Adapting and Extending a Typology to Identify Vaccine Misinformation on Twitter

Amelia Jamison, MAA, MPH, David A. Broniatowski, PhD, Michael C. Smith, PhD, Kajal S. Parikh BS, Adeena Malik BA, BS, Mark Dredze, PhD, and Sandra C. Quinn, PhD

Objectives. To adapt and extend an existing typology of vaccine misinformation to classify the major topics of discussion across the total vaccine discourse on Twitter.

Methods. Using 1.8 million vaccine-relevant tweets compiled from 2014 to 2017, we adapted an existing typology to Twitter data, first in a manual content analysis and then using latent Dirichlet allocation (LDA) topic modeling to extract 100 topics from the data set.

Results. Manual annotation identified 22% of the data set as antivaccine, of which safety concerns and conspiracies were the most common themes. Seventeen percent of content was identified as provaccine, with roughly equal proportions of vaccine promotion, criticizing antivaccine beliefs, and vaccine safety and effectiveness. Of the 100 LDA topics, 48 contained provaccine sentiment and 28 contained antivaccine sentiment, with 9 containing both.

Conclusions. Our updated typology successfully combines manual annotation with machine-learning methods to estimate the distribution of vaccine arguments, with greater detail on the most distinctive topics of discussion. With this information, communication efforts can be developed to better promote vaccines and avoid amplifying antivaccine rhetoric on Twitter. (*Am J Public Health.* 2020;110:S331–S339. https://doi.org/10.2105/AJPH.2020.305940)

See also Chou and Gaysynsky, p. S270.

At present, one of the greatest risks to human health comes from the deluge of misleading, conflicting, and manipulated information currently available online.[1] This includes health misinformation, defined as any "health-related claim of fact that is currently false due to a lack of scientific evidence."[2(p2417)] Vaccination is a topic particularly susceptible to online misinformation, even as the majority of people in the United States endorse the safety and efficacy of vaccines.[2,3] The reduction of infectious diseases through immunization ranks among the greatest health accomplishments of the 20th century, yet as the 21st century progresses, vaccine misinformation threatens to undermine these successes.[4] The rise in vaccine hesitancy—the delay and refusal of vaccines despite the availability of vaccination services—may be fueled, in part, by online claims that vaccines are ineffective, unnecessary, and dangerous.[5] While opposition to vaccines is not new, these arguments have been reborn via new technologies that enable the spread of false claims with unprecedented ease, speed, and reach.[2,6]

Combating vaccine misinformation requires an understanding of the prevalence and types of arguments being made and the ability to track how these arguments change over time. One of the earliest inventories of online vaccine misinformation comes from Kata's 2010 content analysis of antivaccine Web sites.[7] In this work, 8 Web sites were labeled for 6 "content attributes": alternative medicine; civil liberties; conspiracies and search for truth; morality, religion, and ideology claims; safety and effectiveness concerns; and misinformation. All Web sites shared content from more than 1 area: 100% promoted safety concerns and conspiracy content, 88% also promoted civil liberties and alternative medicine content, and 50% also promoted morality claims.[7] Misinformation and antivaccine arguments were nearly synonymous, with 88% of Web sites relying on outdated sources, misrepresenting facts, self-referencing "experts," or presenting unsupported falsehoods.[7]

Since 2010, both the Internet and the nature of vaccine misinformation have changed profoundly. The static Web sites Kata analyzed have been supplanted by interactive social media platforms as the primary channels for antivaccine information dissemination.[8] Unlike Web sites, which feature a single perspective, social media platforms were designed to encourage "dialogue" and feature a plurality of perspectives.[9] Social media also introduces new challenges, as opportunistic actors including automated bots and state-sponsored trolls flood channels with information designed to manipulate, provoke, or scam genuine users.[10]

Recognizing these changes, scholarly efforts to characterize vaccine misinformation

ABOUT THE AUTHORS

Amelia M. Jamison, Kajal S. Parikh, and Adeena Malik are with the Maryland Center for Health Equity, School of Public Health, University of Maryland, College Park. David A. Broniatowski and Michael C. Smith are with the Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, and Institute for Data, Democracy, and Politics, The George Washington University, Washington, DC. Mark Dredze is with the Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD. Sandra C. Quinn is with the Department of Family Science and Maryland Center for Health Equity, School of Public Health, University of Maryland, College Park.

Correspondence should be sent to Amelia Jamison, University of Maryland at College Park, Maryland Center for Health Equity, 1304 School of Public Health Building, 4200 Valley Drive, College Park, MD 20742 (e-mail: ajam1@umd.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted August 19, 2020.
https://doi.org/10.2105/AJPH.2020.305940

on social media have taken many forms. These include content analyses of vaccine posts on platforms including Twitter,[11] Facebook,[12] Instagram,[13] Pinterest,[14] and YouTube.[15] Although research questions varied—often tied to specific vaccines (e.g., human papillomavirus, influenza), temporal events (e.g., outbreaks, policy changes), or claims (e.g., debunked claim that vaccines cause autism)—the presence of misleading antivaccine content is near universal. The universality of Kata's broad categories endures, with many of these studies highlighting antivaccine content, questioning vaccine safety concerns, and promoting conspiracy theories.

More recently, computational advances have made big data and machine-learning methods popular. A common approach has been to use automated classification schemes to label posts by vaccine sentiment, either broad categorical analyses (e.g., sorting content into positive, negative, and neutral) or into topical categorization schemes (e.g., sorting content by topics such as safety, efficacy, and cost).[16–18] Other applications have included mapping semantic networks,[19] detecting network and community structures,[20] using topic modeling to infer areas of discussion,[21–23] classifying images,[24] and using machine-learning classifiers to infer geographic and demographic information.[25] Topic modeling has been particularly successful in surveillance of content shared by social media users and can be deployed in a variety of contexts, from monitoring key topics in human papillomavirus vaccine discussions on Reddit to identifying topical links between content posted by Russian Twitter troll accounts.[22,23] A new study used latent Dirichlet allocation (LDA) topic modeling to track 10 key influenza vaccine–related Twitter topics over time and measure how they correlated with vaccine attitudes.[26] The strength of automated approaches is in the ability to quickly analyze millions of messages; however, the results tend to be tied to specific data sets and often lack the broad applicability and simplicity of Kata's framework.

While these studies have expanded scholarly knowledge, big questions remain: What is the prevalence of both pro- and antivaccine content on social media platforms? What topics dominate the general vaccine discourse? And what topics are spreading misleading vaccine information? To answer these questions, we introduce a new typology, building upon Kata's 2010 work, but updating it for Twitter and introducing automated approaches to replicate our findings at scale. We chose Twitter as one of the key platforms sharing vaccine misinformation, but also as the most accessible for research.[27] Twitter may not be reflective of the attitudes held by the general public, but as a communication channel it plays a powerful role in amplifying vaccine messages and can foster online communities with shared interests. The resulting typology categorizes major themes and amplification strategies in the discourse of both vaccine opponents and vaccine proponents on Twitter. We believe this is a necessary first step toward developing a comprehensive survey of online vaccine discourse and foundational to developing successful efforts to fight misinformation.

METHODS

Our analysis followed 3 stages: first we conducted a manual content analysis on a subsample of vaccine-relevant tweets; then we utilized LDA, a type of probabilistic topic modeling, to infer the major topics of discussion in the total vaccine discourse; and, finally, we conducted a second manual content analysis of representative tweets from each of the 100 topics generated in stage 2.

Data

Our data set contained 1.8 million vaccine-relevant tweets collected between 2014 and 2017 through the Twitter public streaming keyword application programming interface. Tweets were English language, contained vaccine keywords (substrings "vax" or "vacc"), and had been filtered by using a machine-learning classifier trained to exclude tweets not relevant to vaccination (e.g., metaphors).[28]

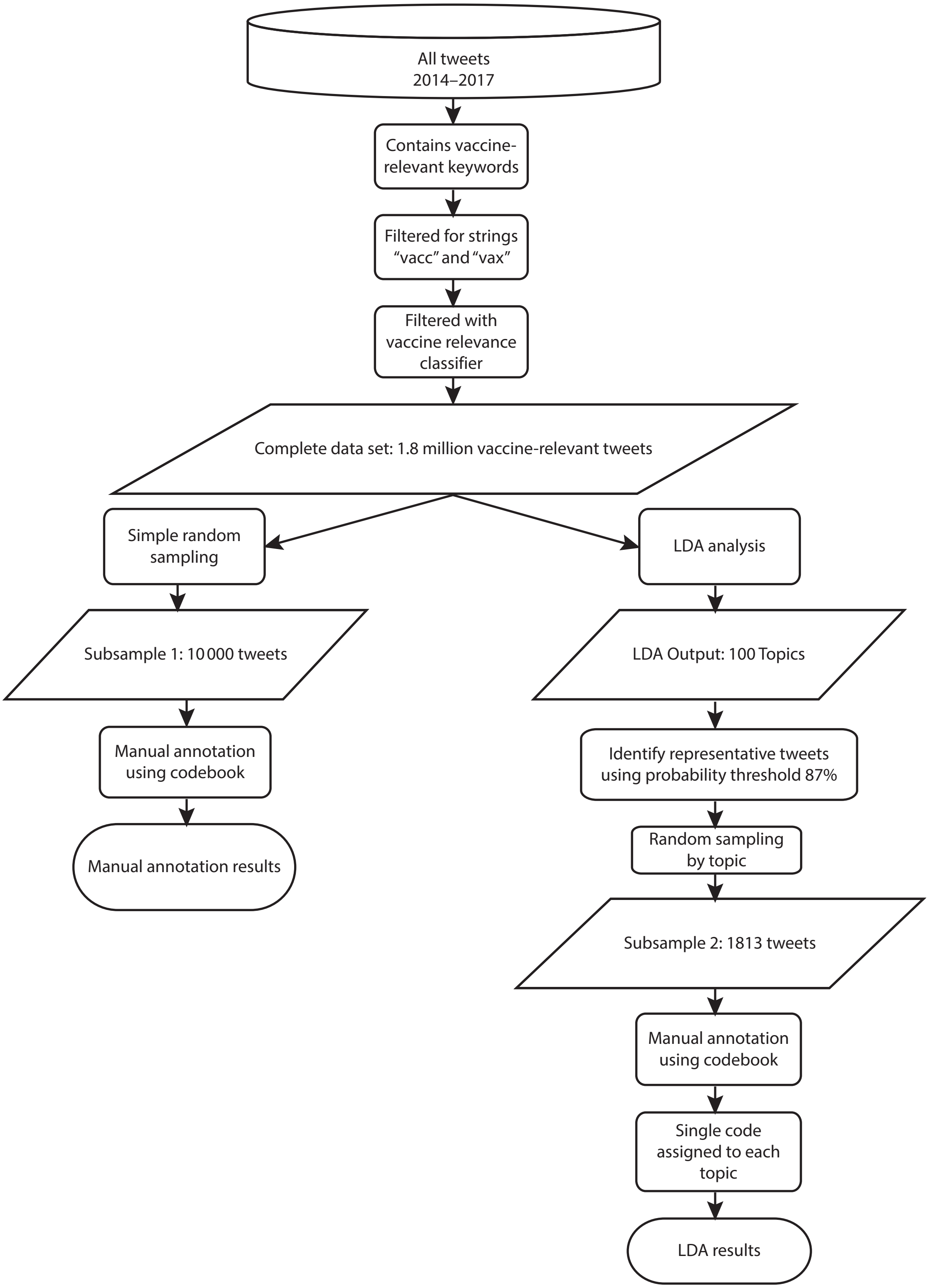
Content Analysis

Our first aim was to adapt Kata's typology to Twitter data by conducting a content analysis of 10 000 randomly selected vaccine-relevant tweets (Figure 1). We designed our approach to comply with an emerging set of best practices to ensure rigor and accuracy.[29] Tweets had been manually annotated for vaccine sentiment as part of an earlier project.[30] Two independent annotators (A. M., K. P.) then coded each nonneutral tweet into 1 or more thematic categories. On a random sample of 100 antivaccine tweets, annotators agreed 88.75% of the time on primary codes (Scott's $\pi = 0.85$; 95% confidence interval [CI] = 0.78, 0.93). On a random sample of 100 provaccine tweets, annotators agreed only 48% of the time for primary codes (Scott's $\pi = 0.38$; 95% CI = 0.26, 0.49), suggesting a much harder task. To address low reliability, a third team member (A. J.) reconciled discrepancies for both data sets and assigned final codes for each tweet.

The antivaccine codebook (see the box on p. S334) included adapted versions of 5 of Kata's 6 original content categories.[7] The final attribute, misinformation, was widespread across all categories and was not coded separately. We did not identify a provaccine equivalent to Kata's typology during our literature search and chose to develop our own. The annotation team created a set of deductive codes to mirror Kata's categories, using examples from the data set to justify each new code. For instance, "safety and efficacy" was determined to be the provaccine counterpart to antivaccine concerns about vaccine safety. In this way, we developed codes for pro-science, provaccine policy, criticizing antivaccine beliefs, and safety and effectiveness. Morality-based provaccine content (e.g., vaccinate to protect others) did not emerge as distinct theme. More common were tweets promoting vaccines without an underlying argument, prompting our fifth theme, "vaccine promotion."

Latent Dirichlet Allocation Topic Modeling

To infer distinctive topics of conversation across the entire sample of vaccine-relevant tweets, we used LDA, a widely used type of probabilistic topic model designed to identify underlying topics in a text data set by identifying groups of words that often co-occur (for more details on probabilistic topic models



Note. LDA = latent Dirichlet allocation.

FIGURE 1—Data Sources and Analysis Process Flowchart

CODEBOOK OF PRO- AND ANTIVACCINE THEMES FOR CONTENT ANALYSIS

Antivaccine content	
Alternative medicine[a]	Content that promotes alternatives to vaccination or alternative or complementary health systems or critiques biomedicine; content that promotes the benefits of "natural" immunity.
Civil liberties[a]	Content that opposes vaccination as an infringement of personal liberty, including opposition to vaccine mandates, parental choice narratives, vaccines as government overreach, and fears of punishment related to nonvaccination.
Conspiracy[a]	Content that presents specific conspiracy theories or conveys a broader "search for truth." Includes stories of fraud, cover-up, or collusion between government, doctors, and pharmaceutical companies; "rebel" spokespeople who speak "truth" at odds with the medical establishment; and unusual theories related to vaccines.
Morality[a]	Content that opposes vaccination for specific ideological reasons. Includes religious beliefs or morally loaded topics.
Safety concerns[a]	Content that critiques the safety or effectiveness of vaccines. Includes notion that vaccines cause harm, injury, or death; that vaccines are toxic or contain poison; and that vaccines fail to provide immunity.
Provaccine content	
Pro-science	Content that promotes vaccine science or science in general. Includes defending science against pseudoscientific claims.
Provaccine policy	Content supporting expanded vaccine policies. Includes support for mandatory vaccination, opposition to nonmedical vaccine exemptions, and opposition to "vaccine choice."
Criticizing antivaccine beliefs	Content focused on refuting antivaccine arguments or blaming/shaming antivaccine individuals. Includes ideas that antivaxxers are uninformed, bad parents, endangering others, etc. Also includes aggressive appeals to vaccinate.
Promotion	Content focused on specific vaccine campaigns. Includes recommendations for vaccines, philanthropic vaccination campaigns, and details on when, where, or how to get a vaccine.
Safety and efficacy	Content that describes the vaccines as safe or effective. Includes successes of vaccines, and reduction in disease. Also includes benefits of vaccination or risks of nonvaccination.
Amplification strategies	
Retweets	Retweets make up majority of topic, suggesting organized retweeting effort (possibly bot-driven).
Hashtags	Use of common hashtags to promote content.
@mentions	Inserting @ to high-profile individuals and organizations to gain attention.
Politics	Engaging in partisan political debate, referencing political candidates.

[a]Adapted from Kata typology.[7]

see Blei[31]).[32] LDA is increasingly common in health informatics research as a method to assess large text-based data sets (see also Walter et al.[23] and Chan et al.[26]). LDA assumes that each document (in this case, a tweet) contains an underlying mixture of topics and that each topic can be captured by an underlying mixture of words. We trained LDA with 100 topics on a subset of 1 million tweets, then inferred the topics on the remaining 800 000 tweets by using the trained model. In training the model, we preselect the number of topics we expect to find and then optimize the model for the most likely arrangement of words in each topic and topics in each document. We used the implementation of LDA from the MALLET topic modeling toolkit and used the default parameter settings unless otherwise noted.[33] Every tweet receives scores reflecting probabilities for all underlying topics; the highest scoring topic is then taken as the primary topic for that tweet. For each topic, we aggregated tweets with the highest topic probabilities (87%–95%). After excluding topics that returned fewer than 100 tweets or non-English content, the new data set contained 26 542 tweets, with an average of 285 tweets per topic (Figure 1).

Integrating the Typology

To understand how LDA topics fit within the updated typology, we conducted a second content analysis, randomly selecting up to 20 of the most relevant tweets from each LDA topic (Figure 1). LDA outputs provide keywords for each topic, but these can sometimes include co-occurring words that may not be truly conceptually related; therefore, it is important to assess highly representative full-length tweets (for topic keywords see Appendix A, available as a supplement to the online version of this article at http://www.ajph.org). Three annotators (A. J., A. M., K. P.) independently labeled each tweet for vaccine sentiment and theme. Across 100 randomly selected tweets, we observed 79% agreement on vaccine sentiment (Fleiss's $\kappa = 0.69$; 95% CI = 0.59, 0.78) and 82% agreement on content labels for nonneutral tweets (Fleiss's $\kappa = 0.78$; 95% CI = 0.66, 0.89). Topics were then arranged by sentiment and

TABLE 1—Topics Labeled by Vaccine Sentiment and Theme: Twitter, 2014–2017

Topic Label, Topic No.	% Pro/%Anti	Theme	Amplification
Provaccine (>70%) 18 topics			
Antivaccine beliefs are stupid	100/0	Criticizing antivaccine	
Global philanthropic campaigns	100/0	Promotion	
Benefits of influenza vaccine	100/0	Safe and effective	
Promoting vaccine clinics	100/0	Promotion	
Recommendations for kids and adults	100/0	Promotion	
Antivaxxers are bad parents	90/0	Criticizing antivaccine	Hashtags
Philanthropic funding for vaccines	90/0	Promotion	
Vaccinate to protect others	90/5	Safe and effective	
Recommendations for students	85/0	Promotion	Hashtags
Human papillomavirus vaccine reduces cancer	85/0	Safe and effective	
Expanding UK meningitis vaccine	85/10	Provaccine policy	
Great moments in vaccine history	80/0	Pro-science	
Antivaxxers are crazy	80/5	Criticizing antivaccine	
Nonvaccination spreads disease	80/15	Safe and effective	
Population impact of vaccines	75/0	Safe and effective	
Debunking myths about vaccines (provaccine hashtags)	75/0	Safe and effective	Hashtags
Human papillomavirus vaccine recommendations	75/0	Promotion	Hashtags
Defending science	75/0	Pro-science	
Provaccine/neutral (20%–70%), 21 topics			
Philanthropic campaigns for polio	70/0	Promotion	
Political backlash to antivaccine views	70/0	Criticizing antivaccine	Politics
Vaccines reduce disease (measles in European Union)	65/5	Safe and effective	
Influenza vaccine recommendations	65/10	Promotion	
New Ebola vaccine successes	60/0	Safe and effective	
Republican primary debate	60/0	Criticizing antivaccine	Politics
Vaxxed screening debate with DeNiro	60/10	Criticizing antivaccine	
Tropical diseases campaigns	55/0	Promotion	
National Immunization Awareness Week	55/0	Promotion	Hashtags
Support for passing of Senate Bill 277 in California	55/0	Provaccine policy	
Science to address misperceptions	55/5	Pro-science	
Unfollowing/blocking antivax	50/5	Criticizing antivaccine	
Promoting vaccines during pregnancy	40/0	Promotion	
Blaming antivaxxers for disease	40/5	Criticizing antivaccine	
Influenza vaccine efficacy	40/5	Safe and effective	
Opposing "parental rights" arguments	40/15	Provaccine policy	
Avian influenza vaccine efficacy	30/5	Safe and effective	
Rejecting alternative schedules	30/5	Criticizing antivaccine	
Linking outbreaks to unvaccinated	25/0	Criticizing antivaccine	
Vaccine hesitancy research	20/0	Safe and effective	
Vaccines 2016 election	20/15	Criticizing antivaccine	Politics
Both (> 20% each), 9 topics			
Debating parental choice arguments	70/20	Criticizing antivaccine/civil liberties	
Discussing immune response	50/20	Safe and effective/safety concerns	
Vaccine mandates	35/25	Policy/conspiracy	
Other side does not understand science	35/45	Pro-science/conspiracy	
Debating epidemiological evidence	30/25	Criticizing antivaccine/safety concerns	@messages
Policy debate on Trump's views	25/25	Safe and effective/civil liberties	Politics, retweets
Debating safety and evidence	25/30	Safe and effective/safety concerns	@messages
Debating school vaccine requirements	25/35	Policy/civil liberties	

Continued

TABLE 1—*Continued*

Topic Label, Topic No.	% Pro/%Anti	Theme	Amplification
Vaccine harms debate	25/40	Safe and effective/safety concerns	@messages
Antivaccine/neutral (20%–70%), 9 topics			
Influenza vaccine efficacy and side effects	0/20	Safety concerns	
Antivaccine "expert" opinions	0/20	Conspiracy	
Enzyme in vaccines causes cancer	0/25	Safety concerns	
Safety Commission rumors	0/25	Conspiracy	
Babies dying after vaccination	10/25	Safety concerns	
Zika vaccine conspiracy	5/25	Conspiracy	
Political spam tweets (e.g., #BLM, #Latino)	0/25	Civil liberties	Politics, retweets
Race-based sterilization in Africa	5/35	Conspiracy	
Vaccine mandates violate rights	0/40	Civil liberties	Politics
Antivaccine (75%–100%), 10 topics			
Vaccine industry corruption	0/75	Conspiracy	
Gates/Clinton conspiracies	5/75	Conspiracy	Politics
#CDCWhistleblower coverup	5/80	Conspiracy	Hashtags
Doctor reveals "truth"—found dead	0/90	Conspiracy	Retweets, @messages
Vaccines cause severe harm (#Vaxxed)	0/95	Safety concerns	Retweets
Parents of vaccine-injured children	0/95	Safety concerns	@messages, retweets, hashtags
Environmental toxins	0/95	Safety concerns	
Vaccines contain disgusting things	0/95	Safety concerns	
Vaccines cause autism (antivaccine hashtags)	0/90	Safety concerns	Hashtags
Shaken baby syndrome cover-up	0/100	Conspiracy	Retweets, @messages

Note. Neutral, 33 topics: vaccines in development (8 topics), not English (6 topics), pet or animal vaccines (4 topics), pharmaceutical or technical information (3 topics), jokes or memes (2 topics), and The Vaccines band, general public health, vaccines are painful, Chinese vaccine scandal, news coverage on HPV, pricing for vaccines, links to porn, personal stories, unclear hashtags, vaccine supply chain (1 topic each).

divided into categories: majority pro- or antivaccine (>70%nonneutral), neutral and pro- or antivaccine (20%–70% nonneutral), majority neutral (< 20% nonneutral), or both (> 20% both provaccine and antivaccine; Table 1).[7,9] In addition, we used this space to incorporate labels for Twitter-specific information including hashtags, @mentions, and retweet campaigns.

TABLE 2—Manual Annotation and Topic Labels by Theme: Twitter, 2014–2017

Broad Theme	Manual Annotation, No. (%)	LDA Topics, No. (%)	P^a
Antivaccine	2241	28	.03
Alternative medicine	50 (2)	0 (0)	
Civil liberties	248 (11)	5 (18)	
Conspiracy	903 (40)	11 (39)	
Morality	105 (5)	0 (0)	
Safety concerns	1322 (59)	12 (43)	
Provaccine	1744	48	.24
Pro-science	85 (5)	4 (8)	
Provaccine policy	92 (5)	5 (10)	
Criticizing antivaccine	538 (31)	13 (27)	
Promotion	648 (37)	11 (23)	
Safety and effectiveness	523 (30)	15 (31)	
Neutral	6015	33	

Note. LDA = latent Dirichlet allocation.
[a]Fisher exact test or χ^2.

RESULTS

First we present results from the content analysis, then we present results integrating LDA and manual coding.

Prevalence of Vaccine Themes

Of the 10 000 messages annotated in subsample 1, 22% (n = 2241) were antivaccine, 17% (n = 1744) were provaccine, and the remaining 61% (n = 6015) were neutral or not relevant (Table 2). Among antivaccine tweets, safety concerns was the single most common theme (59%; n = 1320) followed by conspiracies (41%; n = 930), civil liberties (11%; n = 248), morality claims (5%; n = 105), and alternative medicines (2%; n = 50). Most tweets (68%; n = 1666) were labeled with a single topic. Co-occurrence was most common between safety concerns and conspiracies (n = 411).

For provaccine tweets, vaccine promotion (37%; n = 648) was the most common theme, followed by criticizing antivaccine beliefs (31%; n = 538) and safety and effectiveness

(30%; n = 523). Fewer messages were provaccine policy (5%; n = 92) and pro-science (5%; n = 85). Most tweets (89%; n = 1550) had a single label. The most common co-occurrence was between criticizing antivaccine beliefs and vaccine safety and effectiveness (n = 182).

Characterizing Latent Dirichlet Allocation Topics

Of the 100 LDA topics, 28 topics included significant antivaccine content, 48 included provaccine content, and 33 were neutral or not relevant (Table 1). Within each of these categories, we recognized a spectrum: 10 topics were majority antivaccine, 9 combined antivaccine and neutral content, 9 included both provaccine and antivaccine content, 20 combined provaccine and neutral content, and 18 were majority provaccine.

Although the proportions of nonneutral tweets and nonneutral topics were significantly different ($X^2 = 23.50$; $P < .001$) the distribution of themes was roughly equivalent (Table 2). For provaccine topics, the same 3 topics—safety and efficacy, vaccine promotion, criticizing antivaccine beliefs —were the most represented, with slightly greater representation of provaccine policy among topics ($P = .03$; Fisher's test). For antivaccine topics, we observed no significant differences between the distributions of themes ($X^2 = 5.54$; $P = .24$).

Topics that were primarily antivaccine consisted entirely of conspiracy and safety concerns (5 topics each, 50%). Conspiracy claims tended to focus on governmental and pharmaceutical fraud, while safety concerns included claims of vaccine-induced idiopathic illnesses and vaccines as poison. Among these 10 topics, we found the highest concentration of retweet activity in the data set, with 3 topics dominated by nearly verbatim retweets (possibly indicating bot-like activity). Other amplification strategies included use of antivaccine hashtags and @messages to celebrities and public officials for attention. Topics that combined antivaccine and neutral content (n = 9) included conspiracies (4 topics, 44%) and safety concerns (3 topics, 33%), but also civil liberties (2 topics, 22%). In these topics, neutral news content appeared alongside antivaccine claims and sometimes political content.

Majority provaccine topics (n = 18) included all 5 themes. Vaccine promotion efforts included a mix of event promotion, philanthropy efforts, and vaccine recommendations. Safety and efficacy topics emphasized the risks of not vaccinating, benefits of vaccines, or simply proclaimed #vaccineswork. Antivaccine-critical topics shamed antivaccine parents as crazy, stupid, and neglectful parents—sometimes relying on satire or parody. Pro-science topics included celebrations of vaccines as a major public health accomplishment but also included defending science against "fake news." Like antivaccine-critical topics, some of these claims relied on humor. Provaccine policy topics included discussion of vaccine mandates. Topics that combined provaccine and neutral content (n = 20) also included all 5 themes but included topics that were more controversial or polarizing like political discussions and influenza vaccine topics.

The 9 topics that combined significant provaccine and antivaccine sentiment highlighted areas of overlap in the discourse. This included debated topics in which users repeated and refuted arguments, such as differing interpretations of epidemiological evidence or the legality of mandates. It also included arguments with parallel structure; antivaccine arguments that claim vaccine science is "bad science" appeared alongside provaccine arguments describing vaccine opposition as pseudoscience. These conversations included more neutral hashtags (e.g., #immunity, #vaccine) and reliance on @messages to directly contact other users.

DISCUSSION

The sheer volume of vaccine information on Twitter presents major challenges for researchers trying to systematically address misleading information. With this analysis, we introduce an innovative approach to estimate the prevalence of vaccine themes and classify major topics, providing a comprehensive assessment of the vaccine discourse on Twitter. We found a slightly greater proportion of antivaccine messages than provaccine messages (22% to 17%), with many more messages neutral on vaccines—findings in line with previous work.[17] However, topic modeling demonstrated that distinctive topics of conversation tend to be nonneutral, with a greater diversity of topics containing provaccine content from all 5 thematic areas, while topics containing antivaccine content concentrated on safety concerns and conspiracy theories. Neutral tweets represented most of the data set, but topic modeling demonstrated how they can serve as the foundation for both provaccine and antivaccine arguments, with a roughly one third of all topics mixing both neutral and polarized content.

Although very different in tone and sentiment, provaccine and antivaccine messages were more structurally similar than we anticipated. Because LDA analysis depends on word choice and language structure to identify coherent topics, that 9 topics included significant proportions of both provaccine and antivaccine content suggests use of similar language and rhetorical strategies. This does not necessarily mean that vaccine opponents and proponents were directly engaged; indeed, previous work has highlighted echo-chamber effects that limit exposure between outside viewpoints at work in vaccine communities on Twitter.[34] However, the 2 communities may be indirectly influencing each other's arguments, as evidenced by similar use of semantic strategies.

The Twitter features that allow for the spread of antivaccine content have likely also reshaped how provaccine content spreads. The prevalence of straightforward vaccine promotion content suggests that Twitter is a useful platform to easily share recommendations, remind patients to get vaccinated, and provide links to events. The increased visibility of the antivaccine movement has also likely shaped the ways Twitter users use the platform to defend vaccines. This is most clear in use of the platform to debunk antivaccine conspiracies, vent anger, or otherwise shame, blame, or complain about antivaccine parents. However, many debunking efforts tended to focus on a narrow set of outdated antivaccine claims suggesting that those most critical of antivaccine arguments are responding to an abstract idea of the antivaccine population and not engaging with antivaccine topics directly on Twitter. Defending vaccines also manifested in more subtle ways, like the #vaccineswork hashtag, where users felt the need to tweet in support of vaccines, making visible a sentiment that until recently many viewed as standard. This response is mirrored in the broader "defense of science" debates happening on Twitter as users see antivaccine arguments as part of a broader antiscience trend.

In addition to characterizing topics, we were able to observe how different arguments aligned with misinformation and amplification strategies. While antivaccine arguments are using many of the same strategies Kata detailed in 2010, including presenting unsupported falsehoods and misrepresenting scientific evidence, we saw some newer strategies that are tailored to Twitter. With strict character limits, tweets do not allow for contextualization, making it much easier to mislead by using sensational falsehoods or manipulations of real data. Some antivaccine claims are presented as facts, mimicking the language of mainstream news or science. In these instances, source credibility may be more important for users to gauge validity. Although both vaccine opponents and proponents have successfully utilized hashtags, we found @messaging and retweet campaigns were more common in antivaccine topics. Political language also appeared in both pro- and antivaccine content.

We identified a Twitter-specific amplification strategy that relied on massive retweet campaigns, suggesting evidence of concentrated effort by 1 or a group of actors. These campaigns may be driven by genuine users but could also indicate networks of automated accounts. Unlike previous studies that have characterized users as likely bots,[30] focusing on messaging led us to look for evidence of bot-like behavior, of which these massive retweet campaigns were the most egregious.[10] Topics consisting largely of retweets were among the most clearly misleading or political, including claims that shaken baby syndrome was a cover-up for vaccine injury and stories of an alternative medicine doctor mysteriously found dead after "exposing the truth" about vaccines. This suggests that amplification and automation have been successfully used to artificially inflate the appearance of antivaccine and political topics.

Limitations

This research is not without limitations. LDA methods have their own drawbacks: researchers must preselect the number of topics to be inferred, , setting too many or too few topics can produce different results, and the resulting topics can be too specific or too broad depending on selected parameters.[35] For these reasons, LDA is best at providing broad overviews of content, not nuanced analysis of specific topics. More broadly, by focusing our analyses on text only, we cannot ascertain the identity of the user, network features, the source of the linked content, or the impact of embedded images, which undoubtedly influence how information is perceived.

In the manual analysis, intercoder reliability for provaccine annotations was quite low; we believe this reflects the high level of similarity between our chosen categories. Vaccine opponents typically level specific claims against vaccines, but vaccine proponents tend to use very general language in support of vaccines, making it difficult to select a specific provaccine code. Future research is needed to refine this coding strategy. The challenges from the first round of annotation were largely absent during the second round of annotation, suggesting that annotators improved over time or that the nonrandom distribution of tweets from topic models aided in comprehension.

Public Health Implications

This updated typology was designed to distill relevant information from across the entire vaccine discourse on Twitter quickly and accurately. Mapping the proportion of tweets was necessary first step, but we believe understanding how these themes play out in online conversation can better inform communication efforts on how users engage with vaccine topics on Twitter. At this stage, our findings remain quite general but lend themselves to several specific recommendations, particularly for provaccine messaging. While vaccine proponents are already using the platform to debunk specific antivaccine claims, these are often not the same claims promoted in antivaccine topics. Rather than address rumors directly and risk amplifying them further, it may be more beneficial for vaccine advocates to continue to emphasize the safety and efficacy of vaccines in general terms. Similarly, engaging with a bot-driven narrative only further amplifies the message. It is also important to communicate to the Twitter users eager to defend vaccines that the humor used to criticize antivaccine tweets and anti-science tweets may inadvertently mislead and further provoke.[36] This updated typology serves as a proof of concept. Future research efforts should explore specific communication strategies and extend similar approaches to map vaccine discourse and associated misinformation on additional platforms. AJPH

CONTRIBUTORS

A. M. Jamison designed the content analysis portion of the study, lead the manual annotation team, and wrote the article. D. A. Broniatowski designed the computational portion of the study and provided critical revision of the article. M. C. Smith performed the computational analysis. K. S. Parikh and A. Malik served as annotators for the content analysis and assisted with codebook development. M. Dredze aided in data collection and provided critical revision of the article. S. C. Quinn provided health communication context and provided critical revision of the article.

ACKNOWLEDGMENTS

Research reported in this publication was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under award 5R01GM114771.

Note. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICTS OF INTEREST

M. Dredze holds equity in Sickweather Inc and has received consulting fees from Bloomberg LP and Good Analytics Inc. These organizations did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the article.

D. A. Broniatowski received a $500 honorarium for presenting a talk to the United Nations Shot@Life Foundation, a nonprofit organization that promotes immunization for children worldwide. This organization was not involved in any way with this study's design, analysis, or results.

HUMAN PARTICIPANT PROTECTION

The data used in this article are from a publicly available online source, the uses of which are deemed institutional review board–exempt by the University of Maryland institutional review board (1363471-1).

REFERENCES

1. Larson HJ. The biggest pandemic risk? Viral misinformation. *Nature*. 2018;562(7727):309. https://doi.org/10.1038/d41586-018-07034-4

2. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. *JAMA*. 2018; 320(23):2417–2418. https://doi.org/10.1001/jama.2018.16865

3. Pew Research Center. Vast majority of Americans say benefits of childhood vaccines outweigh risks. 2017. Available at: https://www.pewresearch.org/science/2017/02/02/vast-majority-of-americans-say-benefits-of-childhood-vaccines-outweigh-risks. Accessed: June 25, 2019.

4. World Health Organization. Ten threats to global health in 2019. 2019. Available at: https://www.who.int/emergencies/ten-threats-to-global-health-in-2019. Accessed January 4, 2019.

5. MacDonald NE. Vaccine hesitancy: definition, scope and determinants. *Vaccine*. 2015;33(34):4161–4164. https://doi.org/10.1016/j.vaccine.2015.04.036

6. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science*. 2018;359(6380):1146–1151. https://doi.org/10.1126/science.aap9559

7. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine*. 2010;28(7):1709–1716. https://doi.org/10.1016/j.vaccine.2009.12.022

8. Betsch C, Brewer NT, Brocard P, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine*. 2012;30(25):3727–3733. https://doi.org/10.1016/j.vaccine.2012.02.025

9. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm—an overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*. 2012;30(25):3778–3789. https://doi.org/10.1016/j.vaccine.2011.11.112

10. Jamison AM, Broniatowski DA, Quinn SC. Malicious actors on Twitter: a guide for public health researchers. *Am J Public Health*. 2019;109(5):688–692. https://doi.org/10.2105/AJPH.2019.304969

11. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*. 2010;5(11):e14118. https://doi.org/10.1371/journal.pone.0014118

12. Hoffman BL, Felter EM, Chu KH, et al. It's not all about autism: the emerging landscape of anti-vaccination sentiment on Facebook. *Vaccine*. 2019;37(16):2216–2223. https://doi.org/10.1016/j.vaccine.2019.03.003

13. Kearney MD, Selvan P, Hauer MK, Leader AE, Massey PM. Characterizing HPV vaccine sentiments and content on Instagram. *Health Educ Behav*. 2019;46(2 suppl):37–4.8 https://doi.org/10.1177/1090198119859412

14. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine*. 2015;33(39):5051–5056. https://doi.org/10.1016/j.vaccine.2015.08.064

15. Donzelli G, Palomba G, Federigi I, et al. Misinformation on vaccination: a quantitative analysis of YouTube videos. *Hum Vaccin Immunother*. 2018;14(7):1654–1659. https://doi.org/10.1080/21645515.2018.1454572

16. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. *J Biomed Semantics*. 2017;8(1):9. https://doi.org/10.1186/s13326-017-0120-6

17. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLOS Comput Biol*. 2011;7(10):e1002199. https://doi.org/10.1371/journal.pcbi.1002199

18. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. . Using social connection information to improve opinion mining: identifying negative sentiment about HPV vaccines on Twitter. *Stud Health Technol Inform*. 2015;216:761–765. https://doi.org/10.3233/978-1-61499-564-7-761

19. Kang GJ, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine*. 2017;35(29):3621–3638. https://doi.org/10.1016/j.vaccine.2017.05.052

20. Bello-Orgaz G, Hernandez-Castro J, Camacho D. Detecting discussion communities on vaccination in Twitter. *Future Gener Comput Syst*. 2017;66:125–136. https://doi.org/10.1016/j.future.2016.06.032

21. 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. https://doi.org/10.2196/jmir.6045

22. Lama Y, Hu D, Jamison A, Quinn SC, Broniatowski DA. Characterizing trends in human papillomavirus vaccine discourse on Reddit (2007–2015): an observational study. *JMIR Public Health Surveill*. 2019;5(1):e12480. https://doi.org/10.2196/12480

23. Walter D, Ophir Y, Jamieson KH. Russian Twitter accounts and the partisan polarization of vaccine discourse, 2015–2017. *Am J Public Health*. 2020;110(5):718–724. https://doi.org/10.2105/AJPH.2019.305564

24. Chen T, Dredze M. Vaccine images on Twitter: analysis of what images are shared. *J Med Internet Res*. 2018;20(4):e130. https://doi.org/10.2196/jmir.8221

25. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009–15. *Soc Sci Med*. 2017;191:168–175. https://doi.org/10.1016/j.socscimed.2017.08.041

26. Chan MS, Jamieson KH, Albarracin D. Prospective associations of regional social media messages with attitudes and actual vaccination: a big data and survey study of the influenza vaccine in the United States. *Vaccine*. 2020; Epub ahead of print. https://doi.org/10.1016/j.vaccine.2020.07.054

27. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center. 2019. Available at: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users. Accessed January 21, 2020.

28. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need social media now. *Am J Prev Med*. 2016;50(4):550–552. https://doi.org/10.1016/j.amepre.2015.10.002

29. Colditz JB, Chu KH, Emery SL, et al. Toward real-time infoveillance of Twitter health messages. *Am J Public Health*. 2018;108(8):1009–1014. https://doi.org/10.2105/AJPH.2018.304497

30. Broniatowski DA, Jamison AM, Qi SH, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Am J Public Health*. 2018;108(10):1378–1384. https://doi.org/10.2105/AJPH.2018.304567

31. Blei DM. Probabilistic topic models. *Commun ACM*. 2012;55(4):77–84. https://doi.org/10.1145/2133806.2133826

32. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003;3:993–1022.

33. McCallum AK. Mallet: Machine Learning for Language Toolkit. University of Massachusetts Amherst. 2002. Available at: http://mallet.cs.umass.edu/about.php. Accessed January 6, 2020.

34. Himelboim I, Xiao X, Lee DKL, Wang MY, Borah P. A social networks approach to understanding vaccine conversations on Twitter: network clusters, sentiment, and certainty in HPV social networks. *Health Commun*. 2020;35(5):607–615. https://doi.org/10.1080/10410236.2019.1573446

35. van Kessel P. Making sense of topic models. *Decoded*. Pew Research Center. August 13, 2018. Available at: https://medium.com/pew-research-center-decoded/making-sense-of-topic-models-953a5e42854e. Accessed August 13, 2020.

36. Wardle C. The different types of mis and disinformation. *First Draft*. February 16, 2017. Available at: https://firstdraftnews.org/latest/fake-news-complicated. Accessed April 29, 2020.