2022

A comparative analysis of anti-vax discourse on twitter before and after COVID-19 onset

Tareq Nasralah
Northeastern University

Ahmed Elnoshokaty
Northern Michigan University

Omar F. El-Gayar
Dakota State University

Mohammad Al-Ramahi
Texas A&M University

Abdullah Wahbeh
Slippery Rock University of Pennsylvania

Original Research Article

Health Informatics Journal

Health Informatics Journal
1–17

DOI: 10.1177/14604582221135831
journals.sagepub.com/home/jhi
SAGE

A comparative analysis of anti-vax discourse on twitter before and after COVID-19 onset

Tareq Nasralah
Supply Chain and Information Management Group, D'Amore-McKim School of Business, Northeastern University, Boston, MA, USA

Ahmed Elnoshokaty
Northern Michigan University, Marquette, MI, United States

Omar El-Gayar
Dakota State University, Madison, SD, United States

Mohammad Al-Ramahi
Texas A&M University-San Antonio, San Antonio, TX, United States

Abdullah Wahbeh
Slippery Rock University of Pennsylvania, Slippery Rock, PA, United States

Abstract

This study aimed to identify and assess the prevalence of vaccine-hesitancy-related topics on Twitter in the periods before and after the Coronavirus Disease 2019 (COVID-19) outbreak. Using a search query, 272,780 tweets associated with anti-vaccine topics and posted between 1 January 2011, and 15 January 2021, were collected. The tweets were classified into a list of 11 topics and analyzed for trends during the periods before and after the onset of COVID-19. Since the beginning of COVID-19, the percentage of anti-vaccine tweets has increased for two topics, "government and politics" and "conspiracy theories," and decreased for "developmental disabilities." Compared to tweets regarding flu and measles, mumps, and rubella vaccines, those concerning COVID-19 vaccines showed larger percentages for the topics of conspiracy theories and alternative treatments, and a lower percentage for developmental disabilities. The results support existing

Corresponding author:
Abdullah Wahbeh, Slippery Rock University, ATS 256, 1 Morrow Way, Slippery Rock, PA 16057, USA.
Email: abdullah.wahbeh@sru.edu

anti-vaccine literature and the assertion that anti-vaccine sentiments are an important public-health issue.

Keywords
analytics, anti-vaxxers, COVID-19, social media, vaccines

Introduction

The Coronavirus Disease 2019 (COVID-19) Pandemic has dramatically influenced many people's lives. Globally, social guidelines and restrictions have been implemented to contain the virus' spread, and several vaccines have been authorized. However, despite increasing evidence of the vaccines' efficacy and safety, many people remain reluctant to receive vaccination. Members of this group are known as "vaccine hesitant" or "anti-vaccinationists (anti-vaxxers). The World Health Organization recognizes vaccine hesitancy as a major threat to global health.[1]

Understanding the threat posed by antivaxxers on social media is critical to any vaccination program[2] as the uptake of many vaccines continues to be suboptimal.[3] With social media platforms playing a growing role in health communities,[4] a growing interest in these platforms has emerged by health informaticians[3] to understand how social media platforms become a reflection of anti-vaxxers beliefs and practices.[4] In facts, health informatics in conjunction with machine learning and data science have been applied in different real-life applications including healthcare analytics.[5] The text analytic framework proposed in this study sheds new light on social media data and its potential in public health surveillance.

Social-media platforms afford information-sharing regarding vaccines;[6] however, such information can increase vaccine-hesitant behaviors and anti-vaccine sentiments.[7] Anti-vaxxers utilize social media to manipulate public emotions, promote conspiracy theories and misinformation, and create divisions among the public.[8]

To obtain a holistic view and insights into prevailing vaccine-related topics and issues, it is necessary to analyze anti-vaxxers' discussions on social media. Such insights could inform public awareness campaigns for reducing social-media-based anti-vaccine movements' impact. Few studies have considered COVID-19-vaccine hesitancy in the context of social media, with those that have being limited in regard to sample data size and data-analysis methods.

Accordingly, this study aimed to leverage text analytics (specifically, topic-mining) to analyze negative discourse regarding COVID-19 vaccines in the US. In contrast to prior research, the current study aims to analyze shifts in the relative prevalence of various topics related to vaccine hesitancy. Specifically, we examine a relatively large dataset extracted from Twitter and seek to identify, track, and analyze topics associated with COVID-19-vaccine hesitancy and rejection. We compare the popularity of such topics before and after the onset of COVID-19, and against sentiments towards the well-known vaccines for influenza (flu) and measles, mumps, and rubella (MMR).

Background and significance

Sharing concerns about vaccination on social-media platforms could negatively affect the vaccination process.[9] Vaccine scare has been considered as a major health issue over the past decade.[10] Such scares is a recent phenomenon that is characterized by mass media posting that generates panic about health interventions, such as vaccines.[11] Accordingly, several studies have aimed to analyze

the social-media-based vaccine movement and identify best practices and guidelines for increasing public trust in vaccines.[1,9,12–14] Dhaliwal and Mannion[1] explored public perception of vaccination through analysis of social-media platforms, categorizing the obtained data into truth, consequences, and myths, respectively. The analysis showed that claims about vaccines ranged from questioning the ethics of vaccination to vaccine's benefits, truths consist of information's supported by scientific evidence, and autism was the main concern when to comes to vaccination. Meanwhile, Tara and Rubinstein[15] examined the major themes present in a popular anti-vaccine broadcast, identifying "they are lying to you," "civil liberties," "everyone is an expert," "science will not save us," "skew the science," and "they are out to harm you."

Gunaratne, Coomes, and Haghbayan[16] analyzed trends in pro-vaccine and anti-vaccine discussions on Twitter, finding a lower number of anti-vaccine than pro-vaccine tweets, and despite an increase in anti-vaccine Twitter users, no increase in anti-vaccine tweets since 2014. Massey et al.[17] characterized pro-human-papilloma-virus (HPV) and anti-HPV vaccine networks on Instagram, finding that, in contrast to pro-HPV-vaccine posts, anti-HPV-vaccine posts originated from individuals and included personal narratives. Topics among anti-HPV-vaccine posts included misinformation, vaccine debate, evidence base, and health beliefs.

Kang et al.,[14] analyzing vaccine-related data on Twitter using semantic networks, found that the positive-sentiment network centered on parents and emphasized communicating health risks and benefits, while the negative-sentiment network centered on children and emphasized organizational bodies. Ruiz, Featherstone, and Barnett[7] analyzed Twitter data regarding three vaccines and identified vaccine influencers and their online communities. Sentiment analysis revealed 3 influencer communities: focusing on the dangers of childhood vaccines and showing negative sentiments; focusing on promoting vaccines and showing a neutral sentiment; and focusing on increasing and encouraging vaccination rates and showing positive sentiments.

Social media and COVID-19 vaccines

Lyu et al.,[9] using Twitter data, analyzed public opinions concerning the potential of COVID-19 vaccines. They found that socioeconomically disadvantaged users held polarized opinions on vaccines, and that anti-vaccine opinions were strongest among users with the worst pandemic experience (e.g. sickness in one's family). The major topics identified were safety, effectiveness, and politics.

Wu et al.[12] examined COVID-19-vaccine concerns by analyzing active users on Reddit, finding that the top-10 topics were skeptical/aggressive remarks, clinical trials/research/testing, life/family/kids, people/vaccine efficacy/risks, governments/big companies, symptoms/immune systems, time/long-term effects, stock market/sports, politics/news sources, and lockdown/spread/cases. Jamison et al.[18] analyzed 2000 Twitter accounts, of which 45% opposed vaccinations, finding that most of the vaccine opponents' tweets concerned public-health topics, news topics, discussion topics, conspiracy theories, insinuation/rumors, and scams.

Bonnevie et al.[19] evaluated shifts in vaccine opposition by comparing online conversations during the 4 months before and after the COVID-19 outbreak, respectively, finding that vaccine opposition on Twitter increased by 80%. 11 themes were identified: negative health impacts, pharmaceutical industry, policies and politics, vaccine ingredients, federal health authorities, research and clinical trials, religion, vaccine safety, disease prevalence, school, and family. Similarly, Quintana et al.[20] analyzed Twitter data for the 75 days preceding and succeeding the declaration of the COVID-19 Pandemic, respectively, finding that vaccine-related discussions increased during the pandemic, that a small community of unorthodox users were ambivalent regarding vaccines, and

that the moral and non-moral language used by a number of communities suggested a trust-first model of political engagement.

Overall, few existing studies on vaccine hesitancy and anti-vaxxers have explored the changes, from the pre-COVID-19 period to the post-COVID-19 period, in the respective popularities of associated topics, or differences in discourse for different vaccines. Bonnevie et al.[19] reported only increased vaccine opposition on Twitter and more tweets concerning certain topics associated with vaccine hesitancy. Other studies only identified topics associated with vaccine hesitancy in the COVID-19 era.[9,12–14,20,21] No studies have investigated changes in anti-vaccine topics during the pandemic and compared such changes across different vaccines; only one study has compared the shift in vaccine-hesitancy topics before and after the COVID-19 outbreak.[20] The present study used manual content analysis to identify the main anti-vaccine topics. The current study extends the literature by analyzing COVID-19-related anti-vaccine discourse on social media over an extended period of time, and by comparing anti-vaccine topics and trends over time across different vaccines.

Methodology

The present study's methodology comprised 3 activities: collecting relevant social-media data, identifying anti-vaccine-related topics using extant literature and topic-modeling, and analyzing the identified topics (Figure 1).

Data collection and preprocessing

Social media platforms, such as Twitter, have been widely used in health related crises[22] and research,[23] and considered to be the fastest and most convenient source of information[24,25] to address these crises and help understand how the populations respond to them.[22]

Twitter was selected as a data source because it is commonly used by anti-vaxxers.[4] To identify relevant anti-vaccine tweets, a *search query* (Appendix A) was developed by reviewing the relevant literature and identifying a list of search terms that reflect the negative sentiments on vaccines that are commonly presented by anti-vaxxers. Using Brandwatch, a social-media data collection and analytics tool, we collected tweets matching the search query that were posted between 1 January 2011, and 15 January 2021 (excluding retweets and tweets with URLs). Twitter content is available from January 2011 in Brandwatch.

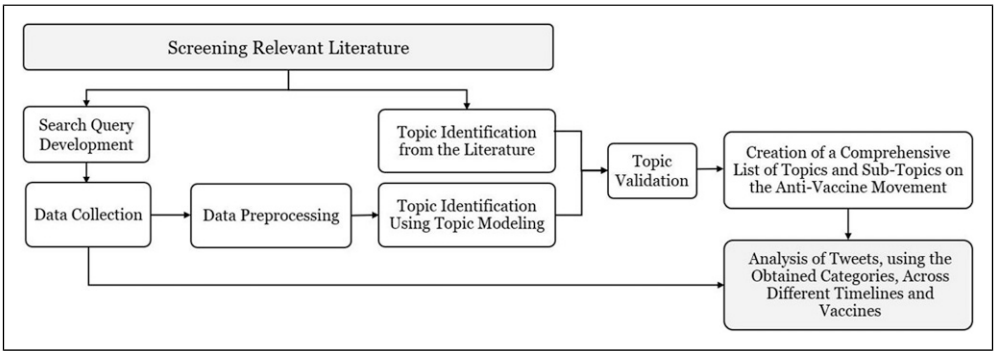

Figure 1. Research methodology.

Next, collected tweets were processed by removing stop words, user identifiers, and hashtags. The tweets were then represented using word-level n-grams;[26] for example, "autism," "conspiracy theories," and "vaccines contain mercury".

Topic identification and validation

We identified relevant topics by screening literature concerning the anti-vaccine movement on social media, performing topic-modeling using the latent Dirichlet allocation (LDA) algorithm,[27,28] visualizing and labeling the topics from LDA using PyLDAVis[29] and t-distributed stochastic neighbor embedding (t-SNE),[30] and combining the resultant topics into one list.

Topic models are statistical-based models for uncovering themes from a large unstructured collection of documents.[27,31] A topic model can help automatically summarize textual data and simplify manual content analysis. We optimized the LDA model using the coherence score measure.[32]

Latent Dirichlet allocation requires specifying the number of topics. According to the literature, the number of topics could be determined using a number of measures such as perplexity and coherence.[33] For LDA applications in which end-users will interact with the generated topics, coherence is considered the best measure[34] since it leads to better human interpretability of topics[33] compared to the perplexity method since it is not stable and the LDA results using perplexity measure could vary with seeds for the same dataset.[35]

To label the topics, the LDA results were visualized using PyLDAVis and t-SNE. The labeling process was based on the 30 most relevant terms returned in the visualization and their estimated overall term frequency within each topic. To ensure the validity and consistency of the topic labels, two independent researchers labeled the topics. Inter-rater reliability (kappa statistic)[36] was evaluated to ensure that the researchers assigning topic labels would eventually obtain similar evaluations.

The final list of topics was generated by merging the list of topics from the literature review and the results of the topic-modeling. This merge involved comparing the listed topics and their meanings and synthesizing the topics into a final list of high-level topics. This process was conducted by one researcher and validated by another.

Topic analysis

The final list of high-level topics was used for analyzing the collected tweets; this was performed using the ReadMe algorithm.[37] The ReadMe algorithm is a supervised learning algorithm that requires sample tweets (training data) to be manually labeled into a list of predefined topics. ReadMe is an automated nonparametric content analysis method[37] that is widely used in social science applications where the interest is to determine the aggregate proportion of all documents that belong to predefined categories.[38] ReadMe estimates the "aggregated distribution of opinions" instead of focusing on individual classification of each single text.[37] The ReadMe deploys a "word-profile of each category" based on the training data set, then the text of the training data set is compared to these profiles, and then a fit estimation for each category is generated for test data.[38]

The algorithm is practical for analysis aiming to show how tweets spread across different topics, and provides an unbiased text classification when compared to traditional supervised learning techniques.[37] We trained the ReadMe algorithm by manually labeling a sample set of tweets from each predefined topic, and then used the trained model to analyze the entire collection of tweets.

A sample of 110 tweets was used to assess the manual-labeling process and ensure the reliability and consistency of the manual training process for the ReadMe algorithm. Two researchers independently assigned labels to each tweet based on the obtained topics from the "topic identification and validation" step. The kappa statistic was again used as a measure of inter-rater reliability.[36]

Based on the results from the ReadMe algorithm, we completed the following analyses: First, we analyzed the distribution of tweets over topics and time. Second, we analyzed the distribution of tweets across different topics by considering tweets before and after 1 February 2020 (February 2020 was chosen because the US Centers for Disease Control confirmed the first US COVID-19 case on 21 January 2020).[21] Third, we analyzed the distribution of tweets across different topics for seasonal influenza, MMR, and COVID-19 vaccines, respectively. The tweets concerning each vaccine were identified by filtering the data based on vaccine-name-related keywords. The keywords for seasonal influenza were (flu OR Influenza); those for measles, mumps, and rubella were (MMR OR MPR OR MMRV OR measle* OR mump* OR rubella); and those for COVID-19 were (coron* OR covid* OR "chinesevirus" OR "china virus" OR "wuhanvirus" OR "SARS-CoV-2").

Results

The search query returned 272,780 tweets posted by 125,461 Twitter users.

Topic identification and validation

According to the literature, and as shown in Table 1, anti-vaxxers have concerns regarding vaccines' safety and potential for harmful side-effects, including death. Anti-vaxxers question vaccines' effectiveness and believe that vaccine mandate contravenes their freedom of choice. They also tend to believe conspiracy theories (e.g. that vaccines are promoted for reasons beyond protecting the population from diseases). Anti-vaxxers can also believe that vaccines are not ethical and against religious beliefs, and that vaccines are not necessary and can be substituted by alternative treatments.

Latent Dirichlet allocation optimization yielded, based on the coherence score, optimal parameter values for 48 topics (Figure 2).

The LDA model results were visualized using PyLDAVis and t-SNE (Figure 3) and analyzed by two independent researchers.

The analysis and labeling process returned 13 topics regarding vaccine hesitancy (Table 2). We achieved a kappa statistic of 0.85, indicating almost perfect agreement between the 2 raters.[36]

The topics identified through the literature and topic-modeling were unified into a single list that represented vaccine hesitancy and anti-vaxxers' perceptions. Table 3 shows a list of 11 high-level topics, examples of topics from the literature and topic-modeling results, and short topic descriptions.

Topic analysis

When labeling the tweets for training the ReadMe algorithm, after several iterations and enhancements in the assigned labels, we achieved a kappa statistic of 0.80, representing substantial agreement among the raters.[36] The trends in the volume of tweets for each topic are shown in Figure 4. Overall, 35% of the total tweets referenced developmental disabilities, followed by "government and politics" (21%) and "conspiracy theory" (13%), respectively. The remaining topics represented less than 10% of the total tweets. "Developmental disabilities" have been

Table 1. Anti-vaccine topics from the literature.

Topics	Example references
Safety	9,19,39
Perceived risks and/or deaths associated with vaccines	12,19,40
Perceived severity	41
Developmental disabilities	1,6,42
Side-effects	6,40
Perceived susceptibility	41
Chemicals/Non-natural vaccine ingredients	6,19
Effectiveness	9
Efficacy	6
Distrust of government, pharmaceutical companies, scientists, and organizations that support vaccination efforts	12,18,19,40
Distrust of government/industries	6,19
Evil government	18,43
Conspiracy theories	39,44
Alternative treatments	12
Natural cures, immune system	12,18
Vaccines are unnecessary	6
Religion	6,19
Ethics	6
School and family	19

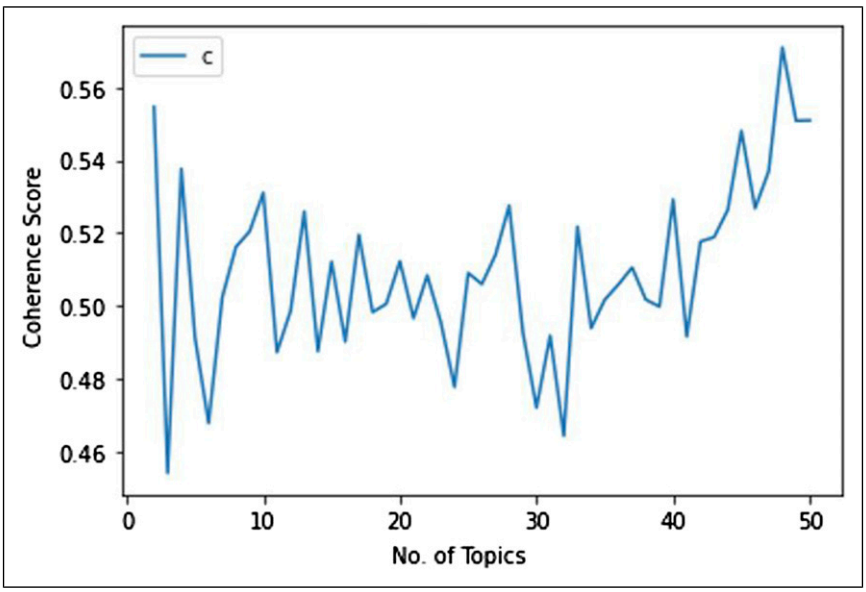

Figure 2. Optimal number of topics based on coherence score.

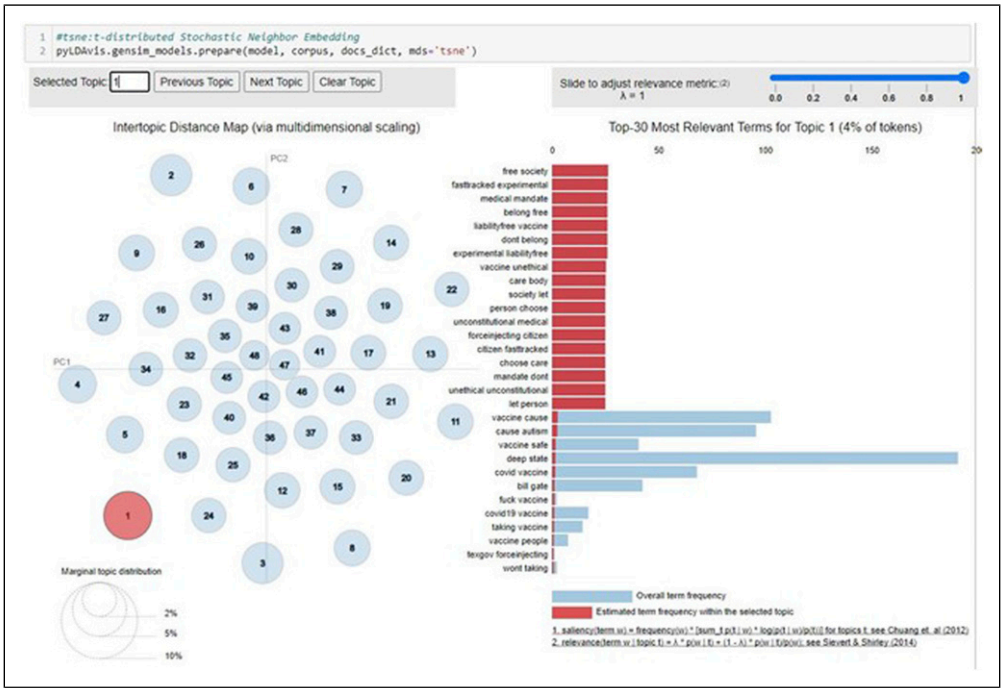

Figure 3. Topic visualization and analysis through PyLDAVis using t-distributed stochastic neighbor embedding.

Table 2. Anti-vaccine topics identified through topic-mining.

Topic	Topic name
1	Big pharma
2	Deep state
3	Vaccines contain microchips
4	Vaccines contain nano-particles
5	Population control
6	Freedom of choice
7	Mistrust of vaccines
8	Vaccines kill people
9	Long-term effects of vaccines
10	Vaccines alter DNA
11	Vaccines contain mercury, toxic chemicals, and heavy metals
12	Vaccines cause autism
13	Vaccines do not work

consistently discussed over the years; however, from March 2020 there was a decrease in the number of such tweets and an increase in the number of tweets discussing "government and policies" and "conspiracy theories." Appendix B shows a list of categories and example tweets classified by the ReadMe algorithm.

Table 3. Combination of anti-vaccine topics identified through literature and topic-mining.

Topic #	Topic	Examples	Description
1	Side-effects	Perceived risks; alleged side-effects and/or deaths caused by vaccines; perceived severity; safety; perceived susceptibility; vaccines kill people	Belief that vaccines can cause severe side-effects and risk serious complications and death
2	Chemicals/non-natural	Chemicals/non-natural; vaccine ingredients; vaccines contain nano-particles; vaccines contain mercury, toxic chemicals, and heavy metals	Belief that vaccines include as immunologic adjuvants chemical and non-natural ingredients such as aluminum compounds, mercury, and metal
3	Developmental disabilities	Long-term effects of vaccines; vaccines cause autism; developmental disabilities	Belief that vaccines can have long-term effects, and lead to developmental disabilities such as autism and brain injuries
4	Effectiveness and Efficacy	Efficacy and effectiveness; mistrust in vaccines; vaccines do not work	Belief that vaccines are not effective and do not protect against diseases
5	Nature is better	Natural cures; immune system	Belief in nature and trust in the body's immune system as the best means of securing protection and achieving herd immunity
6	Alternative treatments	Alternative treatments; vaccines are unnecessary	Use of alternative treatments, therapeutics, and vitamins for treating health-care conditions
7	Government and policies	Mistrust of governments and organizations that support vaccination efforts; evil government	Mistrust in government and politicians promoting vaccination
8	Pharma industry	Mistrust of pharmaceutical companies, scientists, and industries that support vaccination efforts; big pharma	Mistrust in pharmaceutical companies, as they are viewed as being mainly motivated by profit
9	Religion/ethics	Religion; ethics	Belief that vaccines are not ethical and against religion, as the human body is created as it should be, and any external interference is prohibited[6]
10	Civil liberties/ freedom	Freedom of choice	Advocation of freedom of choice and opposition to vaccine mandates
11	Conspiracy theory	Deep state; conspiracy theories; vaccines contain microchips; population control; vaccines alter DNA	Belief that someone influential is responsible for current events

Figure 5 shows a comparison of the percentages of tweets for each topic before and after 1 February 2020. There were 196,200 anti-vaccine tweets before 1 February 2020, and 90,092 tweets afterwards. All but three topics ("government and politics," "conspiracy theories," and "developmental disabilities") showed similar percentages of tweets for before and after 1st February 2020. After 1st February 2020, "government and politics" and "conspiracy theories" showed increased percentages, while "developmental disabilities" showed a decreased percentage. Appendix C shows the changes in the percentages of tweets concerning flu, MMR, and all vaccines,

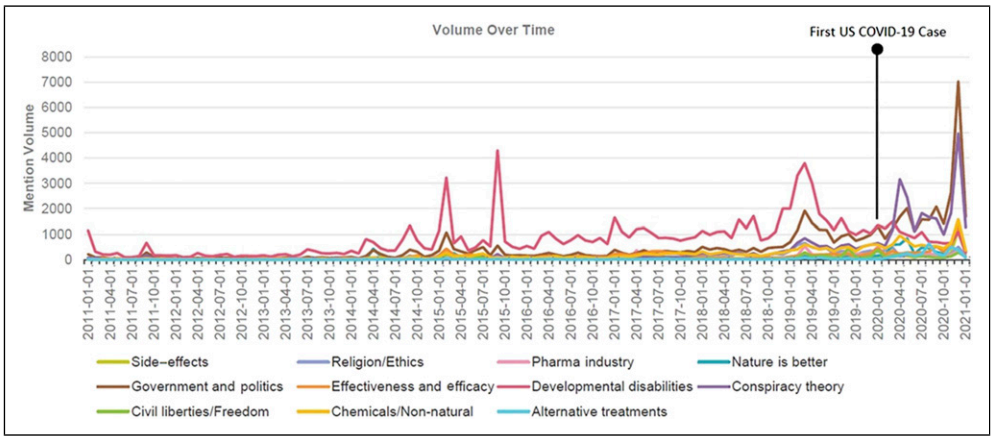

Figure 4. Volume of tweets across topics between 1st January 2011, and 15th January 2021.

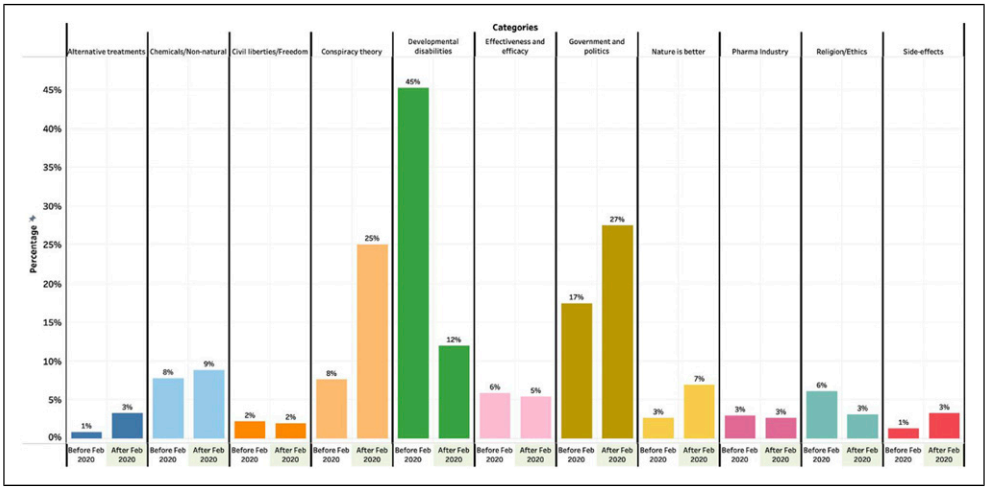

Figure 5. Percentages of tweets for each topic before/after 1st February 2020.

respectively, by topic over time. Most topics show, for all three vaccine types, a similar pattern; this indicates that the COVID-19 outbreak did not significantly impact overall trends in flu and MMR discourse.

Filtering by vaccine, we identified 13,189, 7,225, and 4371 tweets challenging COVID-19, MMR, and flu vaccines, respectively. "Side-effects," "pharma industry," and "civil rights/freedom" showed similar percentages of tweets across the three vaccines (Figure 6). However, for "nature is better," "government and politics," "conspiracy theories," and "alternative treatments" the highest percentages were for COVID-19, followed by flu and MMR, respectively. Additionally, for "effectiveness and efficiency" and "chemical/non-natural" the highest percentages were for flu, followed by MMR and COVID-19, respectively. Finally, for "developmental disabilities" the highest percentage was for MMR, followed by flu and COVID-19, respectively.

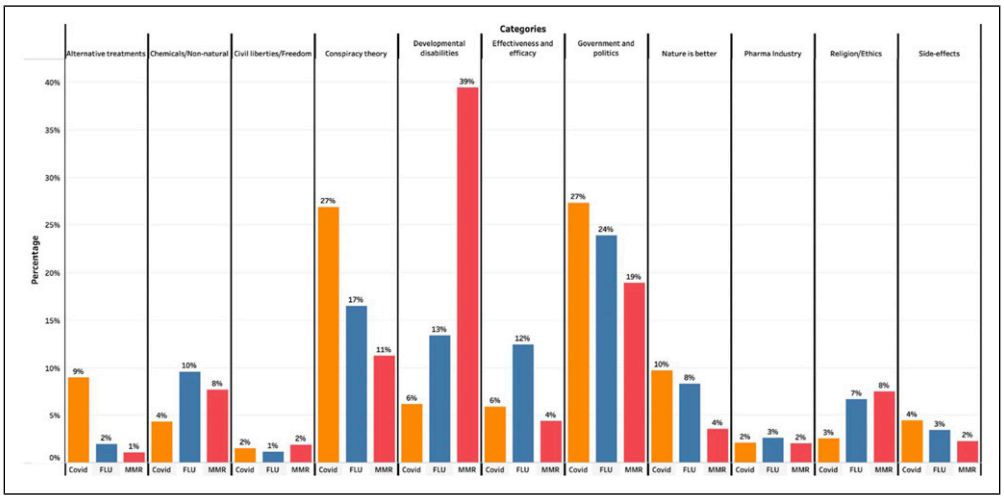

Figure 6. Percentages of tweets regarding flu, MMR, and COVID-19 vaccines between 1st January 2011 and 15th January 2021, across different topics. MMR: Measles, mumps, and rubella.

Appendix D shows the distribution of tweets for each year across different topics. The distribution was similar across all vaccines except those for flu in 2013, at which time the percentages for flu-related tweets regarding "effectiveness and efficacy" increased significantly and decreased for "government and policies". The percentages of tweets concerning "developmental disabilities" remained generally consistent from 2011 to 2019. From 2020 (marking the introduction of COVID-19), the percentage of tweets regarding "government and politics" and "conspiracy theories" increased.

Table 4 shows the number of anti-vaccine tweets from January 2020 to January 2021 that mention both flu and COVID-19. The number of such tweets was low when compared to the total number of anti-vaccine tweets for the same time frame.

Discussion

This study's findings support the literature[45–49] on vaccine hesitancy and the assertion that, despite advancements in vaccine development and the demonstrated efficacy and safety of vaccines, anti-vaccine sentiment remains an important issue. We identified several topics associated with vaccine hesitancy. These topics generally accord with those mentioned in prior research, such as safety/risk and politics,[9,20,21] efficacy/effectiveness,[20,21] governments/big companies,[9,21] immune system,[21] vaccine ingredients and religion,[9] conspiracy theories,[12] side-effects,[6,41] developmental disabilities,[1,43] nature is better and alternative treatments,[5,19] religion/ethics,[6,9] and civil liberties/freedom.[50]

During the COVID-19 Pandemic, of the anti-vaccine tweets concerning the topics "government and politics," "conspiracy theories," and "nature is better," the highest percentages related to COVID-19 vaccines when compared to flu and MMR vaccines. This is not surprising. In the US, the debate on vaccinations has long been highly politicized.[51] In 2015, health-care specialists condemned republican candidates for advancing inaccurate perceptions of vaccines.[52] Furthermore, the

Table 4. Anti-vaccine tweets regarding COVID-19 and flu from Jan 2020 to Jan 2021.

Date	Number of tweets	Tweets mentioning both COVID-19 & flu	Percentages
Jan-20	352	9	2.6%
Feb-20	374	21	5.6%
Mar-20	812	78	9.6%
Apr-20	1156	63	5.4%
May-20	993	79	8.0%
Jun-20	625	25	4.0%
Jul-20	1047	74	7.1%
Aug-20	1233	66	5.4%
Sep-20	1115	77	6.9%
Oct-20	871	53	6.1%
Nov-20	1463	77	5.3%
Dec-20	3373	243	7.2%
Jan-21	852	38	4.5%

present results support existing findings that attitudes toward vaccines are influenced by political beliefs[45] and public mistrust of governments' pandemic responses.[53,54]

Tweets promoting conspiracy theories[39,55] and misinformation[56] are considered a threat to public vaccine acceptance.[55] Many vaccine-related conspiracy theories have emerged during the COVID-19 Pandemic. For example, some believe that COVID-19 vaccines were developed to implant nano-chips in people's bodies so that people can be controlled through 5G technology,[46] that vaccines are a tool by which governments can gain political control,[54] and that mRNA-based COVID-19 vaccines can permanently change human DNA.[47]

Our results also revealed relatively high percentages of "alternative treatments" and "nature is better" tweets regarding COVID-19 vaccines when compared to flu and MMR vaccines. Several factors may explain this. For example, until December 2020 there were no approved COVID-19 vaccines or treatments, and people began considering natural and traditional medicines with known safety profiles.[48] Regarding the increase in "nature is better" tweets, some anti-vaxxers have disseminated a "natural immunity" theory[49] that suggests that the human body can treat itself without vaccines.

The percentage of tweets on developmental disabilities fell after February 2020. This is consistent with the analysis of the distribution of tweets across the different vaccines, which showed that the percentage of tweets on developmental disabilities was highest for MMR when compared to flu and COVID-19. The high level of discussion on developmental disabilities before February 2020 also aligns with the literature. Such popularity is mainly due to discussions on the side-effects of MMR vaccines and the risk of developmental disabilities such as autism[1,49] and epilepsy.[1] A link between MMR and disabilities was considered the main social-media topic for anti-vaxxers prior to COVID-19.[42]

The present results highlight the need for government and health-care agencies to increase transparency in policy development and decision-making before and after the introduction of COVID-19 vaccines. There is also a need to provide updated information to the public on how vaccines are developed and tested before they are administered to the general population.[57] According to Kennedy,[58] vaccine hesitancy and political populism are driven by similar dynamics, characterized by public mistrust of politicians and experts. Thus, it is necessary to build trust between these parties and anti-vaxxers and address the issues underlying vaccine hesitancy.

Limitations and future work

While this study emphasized changes in percentages of tweets on certain topics over time and across vaccines, there is a need for a separate analysis of the number of unique Twitter users post across different topics. Such analysis could clarify whether changes in volume are related to changes in the number of active users or in the numbers of tweets from users. Further, analyzing Twitter threads that comprise a main tweet and replies could clarify the structure of vaccine-related discourse on Twitter. This study examined tweets from US users; future studies could compare our findings with those for users of other nationalities. Further, research could investigate the impact the 2020 US Presidential Election and the associated increase in political discourse on Twitter had on COVID-19-vaccine skepticism, as well as the impact of other controversial issues. Such controversial issues must be identified and understanding of their impact on tweeting activity improved.

Conclusion

In this study, we sought to understand the drivers of hesitancy towards COVID-19 vaccines in the US. We compared tweet-based topics along two dimensions: before and after the onset of COVID-19, and across 3 vaccines (flu, MMR, and COVID-19). Using text analytics, we identified trending themes and topics of public concern regarding vaccine hesitancy in general and COVID-19 vaccines in particular.

Overall, the threat of the COVID-19 Pandemic did not cause the anti-vaccine discussion to shrink, but to shift, with the prevalence of "conspiracy theory," "government and policies," and "alternative treatments"/"nature is better" tweets increasing. Discussion of vaccines' effect on developmental disabilities reduced after the outbreak. The variations between before and after the COVID-19 outbreak regarding the main reasons people oppose vaccines or become vaccine-hesitant show that vaccine hesitancy is context-, time-, place-, and vaccine-specific,[54] and that it rises in prevalence when a pandemic occurs.[59] Such variation could be attributed to pivotal events such as pandemic outbreaks, government responses, and related discourse on social-media platforms. Moreover, the present findings support the theory that vaccine hesitancy has various causes based on several factors. These factors can be grouped into environmental/external, agent/vaccine-specific, and host-specific.[57]

The significance and implications of this research transcend the COVID-19 Pandemic by demonstrating the importance of social-media mining and its potential for supporting public-health-related policies and decisions. Government officials and decision-makers could tailor and fine-tune public awareness campaigns and prioritize policy interventions to increase vaccine acceptance.

Acknowledgements

The authors would like to thank Dakota State University for generously supporting this research by providing access to the BrandWatch platform.

Author Contributions

All authors have contributed to the writing of this research paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data Availability Statement

The data used for this research were obtained using Brandwatch, a social-media data collection and analytics tool. Data are available within Brandwatch and can be obtained by accessing the tool and running the query presented in Appendix A.

ORCID iDs

Tareq Nasralah https://orcid.org/0000-0002-2336-7601
Ahmed Elnoshokaty https://orcid.org/0000-0003-3844-3985
Mohammad Al-Ramahi https://orcid.org/0000-0002-8170-8312
Abdullah Wahbeh https://orcid.org/0000-0002-6894-0192

Supplemental Material

Supplemental material for this article is available online.

References

1. Dhaliwal D and Mannion CA. Anti-vaccine messages on Facebook: a preliminary audit. *JMIR Public Health Surveill* 2020; 6: 10.
2. Wilson SL and Wiysonge C. Social media and vaccine hesitancy. *BMJ Glob Health* 2020; 5: e004206.
3. Puri N, Coomes EA, Haghbayan H, et al. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Hum Vaccin Immunother* 2020; 16: 2586–2593.
4. Deiner MS, Fathy C, Kim J, et al. Facebook and Twitter vaccine sentiment in response to measles outbreaks. *Health Inform J* 2019; 25: 1116–1132.
5. Leung CK, Daniel Mai TH, Thong Tran ND, et al. Predictive analytics to support health informatics on COVID-19 data. In: 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE). 2021, pp. 1–9.
6. UNICEF. *Tracking anti vaccination sentiment in Eastern European social media networks*. New York: UNICEF. https://www.unicef.org/eca/media/1556/file/Tracking-anti-vaccination-sentiment-in-astern-European-social-media-networks.pdf (2013, accessed 14 December 2021).
7. Ruiz J, Featherstone J and Barnett G. Identifying vaccine-hesitant communities on Twitter and their geolocations: a network approach. In: Proceedings of the 54th Hawaii international conference on system sciences, pp. 3964–3969.
8. Chou W-YS and Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health Commun* 2020; 35: 1718–1722.
9. Lyu H, Wang J, Wu W, et al. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. ArXiv201202165 Cs. http://arxiv.org/abs/2012.02165 (2020, accessed 30 December 2020).
10. Guillaume L and Bath PA. A content analysis of mass media sources in relation to the MMR vaccine scare. *Health Inform J* 2008; 14: 323–334.
11. Guillaume LR and Bath PA. The impact of health scares on parents' information needs and preferred information sources: a case study of the MMR vaccine scare. *Health Inform J* 2004; 10: 5–22.
12. Wu W, Lyu H and Luo J. Characterizing discourse about COVID-19 vaccines: a Reddit version of the pandemic story. ArXiv210106321 Cs. http://arxiv.org/abs/2101.06321 (2021, accessed 12 March 2021).

13. Larson HJ, Cooper LZ, Eskola J, et al. Addressing the vaccine confidence gap. *The Lancet* 2011; 378: 526–535.
14. Kang GJ, Ewing-Nelson SR, Mackey L, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine* 2017; 35: 3621–3638.
15. Smith TC and Reiss DR. Digging the rabbit hole, COVID-19 edition: anti-vaccine themes and the discourse around COVID-19. *Microbes Infect* 2020; 22: 608–610.
16. Gunaratne K, Coomes EA and Haghbayan H. Temporal trends in anti-vaccine discourse on Twitter. *Vaccine* 2019; 37: 4867–4871.
17. Massey PM, Kearney MD, Hauer MK, et al. Dimensions of misinformation about the HPV vaccine on Instagram: content and network analysis of social media characteristics. *J Med Internet Res* 2020; 22: e21451.
18. Jamison AM, Broniatowski DA, Dredze M, et al. Not just conspiracy theories: vaccine opponents and proponents add to the COVID-19 'infodemic' on Twitter. *Harv Kennedy Sch Misinformation Rev* 2020; 1. Epub ahead of print 9 September 2020. DOI: 10.37016/mr-2020-38.
19. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, et al. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 Pandemic. *J Commun Healthc* 2021; 14: 12–19.
20. Quintana IO, Klein C, Cheong M, et al. The evolution of vaccine discourse on Twitter during the first six months of COVID-19. https://philpapers.org/rec/QUITEO-12 (2021, accessed 11 March 2021).
21. AJMC. A. Timeline of COVID-19 developments in 2020. American Journal of Managed Care. https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020 (2021, accessed 28 January 2021).
22. Burzyńska J, Bartosiewicz A and Rękas M. The social life of COVID-19: early insights from social media monitoring data collected in Poland. *Health Inform J* 2020; 26: 3056–3065.
23. Zhao Y, Guo Y, He X, et al. Assessing mental health signals among sexual and gender minorities using Twitter data. *Health Inform J* 2020; 26: 765–786.
24. Apuke OD and Omar B. Social media affordances and information abundance: enabling fake news sharing during the COVID-19 health crisis. *Health Inform J* 2021; 27: 14604582211021470.
25. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. *Health Inform J* 2016; 22: 470–478.
26. Cavnar W and Trenkle JM. N-gram-based text categorization. In: Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval. Las Vegas, NV, 1994, pp. 161–175.
27. Blei DM, Ng AY and Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
28. Al-Ramahi MA, Liu J and El-Gayar OF. Discovering design principles for health behavioral change support systems: a text mining approach. *ACM Trans Manag Inf Syst TMIS* 2017; 8: 1–24.
29. Sievert C and Shirley K. LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces. Baltimore, MD, USA: Stroudsburg, PA: Association for Computational Linguistics, pp. 63–70.
30. Laurens M and Geoffrey H. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9: 2579–2605.
31. Mimno D and Blei D. Bayesian checking for topic models. In: Proceedings of the 2011 conference on empirical methods in natural language processing. Edinburgh, Scotland, UKStroudsburg, PA: Association for Computational Linguistics, pp. 227–237.
32. Syed S and Spruit M. Full-Text or abstract? Examining topic coherence scores using latent dirichlet allocation. IEEE International Conference on Data Science and Advanced Analytics (DSAA). Tokyo, Japan: IEEE, 2017, pp. 165–174.
33. Röder M, Both A and Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. New York, NY, USA: Association for Computing Machinery, pp. 399–408.

34. Stevens K, Kegelmeyer P, Andrzejewski D, et al. Exploring topic coherence over many models and many topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Korea: Jeju IslandAssociation for Computational Linguistics, pp. 952–961.
35. Zhao W, Chen JJ, Perkins R, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics* 2015; 16: S8.
36. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
37. Hopkins DJ and King G. A method of automated nonparametric content analysis for social science. *Am J Polit Sci* 2010; 54: 229–247.
38. Simm W, Ferrario M-A, Piao S, et al. Classification of short text comments by sentiment and actionability for VoiceYourView. IEEE Second International Conference on Social Computing. Minneapolis, MN, USA: IEEE, 2010, pp. 552–557.
39. Dunn AG, Surian D, Leask J, et al. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine* 2017; 35: 3033–3040.
40. Blankenship EB, Goff ME, Yin J, et al. Sentiment, contents, and retweets: a study of two vaccine-related Twitter datasets. *Perm J* 2018; 22: 17–138.
41. Du J, Luo C, Shegog R, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. *JAMA Netw Open* 2020; 3: e2022025.
42. Featherstone JD, Barnett GA, Ruiz JB, et al. Exploring childhood anti-vaccine and pro-vaccine communities on Twitter – A perspective from influential users. *Online Soc Netw Media* 2020; 20: 100105.
43. Mitra T, Counts S and Pennebaker JW. Understanding anti-vaccination attitudes in social media. Proceedings of the international AAAI conference on web and social media. Cologne, Germany, 2016, pp. 269–278.
44. Ahmed W, Vidal-Alaball J, Downing J, et al. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *J Med Internet Res* 2020; 22: e19458.
45. Peretti-Watel P, Seror V, Cortaredona S, et al. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *Lancet Infect Dis* 2020; 20: 769–770.
46. Khan YH, Mallhi TH, Alotaibi NH, et al. Threat of COVID-19 vaccine hesitancy in Pakistan: the need for measures to neutralize misleading narratives. *Am J Trop Med Hyg* 2020; 103: 603–604.
47. Reuters. Fact check: COVID-19 vaccines won't alter recipient DNA; frontline workers have suffered directly from the virus. Reuters, 18 December 2020, https://www.reuters.com/article/uk-factcheck-viral-post/fact-check-covid-19-vaccines-wont-alter-recipient-dna-frontline-workers-have-suffered-directly-from-the-virus-idUSKBN28S2V1 (18 December 2020, accessed February 2020).
48. Wang D, Huang J, Yeung AWK, et al. The significance of natural product derivatives and traditional medicine for COVID-19. *Processes* 2020; 8: 937.
49. Butler K. Anti-Vaxxers have a dangerous theory called "natural immunity." Now it's going mainstream. Mother Jones, 12 May, https://www.motherjones.com/politics/2020/05/anti-vaxxers-have-a-dangerous-theory-called-natural-immunity-now-its-going-mainstream/ (2020, accessed 12 May 2020).
50. Hotez PJ. COVID19 meets the antivaccine movement. *Microbes Infect* 2020; 22: 162–164.
51. Khubchandani J, Sharma S, Price JH, et al. COVID-19 vaccination hesitancy in the United States: A rapid national assessment. *J Community Health* 2021; 44(1): 19–28. Epub ahead of print 3 January 2021. DOI: 10.1007/s10900-020-00958-x.
52. Dyer O. Republican candidates cast doubt on vaccines in US presidential debate. *BMJ* 2015; 351: h5006.
53. Schaffer DeRoo S, Pudalov NJ and Fu LY. Planning for a COVID-19 vaccination program. *JAMA* 2020; 323: 2458–2459.

54. Freeman D, Loe BS, Chadwick A, et al. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychol Med* 2020; 11: 1–15. DOI: 10.1017/S0033291720005188.
55. Romer D and Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. *Soc Sci Med* 2020; 263: 113356.
56. Wahbeh A, Nasralah T, Al-Ramahi M, et al. Mining physicians' opinions on social media to obtain insights into COVID-19: mixed methods analysis. *JMIR Public Health Surveill* 2020; 6: 2369–2960.
57. Kumar D, Chandra R, Mathur M, et al. Vaccine hesitancy: understanding better to address better. *Isr J Health Pol Res* 2016; 5: 2.
58. Kennedy J. Populist politics and vaccine hesitancy in Western Europe: an analysis of national-level data. *Eur J Public Health* 2019; 29: 512–516.
59. Williams L, Gallant AJ, Rasmussen S, et al. Towards intervention development to increase the uptake of COVID-19 vaccination among those at high risk: outlining evidence-based and theoretically informed future intervention content. *Br J Health Psychol* 2020; 25: 1039–1054.