, discouragement, anger and politics. Similarly, in the study by Kwok and colleagues (2021), “fear” was identified as the top

negative emotion in tweets. In positive content, news organizations were found to be the most active content writers. Moreover, health alliances and business channels have received more attention. On the other hand, in negative content, personal accounts have been the most active users and news agencies have been mentioned the most. The implication of these results can suggest that authorities should direct their focus to personal accounts to stop the dissemination of negative feelings.

The results presented here regarding opinions toward vaccination have disclosed that anti-vaccine accounts on Twitter were partly Twitter bots generating automatic content as well as political activists, authors and artists. These content writers generated a significant number of posts but attracted low engagement. On the other hand, pro-vaccine accounts were well-known individuals and international organizations that although their posting rate was low, they had a high engagement.

Despite the effort we made to perform a large-scale study on vaccine opinions, the current study might have some potential limitations: (1) Although English is a wide spoken language in social media, it cannot completely represent all vaccine-related discussions on Twitter. (2) The sample data we used in identification of anti-vaccine, pro-vaccine and hesitant tweets might not be a complete representative of vaccine-related data on Twitter. (3) Our data query to pull vaccine-related tweets was confined to COVID-19 and vaccination terms, which might not cover 100% of the desired content. (4) The approach we used to categorize opinions might have missed some posts as we did not review the entire corpus to find phrases. An exhaustive review of the corpus manually would have not been possible in terms of labour and time given that we collected over 4 million tweets. Future research can extend to build a training set with labelled tweets in order to use a machine learning model to gain greater accuracy in classifying vaccine opinions. (5) Despite the effort we made to standardize the geographical information of users, the user-defined profile locations cannot necessarily represent the actual locations that tweets were posted from.

Additionally, in the present study, data duration covers only a certain period of vaccine news and updates. In this duration, Twitter users have been mostly discussing updates on vaccine development phases, prediction of vaccine availability, vaccine approvals and early vaccine injections. Future work can include vaccine-related tweets after January 3, 2021, when vaccines were actively being received by people.

Identification of negative sentiments on social media can help reduce the impact of misinformation. One possible way to achieve this is to tag posts containing misinformation with content warnings so that readers can hide or block those posts. Future research might study the motivations for creating and sharing misinformation. For example, users can share misinformation without believing in the content as self-report data is subject to social desirability bias. The collaboration between researchers and social media platform companies is a way to discard misinformation more effectively. Social media platforms can leverage their approach toward concealing vulnerable individuals and misinformation contents by providing open data to researchers, including people's network relationship and demographic information.

## Funding

This research did not receive any specific funding.

## Authors' contribution

SY collected the data and conducted the experiments. All authors were involved in developing the ideas, design and analysis of the study and writing the manuscript.

## Ethical approval

No ethical approval was required.

## Conflict of interest

The authors have no competing interests to declare.

## Acknowledgments

This research is supported in part by the University of Guelph's Food from Thought initiative.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:[10.1016/j.ijid.2021.05.059](https://doi.org/10.1016/j.ijid.2021.05.059).

## References

- Abd Rahim N, Rafie SM. Sentiment analysis of social media data in vaccination. *Int J* 2020];8(9).
- Aguas R, Corder RM, King JG, Goncalves G, Ferreira MU, Gomes MGM. Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. *medRxiv* 2020];.
- Bello-Organ G, Hernandez-Castro J, Camacho D. Detecting discussion communities on vaccination in twitter. *Future Gen Comput Syst* 2017];66:125–36.
- Bonnevie E, Gallegos-Jeffrey A, Goldberg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *J Commun Healthc* 2020a];1–8.
- Bonnevie E, Goldberg J, Gallegos-Jeffrey AK, Rosenberg SD, Wartella E, Smyser J. Content themes and influential voices within vaccine opposition on Twitter, 2019. *Am J Public Health* 2020b];110(S3):S326–30.
- Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am J Public Health* 2018];108(10):1378–84.
- Burki T. Vaccine misinformation and social media. *Lancet Digital Health* 2019];1(6):e258–9.
- Cossard A, Morales GDF, Kalimeri K, Mejova Y, Paolotti D, Starnini M. Falling into the echo chamber: the Italian vaccination debate on Twitter. *Proceedings of the international AAAI conference on web and social media* 14 2020];130–40.
- DeVerna M, Pierri F, Truong B, Bollenbacher J, Axelrod D, Loynes N, et al. CoVaxxy: a global collection of English Twitter posts about COVID-19 vaccines. 2021. [arXiv:210107694](https://arxiv.org/abs/210107694).
- Elbagir S, Yang J. Twitter sentiment analysis using natural language toolkit and VADER sentiment. *Proceedings of the international multicongference of engineers and computer scientists*, vol. 122 2019];16.
- Ferrer RA, Ellis EM. Moving beyond categorization to understand affective influences on real world health decisions. *Soc Pers Psychol Compass* 2019];13(11):e12502.
- Guess AM, Nyhan B, O'Keeffe Z, Reifler J. The sources and correlates of exposure to vaccine-related (mis) information online. *Vaccine* 2020];38(49):7799–805.
- Hotez PJ, Nuzhath T, Colwell B. Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. *Curr Opin Virol* 2020];41:1–7.
- Huang X, Smith MC, Paul MJ, Ryzhkov D, Quinn SC, Broniatowski DA, et al. Examining patterns of influenza vaccination in social media. *Workshops at the thirty-first AAAI conference on artificial intelligence*. .
- Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. *J Med Internet Res* 2021];23(4):e26627.
- Hutto C, Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*, vol. 8 2014];.
- Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. *Nature* 2020];582(7811):230–3.
- Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine* 2017];35(29):3621–38.
- Kunnenman F, Lambooi M, Wong A, Van Den Bosch A, Mollema L. Monitoring stance towards vaccination in twitter messages. *BMC Med Informatics Decis Making* 2020];20(1):1–14.
- Kwok SWH, Vadde SK, Wang G. Twitter speaks: an analysis of Australian Twitter users' topics and sentiments about COVID-19 vaccination using machine learning. *J Med Internet Res* 2021];.
- Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. *Am J Infect Control* 2013];41(6):568–70.
- Lyu H, Wu W, Wang J, Duong V, Zhang X, Luo J. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. 2020. [arXiv:201202165](https://arxiv.org/abs/201202165).
- Nuzhath T, Tasnim S, Sanjwal RK, Trisha NF, Rahman M, Mahmud SF, et al. COVID-19 vaccination hesitancy, misinformation and conspiracy theories on social media: A content analysis of Twitter data. 2020 doi:1031235/osfio/vc9jb.
- Orenstein WA, Ahmed R. Simply put: vaccination saves lives. *National Acad Sciences*; 2017.
- Owen T, Loewen P, Ruths D, Bridgman A, Mohammad SH, Merkley E, et al. Understanding vaccine hesitancy in Canada: attitudes, beliefs, and the information ecosystem. 2021.
- Peretti-Watel P, Seror V, Cortaredona S, Launay O, Raude J, Verger P, et al. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *Lancet Infect Dis* 2020];20(7):769.
- Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, et al. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. *Vaccines* 2021];9(1):28.
- Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines Immunotherap* 2020];1–8.
- snsrape. snsrape. <https://github.com/JustAnotherArchivist/snsrape> [last accessed 01 February 2021].
- Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. *BMC Public Health* 2019];19(1):1–12.
- Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *J Med Internet Res* 2016];18(8):e232.
- Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, et al. "Mommy blogs" and the vaccination exemption narrative: results from a machine-learning approach for story aggregation on parenting social media sites. *JMIR Public Health Surveill* 2016];2(2):e166.
- Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: a systematic review. *Milbank Quart* 2014];92(1):7–33.
- Yin F, Wu Z, Xia X, Ji M, Wang Y, Hu Z. Unfolding the determinants of COVID-19 vaccine acceptance in China. *J Med Internet Res* 2021];23(1):e26089.
- Yousefinaghani S, Dara R, Poljak Z, Bernardo TM, Sharif S. The assessment of Twitter's potential for outbreak detection: avian influenza case study. *Sci Rep* 2019];9(1):1–17.