Facebook and Twitter vaccine sentiment in response to measles outbreaks

Michael S Deiner
University of California, San Francisco, USA

Cherie Fathy
Vanderbilt University, USA

Jessica Kim
University of California, San Francisco, USA

Katherine Niemeyer
Icahn School of Medicine at Mount Sinai, USA

David Ramirez, Sarah F Ackley, Fengchen Liu, Thomas M Lietman and Travis C Porco
University of California, San Francisco, USA

Abstract

Social media posts regarding measles vaccination were classified as pro-vaccination, expressing vaccine hesitancy, uncertain, or irrelevant. Spearman correlations with Centers for Disease Control and Prevention–reported measles cases and differenced smoothed cumulative case counts over this period were reported (using time series bootstrap confidence intervals). A total of 58,078 Facebook posts and 82,993 tweets were identified from 4 January 2009 to 27 August 2016. Pro-vaccination posts were correlated with the US weekly reported cases (Facebook: Spearman correlation 0.22 (95% confidence interval: 0.09 to 0.34), Twitter: 0.21 (95% confidence interval: 0.06 to 0.34)). Vaccine-hesitant posts, however, were uncorrelated with measles cases in the United States (Facebook: 0.01 (95% confidence interval: −0.13 to 0.14), Twitter: 0.0011 (95% confidence interval: −0.12 to 0.12)). These findings may result from more consistent social media engagement by individuals expressing vaccine hesitancy, contrasted with media- or event-driven episodic interest on the part of individuals favoring current policy.

Corresponding author:
Travis C Porco, Francis I. Proctor Foundation for Research in Ophthalmology, University of California, San Francisco, Box 0412, San Francisco, CA 94143-0412, USA.
Email: travis.porco@ucsf.edu

Introduction

Measles results from infection due to an RNA virus in the genus *Paramyxoviridae*. Transmitted by the respiratory route, measles is one of the most contagious diseases known.[1] Most individuals in the developed world recover without complications, and the death rate of 1–3 per 1000 is largely due to measles pneumonia.[2] The disease causes fever, cough, and rash,[3,4] following an incubation period of 8–12 days.[3,5,6] Subacute sclerosing panencephalitis is a rare but fatal and untreatable complication.[2,7] Measles infection normally confers lifelong immunity.[8]

Vaccination with the measles, mumps, and rubella (MMR) vaccine confers long-lasting protection, which may be life-long.[9] The two-dose MMR schedule shows an efficacy of approximately 94 percent.[10] High vaccination rates achieve herd immunity and are the basis for measles elimination.[8] The World Health Organization estimates that measles vaccination has prevented 20.3 million deaths worldwide during 2000–2015.[11] In the United States, although measles was declared locally eliminated,[12] it is repeatedly reintroduced. In late 2014 to early 2015, a measles outbreak linked to Disneyland theme parks caused approximately 130 cases in California alone; it spread to other states in the United States, as well as Canada and Mexico.[13,14]

Measles transmission in recent years has been linked to vaccine refusal.[15] Opposition to preventive inoculation is hardly new,[16,17] although in recent years a now-retracted[18] study led by A. Wakefield may have reinvigorated anti-vaccination controversy.[19,20] Significant media attention to this research and personal anecdotes from celebrities may have caused a sense of fear and distrust in the vaccine among some individuals, potentially contributing to a decrease in vaccination rates in some locations.[19,21,22] Despite these real concerns about falling vaccination rates, overall US national levels of MMR vaccine coverage appear to have remained high in recent years.[23,24]

Social media sites may now have become a reflection of the beliefs and practices of people in their daily lives.[25] Use of social media is extensive; Twitter reports more than 300 million active users globally in 2016[26] and significant numbers of online American adults interviewed by the Pew Research Center endorsed using Twitter in 2015 (23%) and Facebook (FB; 72%). Furthermore, Twitter and FB provide large amounts of publicly accessible data over a number of years, from very large user bases.

Social media have had a large and growing role in health communication.[27–30] Public health and research organizations have also turned to social media as a form of digital public health surveillance to augment traditional methods of risk or disease monitoring[31–38] or as a communication tool.[39–41] A growing number of non-traditional sources for health information, including blogs and alternative news sites, are available online.[42–46]

Examination of social media can provide insight into public discussion of vaccination[47] and may be predictive of geographical variation in vaccination compliance.[48,49] While Internet communication has played a large role in vaccine opposition for a number of years,[50] social media provide an outlet to share beliefs widely in real time.[42,51,52] Individuals opposing vaccination have also made substantial use of social media and emerging communication technologies.[51,53,54]

Previous studies examining social media and public discourse on measles vaccinations have focused on posts during an outbreak[35,51] and predictors of which articles are shared on social media.[55] Others have found differences in linguistic usage between pro- and anti-vaccination posts in certain settings.[56] Other studies have focused on vaccines other than MMR.[52,57] An important resource for understanding vaccine sentiment worldwide is provided by the recently introduced

Vaccine Sentimeter,[47] which used the public Twitter API and other sources of information (including news sources and blogs), to yield an index of vaccine sentiment (positive, negative, and neutral/unclear). This tool reports data from May 2012 to November 2014.[58] Previous studies have also used machine learning classifiers to simply explore and describe the Twitter conversation during a disease outbreak.[59] Other studies have used social media to relate human mobility to disease transmission,[60] to forecast disease incidence during an outbreak,[61,62] and to track the spread of rumor and perceived risk during an outbreak.[63]

In this report, we examine FB and Twitter social media discussion of vaccination in relation to measles from 2009 to mid-2016, a period including several widely publicized outbreaks. We related volume and machine-learned vaccine-related stance[64] to Centers for Disease Control and Prevention (CDC)-reported weekly measles counts. Vaccine hesitancy refers to "delay in acceptance or refusal of vaccination despite availability of vaccination services."[65] We conjectured that individuals who do not express vaccine hesitancy (who are in favor of current practice) become drawn into the debate during actual outbreaks of measles (which may, for example, be perceived as a threat). Accordingly, we hypothesized that the number of posts exhibiting a given stance—*pro-vaccination* (essentially in favor of current practice), or *vaccine hesitant*—expressed in FB posts and on Twitter regarding vaccinations would differ depending on whether a measles outbreak was occurring in real time or not.

Methods

Overview

We constructed a list of keywords regarding opinion about vaccinations or measles and used a commercial social media analytics platform[66,67] (Crimson Hexagon) to collect FB posts and tweets, from 2009 to mid-2016, that contained these keywords. We trained the BrightView classifier provided by the platform to group FB posts and tweets into four classes: pro-vaccination, expressing vaccine hesitancy, on-topic but unclear, and off-topic. The BrightView classifier is a supervised learning algorithm based, in part, on stacked regression analysis of simplified numerical representation of text.[68] These machine-coded classifications were assessed by comparison with human classifications of a random sample of FB posts and tweets. Finally, we compared the volume of FB posts and tweets to CDC reports of daily measles counts.

Data collection: social media

FB and Twitter data. FB and Twitter data yield spatiotemporal patterns for measles/vaccination posts of interest. Using Boolean search, we searched for FB posts and tweets related to vaccination (the query is found in Appendix 1). To focus posts expressing the original stance of individuals, retweets and posts with URLs (or those likely to be advertisements) were filtered out prior to our primary analyses. To compare those results to the broader conversation, such posts were included in subsequent analysis. Along with the post contents, we collected the date and time the posts were published as well as metadata regarding the Twitter and FB accounts submitting the posts.

Using the primary corpus (having already removed retweets, for example), we assessed the effect of highly repetitive tweets as follows. We computed the edit distance[69] for all pairs of unique tweets found in our search. From this, for each tweet, we identified all tweets in our corpus within an edit distance of 20 for each tweet (chosen to allow posts to differ by, for instance, a short URL or user name). For each author, we assessed whether that author had emitted at least one tweet or

not to which there corresponded another tweet in the corpus which was within a distance of 20 (whether by the same or by an additional author).

Data collection: measles cases. Measles data were acquired from the CDC Morbidity and Mortality Weekly Report (MMWR) website, tabulated *MMWR Reports of infrequently reported notifiable diseases*, for years 2009–2016. The period covered by our study was 4 January 2009–27 August 2016. US measles data are presented by the CDC as (1) weekly reported cases and (2) cumulative to date. Cases are diagnosed based on laboratory findings or confirmed epidemiologic link to a laboratory-confirmed case. Methods of laboratory confirmation include viral isolation, detection of measles-virus specific nucleic acid, IgG seroconversion (or a significant rise), or a positive sero-logic test for measles immunoglobulin M antibody.[70] Since initial samples are ideally collected within the first few days following rash onset[71–77] and possibly within a few weeks,[72,73] diagnosis-related delays longer than this are unlikely. The cumulative column reported by the CDC contains additional cases not reported in the weekly totals, since cases newly reported from past weeks are added to the cumulative total. Thus, neither of these reporting streams provide a complete count of all cases occurring in a given week. For the cumulative series, we removed outliers (reporting error) using a 5-week centered median filter[78] and then computed the final differenced series (number of new cases added to the cumulative total each week).

Stance analysis

Using the built-in BrightView classifier,[66] the media analytics platform was also trained to separate FB posts and tweets as (1) vaccine hesitant, (2) pro-vaccination, or (3) neither; a total of 96 posts were used for each (the minimum recommended being 25 per category). The third category was further subdivided into (3a) unclear whether pro-vaccination or vaccination hesitancy despite being related to vaccination or (3b) unrelated to vaccination. Individuals exhibiting vaccine hesitancy do not form a monolithic bloc,[65,79] and so we viewed vaccine hesitancy in a broad sense. We classified as *vaccine-hesitant* tweets expressing views including universal opposition, opposition to contents of selected vaccines or vaccination schedules, or a conviction that vaccine-related injuries occur at higher rates than commonly believed. Individuals expressing views in favor of vaccination, of current policy, or simply expressing opposition to vaccine hesitancy were classified simply as "pro-vaccination"; no evidence supports the view that "pro-vaccination" individuals form a monolithic bloc either.

To validate the training, three raters manually reviewed and categorized FB posts and tweets into the same four categories and their categorizations were compared to the machine-learned categorizations. The human categorization process consisted of each rater independently reading each tweet or FB posts and determining for each post the following: does the post either (1) clearly express vaccine hesitancy in whole or in part, (2) favor vaccination, or clearly critique vaccine hesitancy or "anti-vaxxers," (3) reference vaccination without clearly expressing or opposing vaccine hesitancy, or is it (4) irrelevant to the debate).

Reviewers first trained together on randomly selected tweets from our corpus (excluding any used in training the machine). A group discussion with all three raters followed to reconcile disagreements in the categorization and to gain familiarity with the discussion on Twitter (i.e. to become aware of the language used in these posts, as well as prominent or potentially polarizing figures in the vaccination debate). The raters subsequently completed a second iteration of the training with a smaller set of tweets followed by a brief discussion. We then provided the raters with Google sheets containing 512 randomly chosen tweets (excluding any used in training either the raters or the machine; the sample size would provide a confidence interval (CI) of half-width approximately 0.02 for estimated proportions near 0.5).

All rated the same tweets, masked to the results of other raters and that of the machine. In addition to the tweet content and URLs to the original tweet, the raters were provided the authors twitter handle (user name) and their actual name, geolocation, and gender if they were available through Crimson Hexagon. For tweets classified by the machine as expressing vaccine hesitancy by the machine, we report the fraction of such posts actually classified as vaccine hesitant by each rater; similarly, for tweets classified as pro-vaccination by the machine, we report the fraction of posts classified as pro-vaccination by the raters.

After completing tweet classifications, we conducted a similar process for the three human raters to classify a set of FB posts, masked to the classifications assigned by other raters and by the machine. Their classifications were analyzed for validation of the machine-learned categorizations. The sample size was halved, since the FB posts were substantially longer than tweets.

Statistical analysis

We report (1) the total weekly volume of identified FB posts and tweets identified as pro-vaccination, vaccine-hesitant, or neither pro- nor vaccine-hesitant (but on topic), (2) the weekly volume of tweets classified by the machine as pro-vaccination, and (3) the weekly volume of tweets classified by the machine as vaccine hesitant. For FB posts and separately for tweets, and for each of these three outcome variables, we computed the Spearman rank correlation with (a) the CDC weekly counts and (b) the first-differenced cumulative count series described above. As a sensitivity analysis, we used our classification of weeks as being in or not in an outbreak (as described in section "Methods") and report the number of FB posts and tweets, advocating and criticizing vaccine hesitancy, during outbreak and non-outbreak periods; we computed the odds ratio for association. CIs in all cases were constructed using time series bootstrap,[80] with a fixed width of 8. All calculations were conducted in R, version 3.2 for MacIntosh (R Foundation for Statistical Computing, Vienna, Austria).

Results

### Identified posts

A total of 58,078 FB posts and 82,993 tweets were identified. Of FB posts, 22,331 were known to be located in the United States. A total of 38,644 tweets were known to be located in the United States. Geolocation was unavailable for 33,335 out of 58,078 (57.40%) of FB posts and for 30,086 of 82,993 (36.25%) of tweets. The earliest post matching our query occurred on 23 June 2009.

Classification

For FB, the algorithm classified 22,306 posts as against vaccination, a total of 17,928 posts as pro-vaccination, and 2431 posts as uncertain but on topic ("neither" in the table). The machine algorithm classified a total of 43,259 tweets as against vaccination, a total of 16,297 tweets as pro-vaccination, and 12,222 tweets as uncertain but on topic ("neither" in the table). The number and stance of FB posts and tweets classified as expressing vaccine hesitancy or expressing pro-vaccination views are shown in Figures 1 and 2, respectively.

The figures also show changes in reported measles cases over the study period. Epidemics occurring during most of 2011 were reported at the time as "the worst measles year in 15 years," with 222 cases provisionally reported across 31 states.[81] Another epidemic period corresponds to epidemics across 16 states, including New York City during the spring and summer of 2013,[82] as
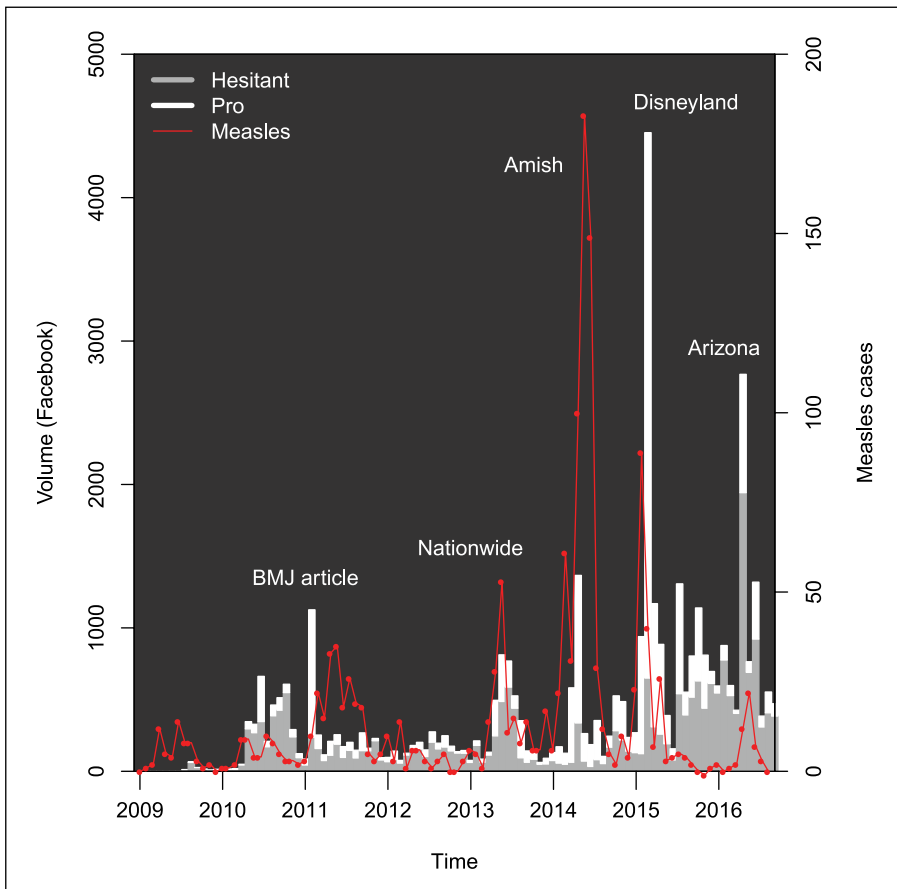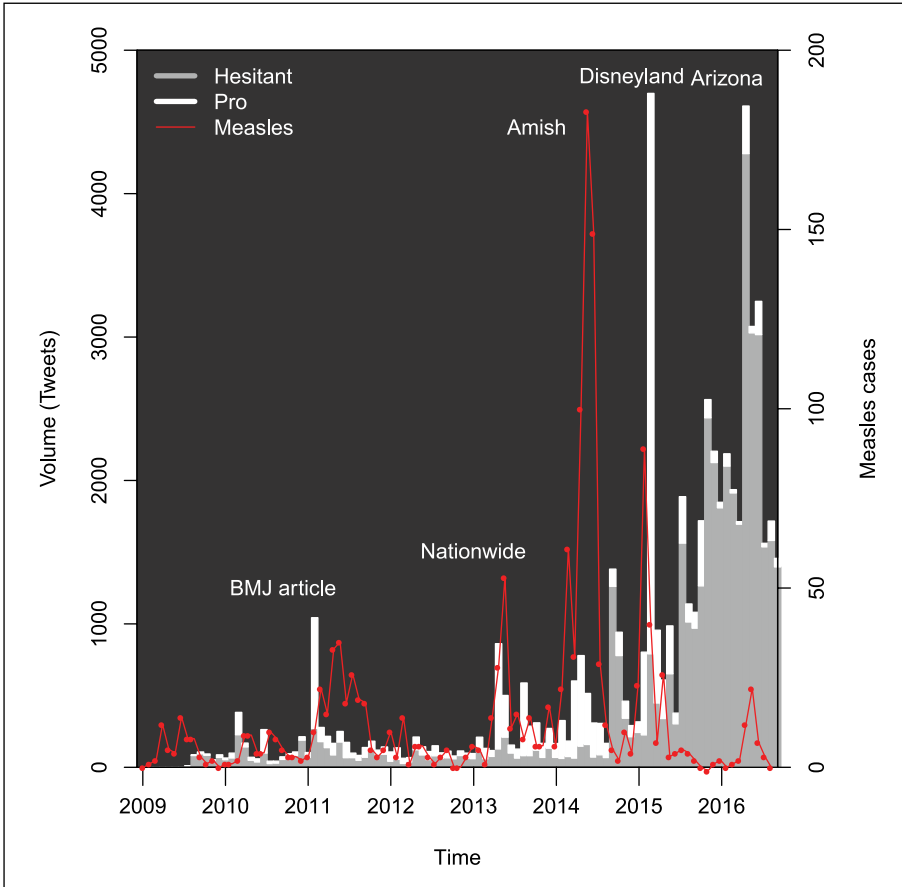
Figure 1. Facebook volume matching query, by stance, 2009–2016, according to classification by as described in the text. Measles cases are shown by the differenced cumulative series derived from CDC reports, as described in the text. Facebook volume and reported measles cases shown in consecutive periods of 4 weeks, with a first period consisting of 3 weeks. *BMJ article* indicates Deer (2011). Additional details on the measles outbreaks are given in Appendix 1.
Source: Crimson Hexagon, CDC.

well as a publicized epidemic which included a large church in Texas.[83] A third includes the outbreak in Ohio largely among Amish residents, which was also the largest outbreak post-elimination.[84] The figure also illustrates the outbreak which apparently began in Disneyland, causing approximately 130 cases in California and spreading to other states.[13] Finally, a recent epidemic period includes an outbreak in Arizona.[85]

Validation

Comparison of the machine algorithm with each human rater is shown in Table 1, and agreement of the human raters with each other is shown in Table 2. The first table shows the fraction of times the human rater agreed with a machine classification of a post as being *vaccine hesitant* versus *pro-vaccination*.

Figure 2. Tweet volume matching query, by stance, 2009–2016, according to classification by BrightView, as described in the text. Measles cases are shown by the differenced cumulative series derived from CDC reports, as described in the text. Tweet volume and reported measles cases shown in consecutive periods of 4 weeks, with a first period consisting of 3 weeks. Additional details on the measles outbreaks are given in Appendix 1.
Source: Crimson Hexagon, CDC.

Social media and measles surveillance

For the total FB post volume identified by our search and the weekly measles reports, we found that the Spearman rank correlation was 0.11 (95% CI: −0.02 to 0.22). For the number of pro-vaccination FB posts, the correlation with the weekly reports was 0.22 (95% CI: 0.09 to 0.34). For the number of FB posts expressing vaccine hesitancy, the correlation with the weekly reports was 0.01 (95% CI: −0.13 to 0.14).

For the total tweet volume identified by our search and the weekly measles reports, we found that the Spearman rank correlation was 0.06 (95% CI: −0.08 to 0.18, time series bootstrap). For the number of pro-vaccination tweets, the correlation with the weekly reports was 0.21 (95% CI: 0.06 to 0.34). For the number of tweets expressing vaccine hesitancy, the correlation with the weekly reports was 0.0011 (95% CI: −0.12 to 0.12). Correlations with the differenced cumulative series and for FB are shown in Table 3.

Table 1. Agreement of stance classification of social media posts by three human raters and automated algorithm.

Platform	Machine	Total	Rater 1 (%)	Rater 2 (%)	Rater 3 (%)
Facebook	Hesitant	92	67.4	60.9	69.6
	Pro	92	59.8	58.7	64.1
Twitter	Hesitant	250	67.2	64.4	66.4
	Pro	106	72.6	72.6	71.7

The columns show the fraction of time each of the individual raters agreed with a machine classification of *vaccine hesitant* versus *pro-vaccination*. All four possible categories were included when computing agreement.

Table 2. Agreement of human raters with one another.

Platform	Reference grader	Reference result	Rater 1 agreement (%)	Rater 2 agreement (%)	Rater 3 agreement (%)
Facebook	Rater 1	Hesitant	–	83.5	85.3
		Pro	–	89.7	97.9
	Rater 2	Hesitant	89	–	85
		Pro	91.6	–	94.7
	Rater 3	Hesitant	89.6	80.2	–
		Pro	88.8	84.1	–
Twitter	Rater 1	Hesitant	–	83.5	85.3
		Pro	–	85.2	83.5
	Rater 2	Hesitant	94.6	–	89.7
		Pro	94.5	–	87.2
	Rater 3	Hesitant	86.4	80.3	–
		Pro	86.9	81.7	–

For each rater (Column 2) who reached the result *vaccine hesitant or pro-vaccination* in Column 3, the fraction of times the other raters agreed is given in the remaining columns.

Table 3. Correlation of social media stance with reported US measles cases, 2009–2016.

Medium	Classification	Weekly measles	Differenced cumulative measles
Facebook	Total	0.11 (−0.02 to 0.22)	0.15 (−0.02 to 0.31)
	Pro	0.22 (0.09 to 0.34)	0.28 (0.13 to 0.41)
	Hesitant	0.01 (−0.13 to 0.14)	0.04 (−0.13 to 0.22)
Twitter	Total	0.06 (−0.08 to 0.18)	0.06 (−0.11 to 0.22)
	Pro	0.21 (0.06 to 0.34)	0.27 (0.11 to 0.41)
	Hesitant	0.0011 (−0.12 to 0.12)	−0.02 (−0.18 to 0.14)

Posts on Twitter and Facebook were classified by the BrightView classifier (Crimson Hexagon) as indicated in the text. Both CDC weekly reported cases and the increase in the cumulative case count were used (see text for details). Spearman rank correlations are reported, with confidence intervals derived from time series bootstrap.

We also conducted the same analysis restricting the posts to those known to be in the United States. Because geolocation information for FB was almost entirely unavailable after March 2015, we restricted the analysis of US-based FB posts to the period ending 31 March 2015. As before, we

found significant Spearman correlation between the weekly case counts and the frequency of posts classified as pro-vaccination (0.269; 95% CI: 0.17 to 0.41), but not for those posts classified as vaccine hesitant (0.0406; 95% CI: −0.14 to 0.24). Similar results were found for the frequency of tweets (for which unavailability of geolocation did not mandate restriction to the period ending 31 March 2015); the Spearman correlation between the weekly case counts and the frequency of tweets classified as pro-vaccination was 0.196 (95% CI: 0.06 to 0.34), while for those classified as vaccine hesitant, we found 0.00111 (95% CI: −0.12 to 0.12).

In the analyses above, we used a dataset filtered to focus on individuals' original expressions (that excluded retweets, URLs, and presumed advertisements). To compare results of that more focused corpus to the broader conversation, we conducted a sensitivity analysis based on a broader query in which we did not remove retweets dollar signs or URLs (see Appendix 1 for details). For this broader query, similar results were obtained, although corpus volumes were larger. We found a Spearman correlation between weekly case counts and non-vaccine-hesitant tweets of 0.173 (95% CI: 0.05 to 0.28). The Spearman correlation between weekly case counts and non-vaccine-hesitant tweets was 0.0804 (95% CI: −0.05 to 0.2). Similar findings were obtained for FB and for the differenced cumulative series (not shown). In all cases, non-vaccine-hesitant tweets and posts showed a greater relationship with measles case counts than was seen for vaccine-hesitant tweets.

As a further sensitivity analysis, we also examined the proportion of pro-vaccination tweets and tweets expressing vaccine hesitancy in the weeks we classified as being in an outbreak or not being in an outbreak. The classification of weeks as corresponding to an outbreak or not is provided in Appendix 1, based on the two CDC measles time series. For FB, the odds ratio for a given post being pro versus hesitant during an outbreak was 3.2 compared with a non-outbreak week (95% CI: 1.21 to 6.62). We found that for Twitter, the odds ratio for a given tweet being pro versus hesitant during an outbreak was 4.3 compared with a non-outbreak week (95% CI: 1.28 to 12.2). We also used the two series to define epidemic or outbreak periods nationally; details are provided in Appendix 1. Note that in general, an "epidemic" or "outbreak" simply indicates more than the usual number of cases, and no unambiguous definition is possible. Topic wheels for vaccine-hesitant and non-vaccine-hesitant Twitter posts prior to and during the 2014–2015 outbreak are provided in Appendix 1.

Excluding the year 2009, for which there is a low volume, we found an overall Spearman rank correlation of the total on-topic tweets with the weekly reported cases of 0.04 (95% CI: −0.12 to 0.19). We found an overall Spearman rank correlation of the total pro-vaccination tweets with the weekly reported cases of 0.2 (95% CI: 0.04 to 0.34), while the Spearman rank correlation of the total vaccine-hesitant tweets with the weekly reported cases was −0.05 (95% CI: −0.2 to 0.09).

Additionally, we computed the Spearman correlation of the measles time series with total tweet volume matching our query, tweets classified as pro-vaccination, and tweets classified as vaccine hesitant, but with the additional step of excluding all accounts which individually contributed more than 0.2 percent (chosen arbitrarily) of the total identified tweets; high-volume accounts may contain contributions from automated scripts. We found an overall Spearman rank correlation of the total on-topic tweets with the weekly reported cases of 0.07 (95% CI: −0.06 to 0.21). We found an overall Spearman rank correlation of the total pro-vaccination tweets with the weekly reported cases of 0.2 (95% CI: 0.05 to 0.34), while the overall Spearman rank correlation of the total vaccine-hesitant tweets with the weekly reported cases was 0.01 (95% CI: −0.11 to 0.13).

When we removed tweets from accounts identified as having produced a highly repetitive tweet (see section "Methods"), the number of tweets classified as pro-vaccination decreased from 16,297 to 13,029, while the number of tweets classified as vaccine hesitant decreased from 43,259 to 13,462 When correlating the weekly number of remaining pro-vaccination and vaccine-hesitant tweets with the reported weekly case series, we found that for the remaining pro-vaccination

tweets, the Spearman correlation was 0.208 (95% CI: 0.059 to 0.339), while for the remaining vaccine-hesitant tweets, the Spearman correlation was 0.0092 (95% CI: −0.11 to 0.131).

Discussion

Social media have, in a few years, gained pervasive importance in many fields[86] and have begun to change healthcare[87] and healthcare research[88] as well. Social media can be used to assess public discussions of health-related issues and behaviors (such as vaccination) and as a vehicle for disseminating public health information, health promotion messages, or other evidence-based interventions.[41,89–91] In contrast, social media may also be a vehicle for the dissemination of views not supported by evidence.[51,92]

We classified social media posts related to measles as vaccine hesitant or not and compared the frequency of such responses with the occurrence of measles in the United States. Across the two social media platforms we examined, the proportion of pro-vaccination versus vaccine-hesitant posts, at times, depended on the presence of an outbreak. For the corpus we identified, posts classified as vaccine-hesitant occurred more regularly on the topic of vaccinations and/or measles than those classified as pro-vaccination. However, during a reported outbreak, pro-vaccination posts overtook the number of vaccine-hesitant posts, a feature which was attenuated in the first year. During these online discussions, vaccine-hesitant individuals may call into question the safety and efficacy of vaccines, while non-vaccine-hesitant individuals criticize and blame those who choose not to vaccinate for the recent outbreaks in vaccine-preventable diseases and for putting immunocompromised and vulnerable populations at risk for these diseases.[93] Our work differs from prior work in that we examined the relationship between volume and stance of the conversation and the occurrence of measles itself, as well as the comparison of Twitter and FB. It is possible that online discussions are driven, in part, by media coverage of outbreaks, an effect not captured by our classification of weeks based only on reported case series. Our findings are consistent with a simple view: many individuals who promote vaccine hesitancy are frequently or continually engaged in online discussions of vaccination, while non-vaccine-hesitant individuals on the whole engage in the debate when current events—such as a measles outbreak—bring measles to public awareness.

Our results suggest an ongoing substantial presence of vaccine hesitancy and vaccination opposition in social media, contrasted with sporadic involvement by individuals favoring vaccination. What effect could this social media involvement be yielding? While many factors affect message credibility,[94–96] of particular concern is the potential importance of bandwagon effects in a social media setting.[96] Is it possible that the large volume of social media content in opposition to vaccines can contribute to such an effect? If so, strategies to increase social media engagement by institutions and individuals in favor of vaccination should be developed, urgently. It may be valuable for medical and public health practitioners to ask not only how they may use social media effectively but how they may empower or encourage social media activism by advocates of evidence-based positions (such as vaccination).

Categorizing social media posts requires inferring a user's stance from a user's post. For example, tweets were capped at 140 characters, which may have forced users to use non-standard abbreviations or phrases that either lack clear context or are poorly phrased. Such texts may be harder to interpret, whether by machine or human. Our machine classification appears to have underestimated the number of pro-vaccination posts, when compared to human raters. This potentially results in underestimating the correlation of pro-vaccination post frequency with outbreak data. We also note that by correlating US measles cases to all available posts, regardless of geocoding, we potentially weaken the effect and bias the analysis toward null hypotheses—an effect perhaps partially offset by international coverage of the 2015 measles outbreak.[97] Finally, we observe that an

article critical of Wakefield's research[98] appears to have caused a large number of pro-vaccination posts, even though no outbreak was occurring. This effect would also dilute the correlation of pro-vaccination posts and outbreak counts. Our use of the BrightView classifier appears to have achieved substantial agreement with the humans, but to have fallen short of human-level performance in its classifications; training on a larger corpus with a more nuanced classification may improve our ability to assess sentiment and stance.[99] Finally, our study was descriptive and retrospective; no predictions of future social media behavior were ventured.

Acknowledgements

Declaration of conflicting interests

Funding

References

1. Henderson DA, Dunston FJ, Fedson DS, et al. The measles epidemic: the problems, barriers, and recommendations. *JAMA* 1991; 266: 1547–1552.
2. Sabella C. Measles: not just a childhood rash. *Clev Clin J Med* 2010; 77: 207–213.
3. Perry RT and Halsey NA. The clinical significance of measles: a review. *J Infect Dis* 2004; 189(Suppl 1): S4–S16.
4. Moss WJ and Griffin DE. Measles. *Lancet* 2012; 379: 153–164.
5. Committee on Infectious Diseases. *Red Book 2015. Report of the Committee on Infectious Diseases*. Elk Grove Village, IL: American Academy of Pediatrics, 2015.
6. Lessler J, Reich NG, Brookmeyer R, et al. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis* 2009; 9: 291–300.
7. Gutierrez J, Isaacson RS and Koppel BS. Subacute sclerosing panencephalitis: an update. *Dev Med Child Neurol* 2010; 52: 901–907.
8. Black FL. Measles active and passive immunity in a worldwide perspective. *Prog Med Virol* 1989; 36: 1–33.
9. Markowitz LE, Preblud SR, Fine PE, et al. Duration of live measles vaccine-induced immunity. *Pediatr Infect Dis J* 1990; 9: 101–110.
10. Uzicanin A and Zimmerman L. Field effectiveness of live attenuated measles-containing vaccines: a review of published literature. *J Infect Dis* 2011; 204(S1): S133–S148.
11. Patel MK, Gacic-Dobo M, Strebel PM, et al. Progress toward regional measles elimination—worldwide, 2000–2015. *MMWR Morb Mortal Wkly Rep* 2016; 65: 1228–1233.
12. Katz SL and Hinman AR. Summary and conclusions: measles elimination meeting, 16–17 March 2000. *J Infect Dis* 2004; 189(Suppl 1): S43–S47.
13. Zipprich J, Winter K, Hacker J, et al. Measles outbreak—California, December 2014–February 2015. *MMWR Morb Mortal Wkly Rep* 2015; 64: 153–154.
14. Blumberg S, Worden L, Enanoria W, et al. Assessing measles transmission in the United States following a large outbreak in California. *PLOS Currents Outbreaks*. 7 May 2015. Edition 1. doi: 10.1371/currents.outbreaks.b497624d7043b1aecfbfd3dfda3e344a.

15. Phadke VK, Bednarczyk RA, Salmon DA, et al. Association between vaccine refusal and vaccine-preventable diseases in the United States: A review of measles and pertussis. *JAMA* 2016; 315: 1149–1158.

16. Best M, Neuhauser D and Slavin L. "Cotton Mather, you dog, dam you! I'l inoculate you with this; with a pox to you": smallpox inoculation, Boston, 1721. *Qual Saf Health Care* 2004; 13: 82–83.

17. Wolfe RM and Sharp LK. Anti-vaccinationists past and present. *BMJ* 2002; 325: 430–432.

18. Dyer C. Lancet retracts Wakefield's MMR paper. *BMJ* 2010; 340: c696.

19. Brown KF, Long SJ, Ramsay M, et al. U.K. parents' decision-making about measles-mumps-rubella (MMR) vaccine 10 years after the MMR-autism controversy: a qualitative analysis. *Vaccine* 2012; 30: 1855–1864.

20. Callender D. Vaccine hesitancy: more than a movement. *Hum Vacc Immunother* 2016; 12: 2464–2468.

21. Larson HJ, Cooper LZ, Eskola J, et al. Addressing the vaccine confidence gap. *Lancet* 2011; 378: 526–535.

22. McClure CC, Cataldi JR and O'Leary ST. Vaccine hesitancy: Where we are and where we are going. *Clin Ther*. August 2017; 39(8): 1550–1562. Epub ahead of print 31 July 2017. doi: 10.1016/j.clinthera.2017.07.003.

23. Hill HA, Elam-Evans LD, Yankey D, et al. Vaccination coverage among children aged 19–35 months—United States, 2015. *MMWR Morb Mortal Wkly Rep* 2016; 65: 1065–1071.

24. Seither R, Calhoun K, Mellerson J, et al. Vaccination coverage among children in kindergarten—United States, 2015–16 school year. *MMWR Morb Mortal Wkly Rep* 2016; 65: 1057–1064.

25. Research P. Mobile messaging and social media 2015. Twitter demographics, http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/2015-08-19_social-media-update_11/. Archived at http://www.webcitation.org/6m9FcIr4T

26. Twitter. Twitter usage/company facts, https://about.twitter.com/company. Archived at http://www.webcitation.org/6m9FjrMn9

27. Chou WS, Hunt YM, Beckjord EB, et al. Social media use in the United States: implications for health communication. *J Med Internet Res* 2009; 11: e48.

28. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Heal Aff* 2009; 28: 361–368.

29. Grajales FJ, III, Sheps S, Ho K, et al. Social media: a review and tutorial of applications in medicine and health care. *J Med Internet Res* 2014; 16: e13.

30. Moorhead SA, Hazlett DE, Harrison L, et al. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res* 2013; 15: e85.

31. Allen C, Tsou MH, Aslam A, et al. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. *PLoS ONE* 2016; 11: e0157734.

32. Velasco E, Agheneza T, Denecke K, et al. Social media and internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q* 2014; 92: 7–33.

33. Hartley DM. Using social media and internet data for public health surveillance: the importance of talking. *Milbank Q* 2014; 92: 34–39.

34. Keller B, Labrique A, Jain KMP, et al. Mind the gap: social media engagement by public health researchers. *J Med Internet Res* 2014; 16: e8.

35. Mollema L, Harmsen IA, Broekhuizen E, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in the Netherlands in 2013. *J Med Internet Res* 2015; 17: e128.

36. Orr D, Baram-Tsabari A and Landsman K. Social media as a platform for health-related public debates and discussions: the polio vaccine on Facebook. *Isr J Heal Policy Res* 2016; 5: 34.

37. Deiner M, Lietman TM, McLeod SD, et al. Eye disease surveillance tools emerging from search and social media data: a study of temporal patterns of conjunctivitis. *JAMA Ophthalmol* 2016; 134: 1024–1030.

38. Surian D, Nguyen DQ, Kennedy G, et al. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *J Med Internet Res* 2016; 18: e232.

39. Harris JK, Mueller NL and Snider D. Social media adoption in local health departments nationwide. *Am J Public Heal* 2013; 103: 1700–1707.

40. Warren KE and Wen LS. Measles, social media and surveillance in Baltimore City. *J Public Health* 2017; 39: e73–e78.

41. Washington TA, Applewhite S and Glenn W. Using Facebook as a platform to direct young Black men who have sex with men to a video-based HIV testing intervention: a feasibility study. *Urban Soc Work* 2017; 1: 36–52.

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

43. Cataldi JR, Dempsey AF and O'Leary ST. Measles, the media, and MMR: impact of the 2014–15 measles outbreak. *Vaccine* 2016; 34: 6375–6380.

44. Powell GA, Zinszer K, Verma A, et al. Media content about vaccines in the United States and Canada, 2012–2014: an analysis using data from the Vaccine Sentimeter. *Vaccine* 2016; 34: 6229–6235.

45. Grant L, Hausman BL, Cashion M, et al. Vaccination persuasion online: a qualitative study of two pro-vaccine and two vaccine-skeptical websites. *J Med Internet Res* 2015; 17: e133.

46. Dunn AG, Leask J, Zhou X, et al. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *J Med Internet Res* 2015; 17: e144.

47. Bahk CY, Cumming M, Paushter L, et al. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Heal Aff* 2016; 35: 341–347.

48. Salathé M and Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamic and control. *PLoS Comput Biol* 2011; 7: e1002199.

49. Dunn AG, Surian D, Leask J, et al. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine* 2017; 35: 3033–3040.

50. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 2010; 28: 1709–1716.

51. Radzikowski J, Stefanidis A, Jacobsen KH, et al. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Heal Surveill* 2016; 2: e1.

52. Becker BF, Larson HJ, Bonhoeffer J, et al. Evaluation of a multinational, multilingual vaccine debate on Twitter. *Vaccine* 2016; 34: 6166–6171.

53. 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: 3778–3789.

54. Kang GJ, Ewing-Nelson SR, Mackey L, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine* 2017; 35: 3621–3638.

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

56. Faasse K, Chatman CJ and Martin LR. A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post. *Vaccine* 2016; 34: 5808–5814.

57. Massey PM, Leader A, Yom-Tov E, et al. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. *J Med Internet Res* 2016; 18: e318.

58. Vaccine sentimeter, http://www.healthmap.org/viss/ (accessed 1 December 2016).

59. Miller M, Banerjee T, Muppalla R, et al. What are people tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. *JMIR Public Health Surveill* 2017; 3: e38.

60. Roche B, Gaillard B, Léger L, et al. An ecological and digital epidemiology analysis on the role of human behavior on the 2014 Chikungunya outbreak in Martinique. *Sci Rep* 2017; 7: 5967.

61. Thapen N, Simmie D, Hankin C, et al. DEFENDER: detecting and forecasting epidemics using novel data-analytics for enhanced response. *PLoS ONE* 2016; 11: e0155417.

62. McGough SF, Brownstein JS, Hawkins JB, et al. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS Negl Trop Dis* 2017; 11: e0005295.

63. van Lent LG, Sungur H, Kunneman FA, et al. Too far to care? Measuring public attention and fear for Ebola using Twitter. *J Medical Internet Res* 2017; 19: e193.

64. Mohammad SN, Kiritchenko S, Sobhani P, et al. A dataset for detecting stance in tweets, http://saif-mohammad.com/WebDocs/StanceDataset-LREC2016.pdf2016. Archived at http://www.webcitation.org/6mfImc9Fp

65. MacDonald NE. Vaccine hesitancy: definition, scope and determinants. *Vaccine* 2015; 33: 4161–4164.

66. Crimson Hexagon, 2016. http://www.crimsonhexagon.com. Archived at http://www.webcitation. org/6mRCbG5Xs

67. Hopkins D and King G. A method of automated nonparametric content analysis for social science. *Am J Polit Sci* 2010; 54: 229–247.

68. Firat A, Brooks M, Bingham C, et al. Systems and methods for calculating category proportions, 2014. US Patent App. 13/804,096, http://www.freepatentsonline.com/y2014/0012855.html. Archived at http:// www.webcitation.org/6mfNftXdX

69. Nugues PM. *Language processing with Perl and Prolog: theories, implementation, and application*, 2nd ed. Berlin: Springer-Verlag, 2014.

70. Measles 2013 case definition. https://wwwn.cdc.gov/nndss/conditions/measles/case-definition/2013/. Archived at http://www.webcitation.org/6m9FGrurN

71. Measles serology: lab tools, CDC. http://www.cdc.gov/measles/lab-tools/serology.html. Archived at http://www.webcitation.org/6m9FPRSNw

72. Helfand RF, Heath JL, Anderson LJ, et al. Diagnosis of measles with an IgM capture EIA: the optimal timing of specimen collection after rash onset. *J Infect Dis* 1997; 175: 195–199.

73. Ratnam S, Tipples G, Head C, et al. Performance of indirect immunoglobulin M (IgM) serology tests and IgM capture assays for laboratory diagnosis of measles. *J Clin Microbiol* 2000; 38: 99–104.

74. Perry KR, Brown DWG, Parry JV, et al. Detection of measles, mumps, and rubella antibodies in saliva using antibody capture radioimmunoassay. *J Med Virol* 1993; 40: 235–240.

75. Fulginiti VA, Eller JJ, Downie AW, et al. Altered reactivity to measles virus: atypical measles in children previously immunized with inactivated measles virus vaccines. *JAMA* 1967; 202: 1075–1080.

76. Matsuzono Y, Narita M, Ishiguro N, et al. Detection of measles virus from clinical samples using the polymerase chain reaction. *Arch Pediat Adol Med* 1994; 148: 289–293.

77. Nakayama T, Mori T, Yamaguchi S, et al. Detection of measles virus genome directly from clinical samples by reverse transcriptase-polymerase chain reaction and genetic variability. *Virus Res* 1995; 35: 1–16.

78. Jayant NS. Average- and median-based smoothing techniques for improving digital speech quality in the presence of transmission errors. *IEEE T Commun* 1976; 24: 1043–1045.

79. Sobo EJ, Huhn A, Sannwald A, et al. Information curation among vaccine cautious parents: Web 2.0, Pinterest thinking, and pediatric vaccination choice. *Med Anthropol* 2016; 35: 529–546.

80. Kreiss J and Lahiri SN. Bootstrap methods for time series. In: Rao TS, Rao SS and Rao CR (eds) *Time series analysis: methods and applications*. Oxford: North-Holland (Elsevier), 2012.

81. CDC. Measles—United States, 2011. *MMWR Morb Mortal Wkly Rep* 2012; 61: 253–257.

82. CDC. Measles—United States, January 1–August 24, 2013. *MMWR Morb Mortal Wkly Rep* 2013; 62: 741–743.

83. Le Tellier A. Only after measles outbreak does Texas megachurch support vaccines. Los Angeles Times 2013. Archived at: http://www.webcitation.org/6m9FVocJG

84. Gastañaduy PA, Redd SB, Fiebelkorn AP, et al. Measles—United States, January 1–May 23, 2014. *MMWR Morb Mortal Wkly Rep* 2014; 63: 496–499.

85. Arizona Department of Health Services. Nation's largest measles outbreak of the year ends August 8, 2016, http://www.azdhs.gov/director/public-information-office/index.php#news-release-080816. Archived at http://www.webcitation.org/6m9DT5zOr

86. Fuchs C. *Social media: a critical introduction*. London: SAGE, 2017.

87. Hors-Fraile S, Atique S, Mayer MA, et al. The unintended consequences of social media in healthcare: new problems and new solutions. *Yearb Med Inform* 2016; 1: 47–52.

88. Sinnenberg L, Buttenheim AM, Padrez K, et al. Twitter as a tool for health research: a systematic review. *Am J Public Health* 2017; 107: e1–e8.

89. Mita G, Mhurchu N and Jull A. Effectiveness of social media in reducing risk factors for noncommunicable diseases: a systematic review and meta-analysis of randomized controlled trials. *Nutr Rev* 2016; 74: 237–247.

90. Balatsoukas P, Kennedy CM, Buchan I, et al. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. *J Med Internet Res* 2015; 17: e141.

91. Isaac BM, Zucker JR, MacGregor J, et al. Use of social media as a communication tool during a mumps outbreak—New York City, 2015. *MMWR Morb Mortal Wkly Rep* 2017; 66: 60–61.
92. Oyeyemi SO, Gabarron E and Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* 2014; 349: g6178.
93. Guidry JP, Carlyle K, Messner M, et al. On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine* 2015; 33: 5051–5056.
94. Westerman D, Spence PR and van der Heide B. Social media as information source: recency of updates and credibility of information. *J Comput Commun* 2014; 19: 171–183.
95. Lachlan KA, Spence PR, Edwards A, et al. If you are quick enough, I will think about it: information speed and trust in public health organizations. *Comput Hum Behav* 2014; 33: 377–380.
96. Lin X, Spence PR and Lachlan KA. Social media and credibility indicators: the effect of influence cues. *Comput Hum Behav* 2016; 63: 264–271.
97. Lesnes C. Aux Etats-Unis, la rougeole devient une maladie de riches. *Le Monde*. 2 October 2015. Archived at http://www.lemonde.fr/ameriques/article/2015/02/10/aux-etats-unis-la-rougeole-devient-une-maladie-de-riches_4573604_3222.html (accessed 31 October 2017).
98. Deer B. How the case against the MMR vaccine was fixed. *BMJ* 2011; 342: c5347.
99. Du J, Xu J, Song H, et al. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. *J Biomed Semant* 2017; 8: 9.

Appendix 1

Query

Our Crimson Hexagon search query is as follows:

(autism OR mercury OR thimerosal OR wakefield OR mccarthy OR immigrant OR obama OR #vax)

AND (vaccine OR measles OR "MMR vaccine" OR sb40 OR polio OR chickenpox OR hepatitis OR "mmr shot")

AND -$

AND -http

AND -RT

To more effectively target conversations between individuals, we excluded dollar signs, http, and retweets from the primary analysis. This minimizes the number of posts which consist of advertisements, forwarded web links, and retweets. To examine the role of retweets and other forwarded content, we conducted a second broader query, identical to the first except no longer excluding dollar signs, http, and retweets. Counts (not post content) were extracted with and without these restrictions separately.

Classification

Topic wheels (produced within the Crimson Hexagon platform) illustrate the classification of posts into pro-vaccination posts and posts which express vaccine hesitancy (Figure 3).

We also include topic wheels showing topics within posts during the period before and during the 2014–2015 outbreak, for posts classified as vaccine hesitant and non-vaccine hesitant (Figure 4).
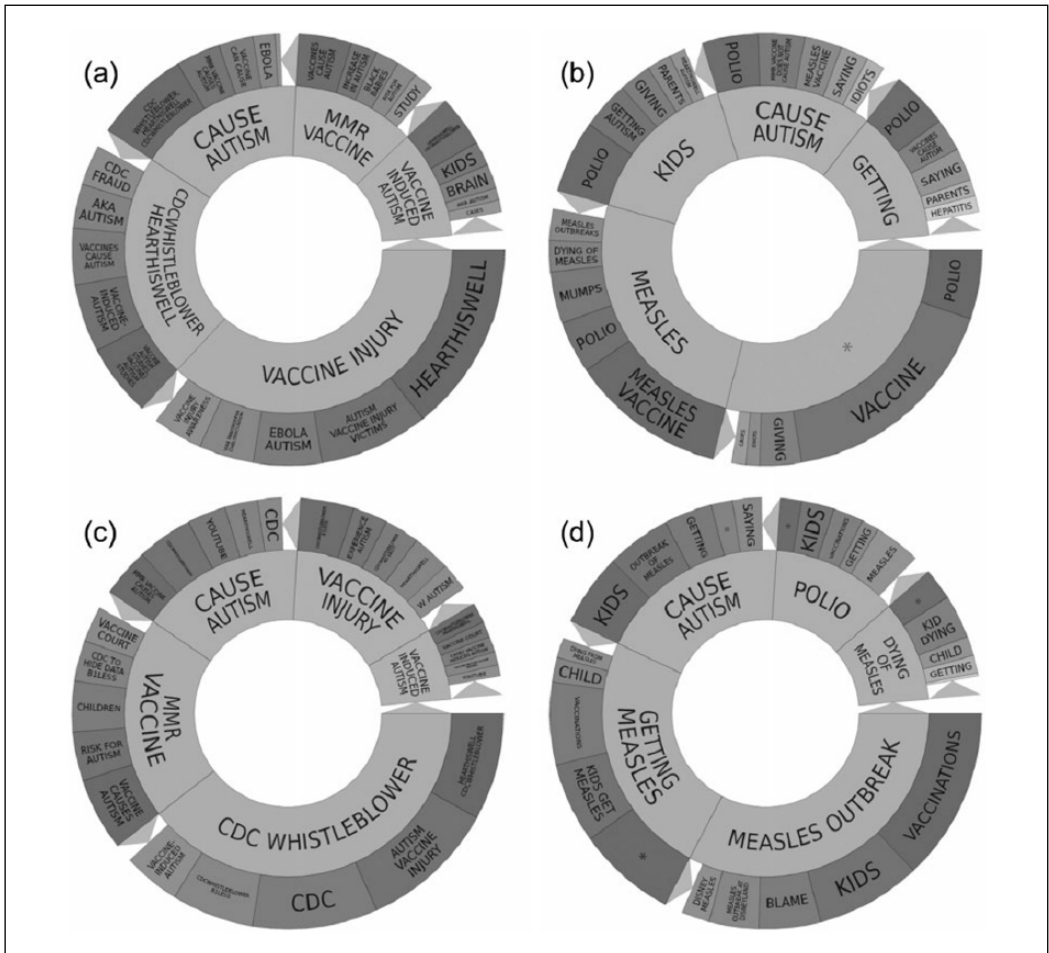
Outbreak periods

To identify starting and ending periods for national outbreak periods in a consistent manner across the study period, we used cumulative case reports (which are more complete than the weekly
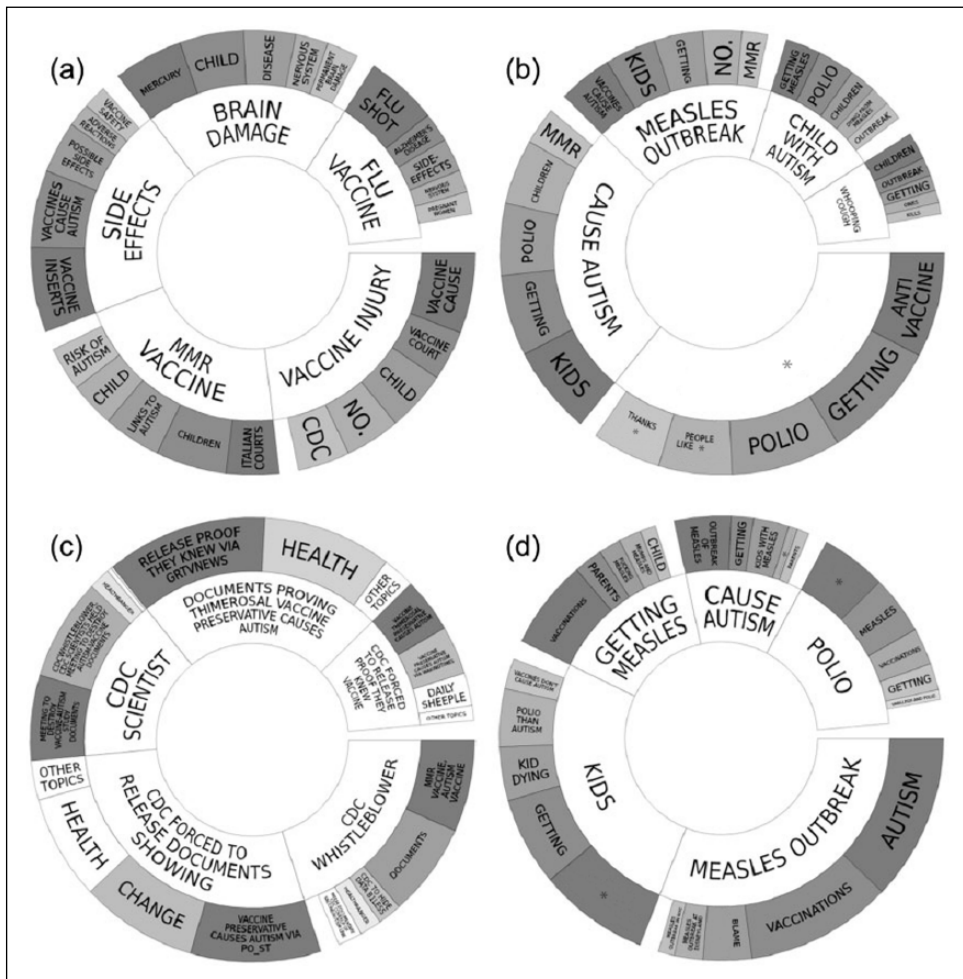
Figure 3. (a) Topic wheel for posts expressing vaccine hesitancy on Facebook, 23 June 2009 to 27 August 2016, derived from Crimson Hexagon. (b) Topic wheel for non-vaccine-hesitant Facebook posts in our corpus, 23 June 2009 to 27 August 2016, derived from Crimson Hexagon. (c) Topic wheel for tweets expressing vaccine hesitancy, 23 June 2009 to 27 August 2016, derived from Crimson Hexagon. (d) Topic wheel for non-vaccine-hesitant tweets in our corpus, 23 June 2009–27 August 2016, derived from Crimson Hexagon. *(*redacted specific famous celebrity's name).*

counts). Specifically, we removed outliers (reporting error) using a 5-week median smooth. Epidemics then appear as peaks in the first-differenced cumulative series. The average weekly increase in this cumulative count is approximately 3.6 cases per week; we chose an epidemic threshold as 250 percent of this, or a count of nine, as indicating an epidemic or outbreak. Weeks where the cumulative count increase by nine or more cases were assumed to be during a candidate epidemic or outbreak; for any such week, we included the 4 previous weeks and the 3 following weeks—the asymmetry reflecting the fact that the cumulative counts contain some reporting delay. This yielded seven candidate epidemics based on the cumulative count series. Two of these contained fewer than four cases during the weekly reported counts and were excluded. This procedure was conducted by a team member masked to the time series of Tweets and Facebook. This procedure yielded the five epidemic periods (see Table 4) discussed in the text.

Figure 4. (a) Topic wheel for tweets expressing vaccine hesitancy during the non-outbreak period 14 September 2014 to 15 November 2014, derived from Crimson Hexagon. (b) Topic wheel for tweets in our corpus not expressing vaccine hesitancy during the non-outbreak period 14 September 2014 to 14 November 2014, derived from Crimson Hexagon. (c) Topic wheel for tweets expressing vaccine hesitancy during the outbreak period 30 November 2014 to 30 May 2015, derived from Crimson Hexagon. (d) Topic wheel for tweets in our corpus not expressing vaccine hesitancy during the outbreak period 30 November 2014–30 May 2015, derived from Crimson Hexagon. *(*redacted specific famous celebrity's name).*

Table 4. Start and end dates for outbreak periods as defined in the appendix.

Outbreak period	Start	End	Total weekly cases
1	20 February 2011	15 October 2011	82
2	31 March 2013	6 July 2013	21
3	26 January 2014	13 September 2014	401
4	30 November 2014	30 May 2015	46
5	17 April 2016	11 June 2016	18

The total number of weekly cases during the period is shown, along with the total change in the cumulative case count as reported by the CDC.