Content Themes and Influential Voices Within Vaccine Opposition on Twitter, 2019

Erika Bonnevie, MA, Jacyln Goldbarg, MPH, Allison K. Gallegos-Jeffrey, MPH, Sarah D. Rosenberg, MPH, Ellen Wartella, PhD, and Joe Smyser, PhD

**Objectives.** To report on vaccine opposition and misinformation promoted on Twitter, highlighting Twitter accounts that drive conversation.

**Methods.** We used supervised machine learning to code all Twitter posts. We first identified codes and themes manually by using a grounded theoretical approach and then applied them to the full data set algorithmically. We identified the top 50 authors month-over-month to determine influential sources of information related to vaccine opposition.

**Results.** The data collection period was June 1 to December 1, 2019, resulting in 356,594 mentions of vaccine opposition. A total of 129 Twitter authors met the qualification of a top author in at least 1 month. Top authors were responsible for 59.5% of vaccine-opposition messages. We identified 10 conversation themes. Themes were similarly distributed across top authors and all other authors mentioning vaccine opposition. Top authors appeared to be highly coordinated in their promotion of misinformation within themes.

**Conclusions.** Public health has struggled to respond to vaccine misinformation. Results indicate that sources of vaccine misinformation are not as heterogeneous or distributed as it may first appear given the volume of messages. There are identifiable upstream sources of misinformation, which may aid in countermessaging and public health surveillance. ([Am J Public Health. 2020;110:S326–S330. https://doi.org/10.2105/AJPH.2020.305901](https://doi.org/10.2105/AJPH.2020.305901))

See also Chou and Gaysynsky, p. S270, and Walsh-Buhi, p. S292.

Vaccine opposition is a threat to global health, with digital and social media a primary source of misinformation and means of organizing vaccine opposition. Misinformation has reached a critical level, with provaccine and vaccine-opposing communities increasingly polarized. "Anti" messaging is increasing in communities that appear to be largely unaffected by traditional health promotion strategies and scientific information. In 2000, measles was declared eradicated in the United States as the result of an effective vaccination campaign; however, in 2019, the Centers for Disease Control and Prevention announced 1282 confirmed cases of measles, the highest since 1992.

Vaccine opposition also has policy implications: dozens of state bills have attempted to supplant established population health practice by prioritizing personal liberties and appealing to ideology, rather than evidence. Misinformation erodes trust in science and public health authorities and is associated with a decrease in vaccination rates, risking further outbreaks and cases of vaccine-preventable disease. There are economic implications as well: treating measles outbreaks costs approximately $32,000 per case, and, in 2017, the reported cost to treat 1 child’s case of tetanus was more than $800,000. Despite the established and evolving threat to public health that vaccine opposition poses, there has been no systematic, sustained effort to identify, track, and routinely report on it in the United States.

In 2019, public health nonprofit The Public Good Projects commenced Project VCTR (Vaccine Communication Tracking and Response) to identify and track vaccine-related communication on digital and social media. This study examines discourse on Twitter, given that the platform is a primary source of online vaccine misinformation. The aims of this study were to (1) determine the volume of conversation around vaccine opposition, (2) explore specific themes in conversation regarding vaccine opposition with a focus on vaccine-related misinformation, and (3) identify accounts that are drivers of vaccine opposition. We compared content themes employed by influential vaccine opposition accounts with general themes in vaccine-opposition discourse to identify message frames top authors use to drive conversation.

**METHODS**

We obtained data through a partnership with a media monitoring platform that collected 100% of publicly available Twitter tweets and retweets containing keywords identified by The Public GoodProjects. The initial data collection process was based on a lengthy keyword search query using English-language Boolean operators to identify information related to vaccination.

**ABOUT THE AUTHORS**

Erika Bonnevie, Jacyln Goldbarg, Allison K. Gallegos-Jeffrey, Sarah D. Rosenberg, and Joe Smyser were with The Public Good Projects, Alexandria, VA, at the time the work was conducted. Ellen Wartella is with The Northwestern School of Communication, Evanston, IL.

Correspondence should be sent to Erika Bonnevie, The Public Good Projects, 2308 Mount Vernon Ave, Suite 758, Alexandria, VA 22301 (e-mail: erika.bonnevie@publicgoodprojects.org). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This article was accepted July 29, 2020.

https://doi.org/10.2105/AJPH.2020.305901
conversation on Twitter in the United States from June 1 to December 1, 2019. Keywords were selected based on a review of previously published scientific, gray, and white literature (Appendix A, available as a supplement to the online version of this article at http://www.ajph.org) and deductive determinations based on familiarity with online vaccine conversation.

Initial data collection followed 2 processes: keywords could either be “standalone” or “co-occurring.” Standalone keywords function so that any mention of a specific word would collect that post. The initial query consisted of 129 standalone words and 129 hashtag equivalents. Terms could also be co-occurring, meaning that a post was collected if 2 terms were present. Shortened forms of “vaccination” were collected if they also included a health condition treated by vaccines or terms referenced in vaccine discourse. The co-occurring search query consisted of 333 health condition– or vaccine-related words and hashtag equivalents, paired with 3 shorthand vaccine terms and hashtag equivalents. We employed 60 exclusion terms to exclude content related to animal vaccinations or medication instructions. Keywords can be found in Appendix B (available as a supplement to the online version of this article at http://www.ajph.org).

Identifying Vaccine Opposition
We gathered Twitter data continuously throughout the data collection period. With data collection ongoing, we selected a random sample of 1000 tweets from the total sample of vaccine-related conversation (0.9% of the data collected at the time, in line with research conducting similar analyses13) and manually coded to identify messages in opposition to vaccines or terms referenced in vaccine discourse. The co-occurring search query consisted of 333 health condition– or vaccine-related words and hashtag equivalents, paired with 3 shorthand vaccine terms and hashtag equivalents. We employed 60 exclusion terms to exclude content related to animal vaccinations or medication instructions. Keywords can be found in Appendix B (available as a supplement to the online version of this article at http://www.ajph.org).

not considered vaccine opposition. These posts generated an additional list of keywords specific to vaccine opposition, which were then added to the full keyword query that generated the total sample, allowing for messages to be identified and analyzed separately as vaccine opposing. All analysis in this study was conducted on posts containing terms related to vaccine opposition.

Theme Generation
We then categorized vaccine-opposing posts into themes. Using a 5-step interpretive process, 2 coders (E. B. and S. D. R.) manually coded 1000 randomly selected posts (step 3, Figure 1). Approximately 200 posts were cross-coded between analysts. Discrepancies were re-examined until reaching agreement on more than 90% of posts. Themes were created, compared, and combined until data saturation was achieved, defined as a theme comprising less than 1% of conversation. For this study, 74.8% of data pertaining to vaccine opposition were coded into a theme. Each theme was assigned its own unique list of keywords that identified a post as having met the criterion of that theme. To test the validity of each theme’s keywords, keywords were turned into queries, as described previously. We reviewed 100 randomly selected posts automatically categorized to each theme. If 90% of automatically categorized posts were accurately coded, that theme’s keyword query was approved and applied to the total sample. Theme definitions and sample keywords can be found in Appendix C (available as a supplement to the online version of this article at http://www.ajph.org). Applications of supervised automatic coding for qualitative analysis have been explored as a practical way of applying lessons from big data sets to public health.13,15

To identify misinformation promoted within themes around vaccine opposition, analysts reviewed 200 posts receiving the most engagement within each theme (step 4, Figure 1). Misinformation was organized into categories, with each category defined by unique keywords. These keywords allowed all posts within the theme to be automatically tagged if they contained a category of misinformation. Analysts manually verified the top 200 posts within each misinformation category, and keywords were amended to ensure that at least 90% of posts were tagged with the correct misinformation category.

The operational definition of “vaccine misinformation” was considered any information that contrasted with the Centers for Disease Control and Prevention’s Immunization Safety Office.16

Top Authors
As with other studies examining Twitter data for vaccine-related information,17,18 this study made use of metadata accompanying
posts to perform social network analyses. We sorted accounts publishing messages by the number of engagements received to determine which accounts had the most influence in the vaccine opposition conversation (termed “top authors”). Engagement was defined as a like, comment, or share of a post. Analysts identified the top 50 authors each month. Defining engagement in this way allowed for discovery of accounts with the most frequent interactions, specific posts receiving the most interactions, and themes most commonly employed across these posts. Previous research has also examined the top 50 Twitter authors as a way of measuring trends.19

Top authors were manually examined to ensure they were promoting vaccine opposition, versus mocking or reporting on vaccine opposition. Results compare conversation from top authors with overall vaccine opposition conversation with top authors removed (“top authors” vs “non–top authors”). We used the $\chi^2$ test to determine statistically significant differences between top authors and non–top authors for each theme.

**RESULTS**

From June 1 to December 1, 2019, we collected 356 594 Twitter posts mentioning vaccine opposition. We identified 129 unique Twitter accounts as top authors within at least 1 of the 6 months, generating 212 018 total engagements and 772.9 million potential impressions (the number of followers of the original author plus the followers of individuals who shared their content). Of those 129 accounts, 15 were top authors for at least 5 months, during which time they generated 124 243 engagements, which was 58.6% of the 212 019 engagements with top authors’ content.

We identified 10 themes within posts about vaccine opposition, with the top 5 themes each comprising over 10% of mentions. (Table 1):

Negative health impacts were shown in 55.4% of mentions from top authors and 49.2% of the general opposition (non–top authors). Within this theme, misinformation around deaths attributable to vaccines and vaccine–caused autism was present in 66.5% and 43.8% of top author posts, respectively. Within general opposition, deaths were mentioned in 14.5% of posts and autism in 26.3%. Across references to death, top authors predominantly shared a journal article citing deaths reported to the Vaccine Adverse Event Reporting System from 1997 to 2013 to claim vaccines cause child death.20 Other misinformation related to health impacts included associations between vaccines and paralysis (5.9% top authors; 0.5% general opposition) and seizures (5.7% top authors; 0.8% general opposition).

Pharmaceutical industry mentions appeared in 16.9% of posts from top authors and 18.9% of general opposition. Vaccines were most often framed as a conspiracy by “Big Pharma” to increase sales revenue. Merck was referenced in 58.1% of posts from top authors, compared with 38.7% by general opposition, because of its manufacturing of the Gardasil vaccine.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Top Authors: Vaccine Opposition, $a$, $b$ % (No.)</th>
<th>Not Top Authors, $a$, $b$ % (No.)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative health impacts attributed to vaccination</td>
<td>55.4 (117 530)</td>
<td>49.2 (71 167)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pharmaceutical industry</td>
<td>16.9 (35 821)</td>
<td>18.9 (27 346)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Research and clinical trials</td>
<td>15.5 (32 819)</td>
<td>5.6 (8 097)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Policies and politics</td>
<td>15.0 (31 723)</td>
<td>17.7 (25 621)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vaccine ingredients</td>
<td>13.8 (29 281)</td>
<td>17.2 (24 858)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Family</td>
<td>7.3 (15 508)</td>
<td>7.4 (10 628)</td>
<td>.68</td>
</tr>
<tr>
<td>Disease prevalence and outbreaks</td>
<td>5.1 (10 885)</td>
<td>3.2 (4 579)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>School</td>
<td>3.6 (7 733)</td>
<td>2.8 (3 997)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Religion</td>
<td>3.2 (6 884)</td>
<td>2.3 (3 997)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Natural alternatives</td>
<td>0.9 (1 953)</td>
<td>1.6 (2 287)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$^a$Calculated out of all posts and engagements on top authors’ content (n = 212 018 out of 356 594 total posts).

$^b$Percentages may add to more than 100%, as one post can be coded across multiple themes.

$^c$Calculated out of all posts and engagements related to vaccination opposition or hesitancy, minus top authors (n = 144 576 out of 356 594 total posts).
Vaccine research was found within 15.5% of posts from top authors and 5.6% of general opposition. Posts most often criticized vaccine research or institutions conducting research or promoted pseudoscience as fact. The most frequently referenced studies were related to the human papillomavirus vaccine and its association with negative health impacts after vaccination. A commonly shared article was retracted in 2019 and is now found on vaccine opposition Web sites (29.0% top authors; 8.9% general opposition). This was followed by research about the influenza vaccine, highlighting studies showing associations with other respiratory infections, renal failure, and suppressed immune responses (20.0% top authors; 7.3% general opposition).

Five of the identified themes amounted to approximately 7% or less of the total conversation:

- Disease prevalence focused on measles outbreaks, with 83.2% of top author posts and 17.4% of general opposition posts mentioning measles or the measles, mumps, and rubella vaccine. Vaccine opponents frequently cited stories about vaccine-driven epidemics, such as the vaccine-derived poliovirus, to suggest the dangers of vaccines (19.5% top authors; 3.8% general opposition).
- Family members typically included mention of individuals who believe they have experienced negative health impacts attributable to vaccination, often from a parent sharing a vaccine adverse health event of their child (70.7% top authors; 57.2% of general opposition).
- School conversation focused on policies related to mandatory vaccinations for enrollment.
- Religion included references to any religion and most often discussed religious exemptions to mandatory vaccines (46.7% top authors; 44.5% general opposition).
- Natural alternatives to vaccines included misinformation about the use of homeopathic alternatives to vaccination and “vaccine detox.”

**DISCUSSION**

This study showed that major talking points used by vaccine opponents originated from a handful of accounts. A total of 129 accounts on Twitter appeared to be driving more than half of all conversation regarding vaccine opposition, and 15 accounts appeared hyperinfluential, generating a majority of engagements on top authors’ posts. When top authors’ posts were compared with other posts, misinformation themes were similar. While there were statistically significant differences in the proportions of most themes, this may have been attributable to the sample size; when themes were ranked by use, the most common themes used by top authors and all other authors were nearly identical.

When we examined themes for specific talking points, top authors promoted similar misinformation within each theme. For example, within conversation about negative health impacts, references to deaths and autism were mentioned in 67% and 44% of posts by top authors, respectively. In posts made by non–top authors, these 2 conditions were both mentioned in approximately 15% of posts. Throughout all themes, results showed how vaccine opponents can manipulate facts and their sources. It can be challenging for even experienced public health researchers to verify each claim made by a vaccine opponent, particularly given the amount, variety, and often misleading nature. For example, information taken from the Vaccine Adverse Event Reporting System, a database created by federal health agencies to monitor vaccine reactions, is used by vaccine opponents as “proof” of the government admitting that vaccines cause child death. Critically important context, such as the fact that an adverse event can be reported even if it is uncertain or unlikely that a vaccine caused it or the role of statistical significance or reporting bias in epidemiology, is lost. Misinformation is a complex issue involving not just what is said but also the intent behind it.

The finding that top authors share the same misinformation suggests that vaccine opponents rely on highly networked communities driven by leaders driving particular narratives. Influential vaccine opponents most likely select their messages based on the receptivity to those messages. By contrast, public health continues to repeat the same vaccination recommendations in the same manner, despite research demonstrating that these messages arrive in an echo chamber, received by those at little risk of vaccine hesitancy. The public health community should think critically and pivot messaging based on themes that receive the most engagement among those likely to be vaccine hesitant.

This study suggests that not only are vaccine opposition talking points discoverable but also that they can be quantified; there are only a handful employed at a given time. This aligns with research showing that a majority of Facebook advertisements opposing vaccination were funded by 2 groups. If these groups are passively monitored, as suggested by other researchers, public health may be able to counter the growing influence of vaccine opposition by quickly identifying and countering talking points.

**Limitations**

The study had limitations. Tweets were collected containing keywords identified (Appendix B). Tweets about vaccines that did not contain these terms were not collected. It is possible that posts were miscoded, particularly for those sarcastically referencing vaccine opposition. Analysts manually checked each theme to ensure at least 90% fidelity and amended keywords to capture sarcasm when possible. In addition, engagement and shared talking points were used as measures of influence, and there are likely other unexplored means of quantifying the influence of individuals in social networks. Furthermore, it is possible that engagements with top author posts were critical of vaccine opposition, rather than supportive. To address limitations, the methodology for automatic coding was tested and checked during this study, and previous research on automatic sentiment analysis for vaccine opposition on Twitter was consulted.

Although outside the scope of this study, research should explore the impact of seasonality on vaccine opposition. The data collection period spanned back-to-school season, flu season, and the legislative cycle. Seasonality was likely a contributing factor to misinformation. In addition, variables such as time or day of the week could be useful in understanding message spread.

**Public Health Implications**

Results highlighted common vaccine-related misinformation used by vaccine opponents. It will continue to be difficult for public health to effectively
counter vaccine opposition without a greater understanding of opposition actors and narratives. It is also important to note that, while this study examined vaccine opposition collectively, pro- and antivaccination beliefs are better represented as a spectrum, not as distinct states. Additional research should segment audiences that may be susceptible to specific messages highlighted in this study. In doing so, researchers can identify ways of utilizing and sharing retrospective, real-time, and predictive media data to create messaging that effectively and quickly reaches individuals who are vaccine hesitant.

CONTRIBUTORS
E. Bonnevie conceptualized the study, oversaw data collection and analysis, and led writing. A. K. Galegos-Jeffrey and S. D. Rosenberg analyzed the data. J. Goldbag, E. Wartella, and J. Snyder conceptualized the study and provided a critical review of the article. All authors approved the final version to be published.

ACKNOWLEDGMENTS
This work was supported by The New York State Health Foundation.

CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

HUMAN PARTICIPANT PROTECTION
No human participants were involved in this study. Institutional review board approval was sought from IntelReview, and the study was found to be exempt from review.

REFERENCES