Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate

David A. Broniatowski, PhD, Amelia M. Jamison, MAA, MPH, SiHua Qi, SM, Lulwah AlKulaib, SM, Tao Chen, PhD, Adrian Benton, MS, Sandra C. Quinn, PhD, and Mark Dredze, PhD

Objectives. To understand how Twitter bots and trolls ("bots") promote online health content.

Methods. We compared bots' to average users' rates of vaccine-relevant messages, which we collected online from July 2014 through September 2017. We estimated the likelihood that users were bots, comparing proportions of polarized and antivaccine tweets across user types. We conducted a content analysis of a Twitter hashtag associated with Russian troll activity.

Results. Compared with average users, Russian trolls ($\chi^2(1) = 102.0$; $P < .001$), sophisticated bots ($\chi^2(1) = 28.6$; $P < .001$), and "content polluters" ($\chi^2(1) = 7.0$; $P < .001$) tweeted about vaccination at higher rates. Whereas content polluters posted more antivaccine content ($\chi^2(1) = 11.18$; $P < .001$), Russian trolls amplified both sides. Unidentifiable accounts were more polarized ($\chi^2(1) = 12.1$; $P < .001$) and antivaccine ($\chi^2(1) = 35.9$; $P < .001$). Analysis of the Russian troll hashtag showed that its messages were more political and divisive.

Conclusions. Whereas bots that spread malware and unsolicited content disseminated antivaccine messages, Russian trolls promoted discord. Accounts masquerading as legitimate users create false equivalency, eroding public consensus on vaccination.

Public Health Implications. Directly confronting vaccine skeptics enables bots to legitimize the vaccine debate. More research is needed to determine how best to combat bot-driven content. (*Am J Public Health.* Published online ahead of print August 23, 2018: e1–e7. doi:10.2105/AJPH.2018.304567)

Health-related misconceptions, misinformation, and disinformation spread over social media, posing a threat to public health.[1] Despite significant potential to enable dissemination of factual information,[2] social media are frequently abused to spread harmful health content,[3] including unverified and erroneous information about vaccines.[1,4] This potentially reduces vaccine uptake rates and increases the risks of global pandemics, especially among the most vulnerable.[5] Some of this information is motivated: skeptics use online platforms to advocate vaccine refusal.[6] Antivaccine advocates have a significant presence in social media,[6] with as many as 50% of tweets about vaccination containing antivaccine beliefs.[7]

Proliferation of this content has consequences: exposure to negative information about vaccines is associated with increased vaccine hesitancy and delay.[8–10] Vaccine-hesitant parents are more likely to turn to the Internet for information and less likely to trust health care providers and public health experts on the subject.[9,11] Exposure to the vaccine debate may suggest that there is no scientific consensus, shaking confidence in vaccination.[12,13] Additionally, recent resurgences of measles, mumps, and pertussis and increased mortality from vaccine-preventable diseases such as influenza and viral pneumonia[14] underscore the importance of combating online misinformation about vaccines.

Much health misinformation may be promulgated by "bots"[15]—accounts that automate content promotion—and "trolls"[16]—individuals who misrepresent their identities with the intention of promoting discord. One commonly used online disinformation strategy, amplification,[17] seeks to create impressions of false equivalence or consensus through the use of bots and trolls. We seek to understand what role, if any, they play in the promotion of content related to vaccination.

Efforts to document how unauthorized users—including bots and trolls—have influenced online discourse about vaccines have been limited. DARPA's (the US Defense Advanced Research Projects Agency) 2015 Bot Challenge charged researchers with identifying "influence bots" on Twitter in a stream of vaccine-related tweets. The teams effectively identified bot networks designed to spread vaccine misinformation,[18] but the public health community largely overlooked the implications of these findings. Rather, public health research has focused on combating online antivaccine content, with less focus on the actors who produce and promote this content.[1,19] Thus, the role of bots' and trolls' online activity pertaining to vaccination remains unclear.

We report the results of a retrospective observational study assessing the impact of

ABOUT THE AUTHORS

David A. Broniatowski is with the Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, Washington, DC. Amelia M. Jamison and Sandra C. Quinn are with the Department of Family Science, School of Public Health, University of Maryland, College Park. Sihua Qi and Lulwah Alkulaib are with the Department of Computer Science, School of Engineering and Applied Science, The George Washington University. Tao Chen, Adrian Benton, and Mark Dredze are with the Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD.

Correspondence should be sent to David A. Broniatowski, 800 22nd St. NW #2700, Washington, DC 20052 (e-mail: broniatowski@gwu.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted May 22, 2018.

doi: 10.2105/AJPH.2018.304567

bots and trolls on online vaccine discourse on Twitter. Using a set of 1 793 690 tweets collected from July 14, 2014, through September 26, 2017, we quantified the impact of known and suspected Twitter bots and trolls on amplifying polarizing and antivaccine messages. This analysis is supplemented by a qualitative study of #VaccinateUS—a Twitter hashtag designed to promote discord using vaccination as a political wedge issue. #VaccinateUS tweets were uniquely identified with Russian troll accounts linked to the Internet Research Agency—a company backed by the Russian government specializing in online influence operations.[20] Thus, health communications have become "weaponized": public health issues, such as vaccination, are included in attempts to spread misinformation and disinformation by foreign powers. In addition, Twitter bots distributing malware and commercial content (i.e., spam) masquerade as human users to distribute antivaccine messages. A full 93% of tweets about vaccines are generated by accounts whose provenance can be verified as neither bots nor human users yet who exhibit malicious behaviors. These unidentified accounts preferentially tweet antivaccine misinformation. We discuss implications for online public health communications.

METHODS

In our first analysis, we examined whether Twitter bots and trolls tweet about vaccines more frequently than do average Twitter users. In a second analysis, we examined the relative rates with which each type of account tweeted provaccine, antivaccine, and neutral messages. Finally, in a third analysis, we identified a hashtag uniquely used by Russian trolls and used qualitative methods to describe its content.

Data Collection

We drew all tweets in our first analysis from 1 of 2 data sets derived from the Twitter streaming application programming interface (API): (1) a random sample of 1% of all tweets ("the 1% sample"), and (2) a sample of tweets containing vaccine-related keywords ("the vaccine stream"; Table A, available as a supplement to the online version of this article at http://www.ajph.org). For each data set, we extracted tweets from accounts known to be bots or trolls and identified in 7 publicly available lists of Twitter user IDs.[20–26] We compared these with a roughly equal number of randomly selected tweets that were posted in the same time frame. We calculated the relative frequency with which each type of account tweeted about vaccines by counting the total number of tweets containing at least 1 word beginning with "vax" or "vacc."

In our second analysis, we collected a random subset of tweets from all users in the vaccine stream containing the strings "vax" or "vacc" and tagged them as relevant to vaccines by a machine-learning classifier developed for that purpose by Dredze et al.[27] We used the Botometer[28] API—a widely used[29] bot-detection tool—to estimate each tweet's "bot score," reflecting the likelihood that its author is a bot. Botometer returns a likelihood score between 0% and 100% for each query and cannot make an accurate assessment for all accounts. Thus, we segmented accounts into 3 categories: those with scores less than 20% (very likely to be humans), greater than 80% (very likely to be bots), and between 20% and 80% (of uncertain provenance). Finally, we applied the same criteria to a subset of tweets from the vaccine stream for each of the 7 types of known bot and troll accounts identified in the first analysis. All data collection procedures are described in detail in Appendix A, available as a supplement to the online version of this article at http://www.ajph.org.

Analysis

Are bots and trolls more likely to tweet about vaccines? We tested the hypothesis that bot and troll accounts generated proportionally more tweets about vaccines. We derived estimates of vaccine tweet frequencies for each account type from the vaccine stream, and we derived base rate estimates for average Twitter users from the 1% sample (Table B, available as a supplement to the online version of this article at http://www.ajph.org).

Are bots and trolls more likely to tweet polarizing and antivaccine content? We next tested the hypothesis that bots and trolls produced higher proportions of polarizing material. Three of the authors (A. M. J., S. Q., and L. A.) coded relevant tweets as "provaccine," "antivaccine," or "neutral" using a codebook developed by 1 of the authors (A. M. J.). When coders disagreed, we employed a second round of annotation. We resolved any remaining disagreements by a fourth annotator (D. A. B.). We compared all users' proportions of polarized (nonneutral) tweets to users with bot scores below 20% (likely humans). We also tested the hypothesis that nonneutral content posted by bots and trolls was more likely to be antivaccine by comparing the relative proportions of polarized tweets that were antivaccine across all user types. We assessed all hypothesis tests for statistical significance using the distribution-free χ^2 goodness of fit test.

Thematic analysis of tweets by Russian trolls. During annotation, an unfamiliar hashtag, #VaccinateUS, appeared 25 times in tweets posted by known Russian troll accounts identified by NBC News and documenting Russian interference in the US political system.[20] Searching Twitter on February 20, 2018, we found only 5 messages including this hashtag, suggesting that #VaccinateUS had been primarily used by suspended accounts and that most instances had been purged. Turning to data stored in the vaccine stream, we identified 253 messages with #VaccinateUS. We conducted an exploratory thematic analysis of these messages to identify and describe major themes. Our goal was to explore unifying patterns in the #VaccinateUS data[30] and to illustrate some of the behaviors that known Russian trolls exhibit on Twitter. Consistent with this aim, we annotated messages as pro- or antivaccine. Next, 1 author (A. M. J.) categorized messages into pro- and antivaccine themes using a combination of inductive and deductive codes.[31] We determined these categories loosely from existing research,[12] and we incorporated emergent themes in the data into them. We compared these tweets with the randomly selected vaccine-relevant tweets we used in the second analysis, which we considered representative of the general vaccine discourse.

RESULTS

Raw counts of tweets by source are shown in Table C (available as a supplement to the

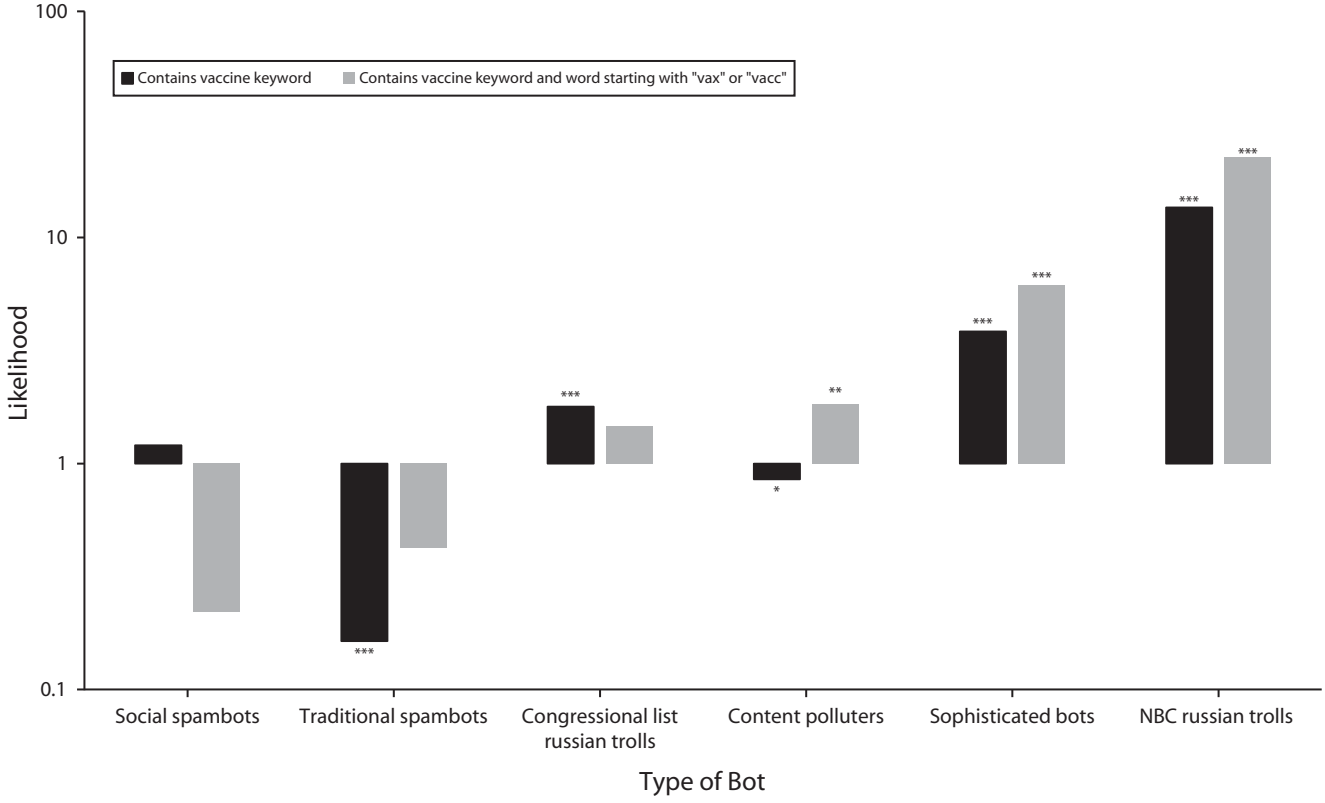
online version of this article at http://www.ajph.org). Figure 1 shows that accounts identified by NBC News as Russian trolls[20] or by Varol et al. as sophisticated bots[25] or content polluters[21] (i.e., accounts that disseminate malware and unsolicited content) are significantly more likely to tweet about vaccination than are average Twitter users. Additionally, accounts the US Congress identifies as Russian trolls[26] were significantly more likely to tweet about vaccine-preventable illnesses (e.g., Zika) but not necessarily about vaccines. Finally, traditional spambots[23,24] (designed to be recognizable as bots) and content polluters were significantly less likely to tweet about vaccine-preventable illnesses than was the average Twitter user.

Antivaccine Content

We collected 899 tweets from the vaccine stream, representing the activity of known bots and trolls. Annotators achieved moderate agreement on the first round of annotation of these tweets (Fleiss $\kappa = 0.48$). In addition, we collected 9895 tweets from the vaccine stream, representing the activity of assorted Twitter users, also with moderate initial agreement between annotators (Fleiss $\kappa = 0.49$). In all cases, annotators reached consensus after 2 more rounds. We segmented these tweets into 3 subsets: 450 (5%) tweets possessing Botometer scores of 20% or lower, 290 (3%) tweets possessing scores of 80% or higher, and 7518 (76%) tweets possessing intermediate scores. A total of 1587 (16%) tweets were associated with users whose scores could not be determined (e.g., because their accounts had been deleted).

One strategy used by bots and trolls is to generate several tweets about the same topics, with the intention of flooding the discourse.[17] Thus, to better understand the behavior of each type of account, we examined the total proportion of tweets that were generated by unique users—a possible indicator of bot- or troll-like behavior—to assess whether accounts with higher bot scores exhibited such behavior. Figure 2 shows that accounts with intermediate bot scores posted more tweets per account overall. Similarly, intermediate-scored accounts posted significantly more polarized and neutral tweets per account; however, their rates of provaccine tweets did not differ significantly from nonbots' after controlling for multiple comparisons. By contrast, accounts with high bot scores posted more neutral, but not polarized, tweets per account.

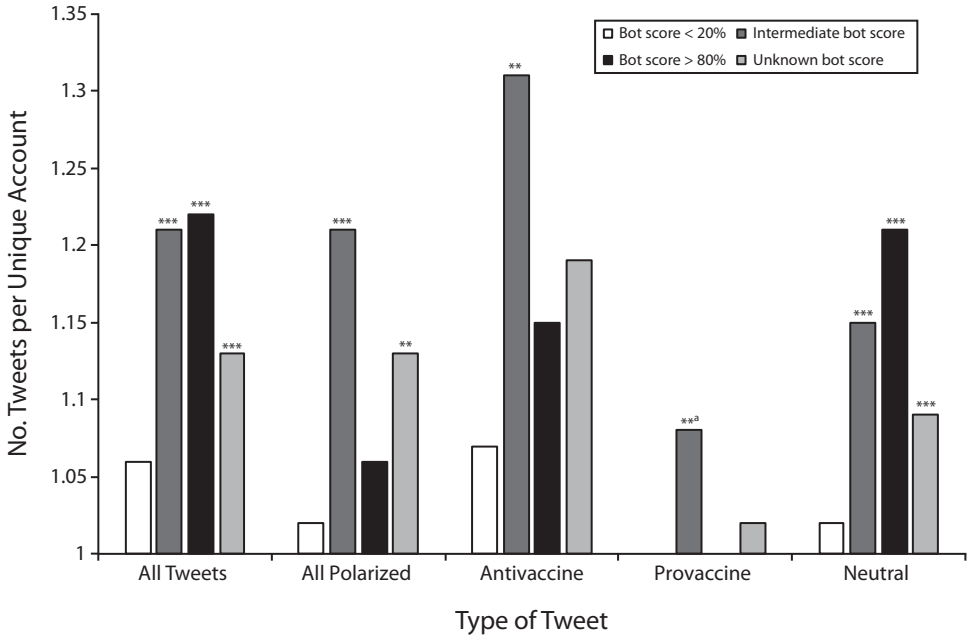
Table 1 shows the proportions of polarized and antivaccine messages by user type. Results show that accounts with intermediate bot scores post content that is significantly more polarized than are nonbots' posts. These accounts, and those whose bot scores could not be determined, posted antivaccine content at a significantly higher rate than did nonbots.



Note. NBC = National Broadcasting Network. All results remained significant after controlling for multiple comparisons using the Holm–Bonferroni procedure. Raw counts are given in Table B (available as a supplement to the online version of this article at http://www.ajph.org).

*$P < .05$; **$P < .01$; ***$P < .001$.

FIGURE 1—Bots' Likelihood of Tweeting About Vaccines Compared With Average Twitter Users: July 14, 2014–September 26, 2017



[a]Not significant after controlling for multiple comparisons using the Holm–Bonferroni procedure. Raw counts are given in Table D (available as a supplement to the online version of this article at http://www.ajph.org).

**$P < .01$; ***$P < .001$.

FIGURE 2—Number of Tweets per Unique Account, Separated by Sentiment and Bot Score Category: July 14, 2014–September 26, 2017

By contrast, known bots and trolls posted messages that were no more polarized than the messages of average nonbot users. Content polluters—malicious accounts identified as promoting commercial content and malware—posted significantly more antivaccine content. Troll accounts and sophisticated bots posted roughly equal amounts of pro- and antivaccine content.

Qualitative Analysis of #VaccinateUS

Of the 253 messages containing #VaccinateUS, 43% were provaccine, 38% were antivaccine, and the remaining 19% were neutral. Whereas most nonneutral vaccine-relevant hashtags were clearly identifiable as either provaccine (#vaccineswork, #vaxwithme) or antivaccine (#Vaxxed, #b1less, #CDCWhistleblower), with limited appropriation by the opposing side, #VaccinateUS is unique in that it appears with very polarized messages on both sides, without other nonneutral hashtags.

Messages containing #VaccinateUS contain a combination of grammatical errors, unnatural word choices, and irregular phrasing. However, the #VaccinateUS messages contain fewer spelling and punctuation errors than do comparable tweets from the general vaccine stream. The #VaccinateUS messages are also distinctive in that they contain no links to outside content, rare @mentions of other users, and no images (but occasionally use some emojis).

Thematically, the messages with #VaccinateUS mirror the general vaccine discourse on Twitter (the box on page e6). Although the authors of these tweets have a fairly comprehensive understanding of the content of both pro- and antivaccine arguments, small differences set the messages apart. The authors of #VacccinateUS messages tend to tie both pro- and antivaccine messages explicitly to US politics and frequently use emotional appeals to "freedom," "democracy," and "constitutional rights." By contrast, other tweets from the vaccine stream focus more on "parental choice" and specific vaccine-related legislation.

Like other antivaccine tweets, antivaccine messages with #VaccinateUS often reference conspiracy theories. However, whereas conspiracy theories tend to target a variety of culprits (e.g., specific government agencies, individual philanthropists, or secret organizations), the #VaccinateUS messages are almost singularly focused on the US government (e.g., "At first our government creates diseases then it creates #vaccines.what's next?! #VaccinateUS"). In general, users of #VaccinateUS talk in generalities and fail to provide the level of detail commensurate with what is found in other vaccine-relevant tweets. For example, the author might summarize an argument (e.g., "#VaccinateUS #vaccines cause serious and sometimes fatal side effects"), whereas tweets from the vaccine stream would typically use as many specifics as possible to sound convincing.

#VaccinateUS messages included several distinctive arguments that we did not observe in the general vaccine discourse. These included arguments related to racial/ethnic divisions, appeals to God, and arguments on the basis of animal welfare. These are divisive topics in US culture, which we did not see frequently discussed in other tweets related to vaccines. For instance, "Apparently only the elite get 'clean' #vaccines. And what do we, normal ppl, get?! #VaccinateUS" appears to target socioeconomic tensions that exist in the United States. By contrast, standard antivaccine messages tend to characterize vaccines as risky for all people regardless of socioeconomic status.

TABLE 1—Proportions of Polarized and Antivaccine Tweets by User Type: July 14, 2014–September 26, 2017

User Type	Polarized, %	Antivaccine, %
Assorted users, bot score, %		
< 20	31	35
20–80	39***	60***
> 80	26	49*,[a]
Unknown	37*,[a]	62***
Known bots and trolls		
NBC Russian trolls[20]	20*,[a]	47
Content polluters[21]	38	60***
Fake followers[22]	0	NA
Traditional spambots[23,24]	3***	0
Social spambots[23,24]	18**	56*,[a]
Sophisticated bots[25]	28	44
Congressional list Russian trolls[26]	39	48

Note. NA = not applicable because of insufficient data; NBC = National Broadcasting Network. A statistically significant result indicates that a certain type of account posts polarized or antivaccine tweets at a rate that differs significantly from that of accounts with bot scores < 20% (likely humans). Polarized proportion is the ratio of all nonneutral tweets to all tweets. Antivaccine proportion is the ratio of antivaccine tweets to polarized tweets. Raw counts are shown in Table E (available as a supplement to the online version of this article at http://www.ajph.org).

[a]No longer significant after controlling for multiple comparisons using the Holm–Bonferroni procedure.

*P < .05; **P < .01; ***P < .001.

#VaccinateUS messages also include several messages that seem designed to provoke a response and prolong an argument, including open-ended items and comments on the debate itself (e.g., "I believe in #vaccines, why don't you? #VaccinateUS"). Comments were also used to bait other users into responding, specifically by posting content that advocates of vaccination would take for granted, such as "#VaccinateUS Major medical organizations state that #vaccines are safe" and "#vaccine injuries are rare, despite parental worrying #VaccinateUS."

DISCUSSION

Results suggest that Twitter bots and trolls have a significant impact on online communications about vaccination. The nature of this impact differs by account type.

Russian Trolls

Russian trolls and sophisticated Twitter bots post content about vaccination at significantly higher rates than does the average user. Content from these sources gives equal attention to pro- and antivaccination arguments. This is consistent with a strategy of promoting discord across a range of controversial topics—a known tactic employed by Russian troll accounts.[20,26] Such strategies may undermine the public health: normalizing these debates may lead the public to question long-standing scientific consensus regarding vaccine efficacy.[13] Indeed, several antivaccine arguments claim to represent both sides of the debate—like the tactics used by the trolls identified in this study—while simultaneously communicating a clear gist (i.e., a bottom-line meaning). We recently found that this strategy was effective for propagating news articles through social media in the context of the 2015 Disneyland measles outbreak.[32]

Commercial and Malware Distributors

Unlike troll accounts, content polluters (i.e., disseminators of malware, unsolicited commercial content, and other disruptive material that typically violates Twitter's terms of service)[21] post antivaccine messages 75% more often than does the average nonbot Twitter user. This suggests that vaccine opponents may disseminate messages using bot networks that are primarily designed for marketing. By contrast, spambots,[3,4] which can be easily recognized as nonhuman, are less likely to promote an antivaccine agenda than are nonbots. Notably, content polluters and traditional spambots are both less likely to discuss vaccine-preventable illnesses than is the average Twitter user, suggesting that when they do tweet vaccine-relevant messages, their specific focus is on vaccines per se, rather than the viruses that require them. Thus, it is unclear to what extent their promotion of vaccine-related content is driven by true antivaccine sentiment or is used as a tactic designed to drive up click-through rates by propagating motivational content ("clickbait").

Unidentified Accounts

Several accounts could not be positively identified as either bots or humans because of intermediate or unavailable Botometer scores. These accounts, together constituting 93% of our random sample from the vaccine stream, tweeted content that was both more polarized and more opposed to vaccination than is that of the average nonbot account. Although the provenance of their tweets is unclear, we speculate that these accounts may possess a higher proportion of trolls or cyborgs—accounts nominally controlled by human users that are, on occasion, taken over by bots or otherwise exhibit bot-like or malicious behavior.[15] Cyborg accounts are more likely to fall into this middle range because they only display bot-like behaviors sometimes. This middle range is also likely to contain tweets from more sophisticated bots that are designed to more closely mimic human behaviors.

Finally, trolls—exhibiting malicious behaviors yet operated by humans—are also likely to fall within this middle range. This suggests that proportionally more antivaccine tweets may be generated by accounts using a somewhat sophisticated semiautomated approach to avoid detection. This creates the false impression of grassroots debate regarding vaccine efficacy—a technique known as "astroturfing"[17] (as in the #VaccineUS tweets shown in the box on page e6). There are certainly standard human accounts that also fall within this middle range. Although technological limitations preclude us from drawing definitive conclusions about these account types, the fact that middle-range tweets tend to post proportionately more antivaccine messages suggests strongly that these antivaccine messages may be disseminated at higher rates by a combination of malicious actors (bots, trolls, cyborgs, and human users) who are difficult to distinguish from one another.

This interpretation is supported by the fact that users within this intermediate range tended to produce more tweets, and especially antivaccine tweets, per account, suggesting that antivaccine activists may preferentially use these channels. In addition, users whose accounts had been deleted posted more polarized messages per user and were also significantly more likely to post

EXAMPLES OF TWEETS WITH #VACCINATEUS AND CORRESPONDING THEMES: JULY 14, 2014–SEPTEMBER 26, 2017

Antivaccine theme	Example tweet
Freedom of choice/antimandatory vaccines	VaccinateUS mandatory #vaccines infringe on constitutionally protected religious freedoms
Can't trust government on vaccines	Did you know there was a secret government database of #vaccine-damaged children? #VaccinateUS
Pharmaceutical companies want vaccine profits	Pharmacy companies want to develop #vaccines to cash, not to prevent deaths #VaccinateUS
Vaccines cause bad side effects	#VaccinateUS #vaccines can cause serious and sometimes fatal side effects
Natural immunity is better	#VaccinateUS natural infection almost always causes better immunity than #vaccines
General vaccine conspiracy theories	Dont get #vaccines. Iluminati are behind it. #VaccinateUS
Vaccines cause autism	Did you know #vaccines caused autism? #VaccinateUS
Vaccine ingredients are dangerous	#VaccinateUS #vaccines contain mercury! Deadly poison!
Diseases aren't so dangerous	#VaccinateUS most diseases that #vaccines target are relatively harmless in many cases, thus making #vaccines unnecessary
Provaccine theme	**Example tweet**
Vaccines work	#VaccinateUS #vaccines save 2.5 million children from preventable diseases every year
Vaccines should be mandatory	Your kids are not your property! You have to #vaccinate them to protect them and all the others! #VaccinateUS
People who don't vaccinate are stupid	#VaccinateUS You can't fix stupidity. Let them die from measles, and I'm for #vaccination!
Vaccination protects herd immunity	#VaccinateUS #vaccines protect community immunity
People who don't vaccinate put me/my kids at risk	#VaccinateUS My freedom ends where another person's begins. Then children should be #vaccinated if disease is dangerous for OTHER children
Vaccines don't cause autism	#vaccines cause autism—Bye, you are not my friend anymore. And try to think with your brain next #VaccinateUS
You deserve bad things if you don't vaccinate	#vaccines are a parent's choice. Choice of a color of a little coffin #VaccinateUS
Alternative medicine doesn't work	Do you still treat your kids with leaves? No? And why don't you #vaccinate them? Its medicine! #VaccinateUS
People died without vaccines	Most parents in Victorian times lost children regularly to preventable illnesses. #vaccines can solve this problem #VaccinateUS

antivaccine messages. Although reasons for account deletion vary, recent movement by Twitter to remove bots,[33,34] trolls,[20] cyborgs, and other violators of Twitter's terms of service suggests that these violators may be overrepresented among the deleted accounts in our sample. We cannot claim that all, or even most, accounts with unknown bot scores are malicious actors; however, we expect that a higher proportion of malicious actors are present in this subset of the data. By contrast, randomly sampled accounts that were easily identifiable as bots generated more neutral, but not polarized tweets per account. Presumably, accounts that are obviously automated are more frequently used to disseminate content such as news and may not be considered credible sources of grassroots antivaccine information.

Public Health Implications

Survey data show a general consensus regarding the efficacy of vaccines in the general population.[35] Consistent with these results, accounts unlikely to be bots are significantly less likely to promote polarized and antivaccine content. Nevertheless, bots and trolls are actively involved in the online public health discourse, skewing discussions about vaccination. This is vital knowledge for risk communicators, especially considering that neither members of the public nor algorithmic approaches may be able to easily identify bots, trolls, or cyborgs.

Malicious online behavior varies by account type. Russian trolls and sophisticated bots promote both pro- and antivaccination narratives. This behavior is consistent with a strategy of promoting political discord. Bots and trolls frequently retweet or modify content from human users. Thus, well-intentioned posts containing provaccine content may have the unintended effect of "feeding the trolls," giving the false impression of legitimacy to both sides, especially if this content directly engages with the antivaccination discourse. Presuming bot and troll accounts seek to generate roughly equal numbers of tweets for both sides, limiting access to provaccine content could potentially also reduce the incentive to post antivaccine content.

By contrast, accounts that are known to distribute malware and commercial content are more likely to promote antivaccination messages, suggesting that antivaccine advocates may use preexisting infrastructures of bot networks to promote their agenda. These accounts may also use the compelling nature of antivaccine content as clickbait to drive up advertising revenue and expose users to malware. When faced with such content, public health communications officials may consider emphasizing that the credibility of the source is dubious and that users exposed to such content may be more likely to encounter malware. Antivaccine content may increase the risks of infection by both computer and biological viruses.

The highest proportion of antivaccine content is generated by accounts with unknown or intermediate bot scores. Although we speculate that this set of accounts contains more sophisticated bots, trolls, and cyborgs, their provenance is ultimately unknown. Therefore, beyond attempting to prevent bots from spreading messages over social media, public health practitioners should focus on combating the messages themselves while not feeding the trolls. This is a ripe area for future research, which might include emphasizing that a significant proportion of antivaccination messages are organized "astroturf" (i.e., not grassroots) and other bottom-line messages that put antivaccine messages in their proper contexts. AJPH

CONTRIBUTORS

D. A. Broniatowski designed the study, conducted the statistical analyses, and wrote the first draft of the article. A. M. Jamison conducted the qualitative analysis. A. M. Jamison, S. Qi, and L. Alkulaib conducted the sentiment coding. A. M. Jamison, S. Qi, L. Alkulaib, T. Chen, A. Benton, S. C. Quinn, and M. Dredze critically revised the article. S. Qi, L. Alkulaib, T. Chen, and A. Benton collected and analyzed Twitter data. S. C. Quinn and M. Dredze assisted with study design.

ACKNOWLEDGMENTS

This research was supported by the National Institute of General Medical Sciences, National Institutes of Health (NIH; award 5R01GM114771).

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

HUMAN PARTICIPANT PROTECTION

The data used in this article are from publicly available online sources, the uses of which the Johns Hopkins Homewood institutional review board deems exempt from institutional review board approval (approval no. 2011123).

REFERENCES

1. 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.

2. Breland JY, Quintiliani LM, Schneider KL, May CN, Pagoto S. Social media as a tool to increase the impact of public health research. *Am J Public Health*. 2017;107(12):1890–1891.

3. Luxton DD, June JD, Fairall JM. Social media and suicide: a public health perspective. *Am J Public Health*. 2012;102(suppl 2):S195–S200.

4. Witteman HO, Zikmund-Fisher BJ. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. *Vaccine*. 2012;30(25):3734–3740.

5. Quinn SC. Probing beyond individual factors to understand influenza and pneumococcal vaccine uptake. *Am J Public Health*. 2018;108(4):427–429.

6. Betsch C, Brewer NT, Brocard P, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine*. 2012;30(25):3727–3733.

7. 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.

8. Smith MJ, Marshall GS. Navigating parental vaccine hesitancy. *Pediatr Ann*. 2010;39(8):476–482.

9. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert Rev Vaccines*. 2015;14(1):99–117.

10. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS One*. 2014;9(2):e89177.

11. Jones AM, Omer SB, Bednarczyk RA, Halsey NA, Moulton LH, Salmon DA. Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions. *Adv Prev Med*. 2012;2012:932741.

12. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. *Vaccine*. 2010;28(7):1709–1716.

13. Dixon G, Clarke C. The effect of falsely balanced reporting of the autism-vaccine controversy on vaccine safety perceptions and behavioral intentions. *Health Educ Res*. 2013;28(2):352–359.

14. Centers for Disease Control and Prevention. Mortality. 2017. Available at: https://www.cdc.gov/nchs/health_policy/mortality.htm. Accessed April 27, 2018.

15. Chu Z, Gianvecchio S, Wang H, Jajodia S. Detecting automation of Twitter accounts: are you a human, bot, or cyborg? *IEEE Trans Depend Secure Comput*. 2012;9(6):811–824.

16. Collins English Dictionary. Troll definition and meaning. Available at: https://www.collinsdictionary.com/dictionary/english/troll. Accessed April 27, 2018.

17. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. *Commun ACM*. 2016;59(7):96–104.

18. Subrahmanian VS, Azaria A, Durst S, et al. The DARPA Twitter bot challenge. *Computer*. 2016;49(6):38–46.

19. Ward JK, Peretti-Watel P, Verger P. Vaccine criticism on the Internet: propositions for future research. *Hum Vaccin Immunother*. 2016;12(7):1924–1929.

20. Popken B. Twitter deleted Russian troll tweets. So we published more than 200,000 of them. Available at: https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731. Accessed March 11, 2018.

21. Lee K, Eoff BD, Caverlee J. Seven months with the devils: a long-term study of content polluters on Twitter. Available at: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2780. Accessed March 11, 2018.

22. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M. Fame for sale: efficient detection of fake Twitter followers. *Decis Support Syst*. 2015;80:56–71.

23. Cresci S, Pietro RD, Petrocchi M, Spognardi A, Tesconi M. Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. *IEEE Trans Dependable and Secure Comput*. 2018;15(4):561–576.

24. Cresci S, Di Pietro R, Petrocchi M, Spognardi A. The paradigm-shift of social spambots: evidence, theories, and tools for the arms race. In: Barrett R; International World Wide Web Conferences Steering Committee, eds. *Proceedings of the 26th International Conference on World Wide Web Companion*.Republic and Canton of Geneva: ACM Press; 2017:963–972.

25. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human–bot interactions: detection, estimation, and characterization. Available at: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587. Accessed March 11, 2018.

26. Frommer D. Twitter's list of 2,752 Russian trolls. 2017. Available at: https://www.recode.net/2017/11/2/16598312/russia-twitter-trump-twitter-deactivated-handle-list. Accessed March 11, 2018.

27. 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.

28. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A System to evaluate social bots. Available at: http://dl.acm.org/citation.cfm?doid=2872518.2889302. Accessed July 25, 2018.

29. Wojcik S, Messing S, Smith A, Rainie L, Hitlin P. Bots in the Twittersphere. 2018. Available at: http://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere. Accessed April 26, 2018.

30. Sandelowski M, Barroso J. Classifying the findings in qualitative studies. *Qual Health Res*. 2003;13(7):905–923.

31. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int J Qual Methods*. 2006;5(1):80–92.

32. Broniatowski DA, Hilyard KM, Dredze M. Effective vaccine communication during the Disneyland measles outbreak. *Vaccine*. 2016;34(28):3225–3228.

33. Flynn K. Twitter influencers suspect a bot "purge." Available at: https://mashable.com/2018/01/29/twitter-bots-purge-influencers-accounts. Accessed April 27, 2018.

34. Madrak S. Wingers melt down as twitter finally purges Russian bots. Available at: https://crooksandliars.com/2018/02/wingers-have-sad-twitter-purges-russian-0. Accessed April 27, 2018.

35. Funk C, Kennedy B, Hefferon M. Vast majority of Americans say benefits of childhood vaccines outweigh risks. 2017. Available at: http://www.pewinternet.org/2017/02/02/vast-majority-of-americans-say-benefits-of-childhood-vaccines-outweigh-risks. Accessed February 14, 2017.