Geographic and Demographic Correlates of Autism-Related Anti-Vaccine Beliefs on Twitter, 2009-15

Abstract

This study examines temporal trends, geographic distribution, and demographic correlates of anti-vaccine beliefs on Twitter, 2009-2015. A total of 549,972 tweets were downloaded and coded for the presence of anti-vaccine beliefs through a machine learning algorithm. Tweets with self-disclosed geographic information were resolved and United States Census data were collected for corresponding areas at the micropolitan/metropolitan level. Trends in number of anti-vaccine tweets were examined at the national and state levels over time. A least absolute shrinkage and selection operator regression model was used to determine census variables that were correlated with anti-vaccination tweet volume. Fifty percent of our sample of 549,972 tweets collected between 2009 and 2015 contained anti-vaccine beliefs. Anti-vaccine tweet volume increased after vaccine-related news coverage. California, Connecticut, Massachusetts, New York, and Pennsylvania had anti-vaccination tweet volume that deviated from the national average. Demographic characteristics explained 67% of variance in geographic clustering of anti-vaccine tweets, which were associated with a larger population and higher concentrations of women who recently gave birth, households with high income levels, men aged 40 to 44, and men with minimal college education. Monitoring anti-vaccination beliefs on Twitter can uncover vaccine-related concerns and misconceptions, serve as an indicator of shifts in public opinion, and equip pediatricians to refute anti-vaccine arguments. Real-time interventions are needed to counter anti-vaccination beliefs online. Identifying clusters of anti-vaccination beliefs can help public health professionals disseminate targeted/tailored interventions to geographic locations and demographic sectors of the population.

Keywords: autism spectrum disorder, beliefs, big data, machine learning algorithms, social media, Twitter, vaccines

Introduction

Recent outbreaks of previously eradicated, vaccine-preventable diseases such as measles and pertussis are a public health concern (Adams et al., 2016; Phadke et al., 2016; Winter et al., 2014; Zipprich et al., 2015). These outbreaks have been linked to parental delay or refusal of vaccines over anti-immunization related beliefs (Gust et al., 2008; Salmon et al., 2005; World Health Organization, n.d.). Anti-vaccination beliefs represent diverse elements and characteristics in relation to vaccines, which manifest themselves in a wide range of negative attitudes ranging from being fully against vaccines to expressing hesitancy about them (Gust et al., 2005). These beliefs are driven by distrust of government/the pharmaceutical industry (Larson et al., 2014), lack of perceived need, and doubts about vaccine safety and potential side effects (Chen & DeStefano, 1998; Chen & Hibbs, 1998; Smailbegovic et al., 2003). Of particular interest are misconceptions linking antigens in children's vaccines, especially thimerosal, to autism spectrum disorder [ASD; (Hviid et al., 2003)]. ASD is a developmental disorder characterized by communication, social, and behavioral impairments (Centers for Disease Control and Prevention, 2015; Phadke et al., 2016) and affects one in 68 eight-year-old children (Christensen et al., 2016).

Delay of vaccines involves individualized vaccine administration schedules against the official recommendations of the Advisory Committee on Immunization Practices (ACIP) on appropriate ages, number of doses, and intervals between doses, which compromises vaccine effectiveness (Centers for Disease Control and Prevention, 2004; Plotkin, 2011). In 2005, missed doses accounted for two-thirds of a 28% non-compliance rate with ACIP's recommendations among children 19-35 months (Luman et al., 2008). Other forms of non-compliance include delays up to six months for four or more vaccines in the first two years of life (Luman et al., 2005).

One measure of parental refusal of vaccines is the rate of nonmedical exemptions from mandated school immunizations (Salmon et al., 2005). Currently, District of Columbia and all states except Mississippi and West Virginia grant religious exemptions and 18 states grant philosophical exemptions (National Conference of State Legislatures, 2016). Data show a 6% annual increase in nonmedical exemptions in states that offer belief-based exemptions but no significant changes in religious exemptions (Omer et al., 2006). Further, nonmedical exemptions vary within states to create geographic clusters where rates of unvaccinated children are likely to increase (Omer et al., 2008; Omer et al., 2009; Richards et al., 2013; Smith et al., 2004). For example, in 2015-2016, county-level exemptions for kindergarten-aged children in Washington ranged from 1.0% to 17.0% (Washington State Department of Health, 2016).

Nonmedical exemptions are associated with acquisition and transmission of vaccine-preventable diseases (Feikin et al., 2000; Salmon et al., 1999). In a nationwide retrospective cohort study, exempt children were 35 times more likely to acquire measles than nonexempt children (Salmon et al., 1999). Another state-level retrospective cohort study showed that exempt children were 22.2 and 5.9 times more likely to acquire measles and pertussis than vaccinated children (Feikin et al., 2000). Beyond risks to exempt children, clusters of nonmedical exemptions pose risks to the community. Omer and colleagues found that geographic/temporal clusters of pertussis cases were 2.7 times more likely to overlap with exemption clusters after covariate adjustments (Omer et al., 2008). Similarly, incidence rates of measles and pertussis in vaccinated children were associated with the frequency of exempt children in a county [relative risk 1.6 and 1.9; (Feikin et al., 2000)].

Anti-vaccination beliefs either directly or indirectly (through vaccination perceived risks) predict underimmunization (Betsch et al., 2010; Brewer et al., 2007; Gust et al., 2004). However, research on the characteristics of individuals who hold anti-vaccination beliefs remains limited (Kata, 2012). Studies show that women and highly educated and high socio-economic parents are more likely to be concerned about vaccine safety and to delay/refuse childhood vaccines (Freed et al., 2010; Smith et al., 2010; Song, 2014). These results are largely based on survey methods that are subject to social desirability biases (Krumpal, 2013). Conversely, Web 2.0 affords an uncensored platform for disseminating vaccine-related beliefs (Witteman & Zikmund-Fisher, 2012). More importantly, parents who are concerned about vaccine safety and delay/refuse vaccines often seek health information online (Gust et al., 2005; Smith et al., 2010). Online sources are considered horizontal media sources (McCombs et al., 2014) where people choose to be exposed to beliefs and opinions similar to their own, creating an echo chamber and increasing the polarization around vaccines (Witteman & Zikmund-Fisher, 2012).

Researchers have documented the prevalence and content of anti-vaccination websites (Bean, 2011; Wolfe & Sharp, 2005). However, little is known about anti-vaccination beliefs on social media sites such as Twitter. The literature thus far has been limited to review articles on the potential role of social media in vaccination beliefs and behavior (Betsch et al., 2012; Dredze et al., 2016; Kata, 2012). Love and colleagues conducted, to our knowledge, the only data-driven study of the source, tone, and accuracy of 2,580 reposted/shared vaccination tweets (Love et al., 2013). The sample included all vaccine-related tweets (e.g., adult vaccines) and was limited to reposted/shared tweets over one week.

Twitter is a platform for health-related information (Scanfield et al., 2010) and exposure to vaccine-related information on social media has been associated with vaccine-related behavior

(Avery & Lariscy, 2014). Further, geo-tagged Twitter data allow researchers to identify geographical regions where anti-vaccination beliefs are predominant. The primary goal of this study is to examine variations in anti-vaccine beliefs that link vaccines to ASD by geographic distribution and demographics on Twitter. Specifically, we examined prevalence of anti-vaccine beliefs tied to ASD from 2009 to 2015 in U.S. micropolitan and metropolitan areas, as well as in entire states. Finally, we examined the association between micro/metro-specific demographic characteristics and geographic distribution of anti-vaccine tweets.

Methods

Data Collection

We used Social Studio's Radian6 (Kim et al., 2013; Stavrakantonakis et al., 2012) application programming interface (API) to identify publicly available tweets that contained at least one ASD and one vaccine-related search keyword. We used search keywords that were culled from previous literature (Diresta & Lotan, 2015; Offit, 2008) to retrieve tweets from Radian6 that mentioned ASD and vaccines. Search keywords were vaccine, vaccinated, immunization, mmr vaccine, mmrvaccine, #b1less, #hearus, heavy metals, leaky gut, mercury, ethylmercury, methylmercury, thimerosal, preservative, dpt, diphtheria-pertussis-tetanus, pharmaceutical companies, big pharma, autism, autistic, asperger. We also included slang and misspellings of search keywords (i.e., vacinne, vacine, antivax, anti vax, aspie, asberger, assberger, asd). Finally, we included hashtags that journalists described in their coverage of anti-vaccination beliefs on Twitter (i.e., cdcwhistleblower, cdc whistleblower, sb277).

A total of 549,972 tweets (including retweets) from 01/01/2009 to 08/21/2015 were returned and downloaded. To ensure search accuracy, two researchers coded a random sample of

550 tweets. We found that 540 tweets mentioned vaccines and ASD, a 98.2% accuracy for the search keywords adopted.

Data Coding

We adopted a machine learning approach to identify tweets that expressed anti-vaccine beliefs. We used "anti-vaccine" as an umbrella term to capture a wide range of negative beliefs about vaccines. This approach allowed us to manually annotate a manageable number of tweets to build an algorithm that then coded the entire dataset.

To train the algorithm, two researchers coded 2,000 tweets into two categories: (1) Anti-vaccine and (2) Other, which consisted of tweets that were pro-vaccines and neutral (i.e., tweets that did not make a judgment about ASD and vaccines). Anti-vaccine tweets portrayed vaccines as dangerous, ineffective or negative, and mentioned a potential causal link to ASD. Examples of tweets that fell into the anti-vaccine category include: "CDC whistleblower confesses to publishing fraudulent data to obfuscate link between vaccines and autism," and "RT @_______________: Autism is primarily caused by mercury present in vaccines." Examples of tweets that fell into the other category include: "NEWS: The Lancet revokes 1998 Wakefield, et. al. paper associating MMR to autism and GI problems. On February 2,..."; "Additional evidence of no link between Autism and thimerosal, a preservative used in vaccines. New national study published in Medscape Today," "Could too many vaccines too early lead to #autism? Latest study says no (please don't shoot the messenger): http://t.co/HIgHbrnGLZ," and "Do you know the truth about the link between vaccines and #autism? Hint: Science knows. (via @upworthy http://t.co/8rbRlOMMfj" (wording was changed slightly to maintain user anonymity).

The two coders read the tweets and discussed the nature of anti-vaccine sentiment, and then made their judgments independently. The two coders had perfect agreement for the anti-vaccination category and, thus, had perfect intercoder reliability. This was due to the straightforward nature of anti-vaccination tweets. The manually annotated data were then used to build a machine learning algorithm.

LightSide (Mayfield & Rosé, 2013), an open-source platform that performs feature-extraction (e.g., text-strings, strings that include characters like hashtags), was used to build a machine learning algorithm. Features used in the final model included: unigrams, bigrams, word occurrence counts, punctuation, and feature hit location tracking. These features were selected by the researchers in an effort to maximize accuracy and precision. The performance of the model was tested against cross-fold validation ($N = 10$). The training and test set size was calculated using the formula (N-1/N). In each test, 90% of the data was used to train and 10% of the data was used to test.

Final accuracy (i.e., number of correct predictions divided by total number of predictions made; $a = .8617$) and Kappa (i.e., comparison of observed accuracy accounting for expected accuracy; $k = .7227$) were acceptable. Models with accuracy $\geq .68$ are generally acceptable (Bradley, 1997). Models with Kappa scores in the range of .61 – .80 reflect substantial agreement (Landis & Koch, 1977) that was not due to random chance or biased classification. Examples of misclassified tweets included “Mercury Levels Same in Autistic, Other Children: Blood levels of mercury are comparable in children with a.. http://bit.ly/2cTqP7 #health” (a false positive) and “RT @____________: #OpTwitterRaid #TheMostCommonLies MMR does not cause Autism. Dr Wakefield and 16 other studies all must be wrong...” (a false negative). The algorithm was then applied to the remaining, unseen tweets. As an extra external validity check, 350 newly classified tweets were randomly selected and coded by a researcher to verify the algorithm correctly classified sentiments of each tweet. Of the 350 tweets, 313 were correctly classified (89.43%). This suggests that the performance metrics were indeed accurate, and the model was externally valid. Overall, 272,546 tweets (49.5% of all downloaded tweets) were classified as containing anti-vaccination beliefs.

Geocoding

We geographically tagged tweets based on census Metropolitan and Micropolitan Statistical areas, which are geographic entities delineated by the Office of Management and Budget (OMB) for use by federal agencies when collecting, tabulating, and publishing federal statistics. Metro areas contain a core urban area of 50,000 or more residents whereas micro areas contain an urban core of at least 10,000 (but fewer than 50,000) residents. Each metro or micro area consists of one or more counties including the core urban area and any adjacent counties that have a high degree of social/economic integration with the urban core (as measured by commuting to work). Each tweet was geolocated to the area from which it originated if (1) the tweet had Global Positioning Systems (GPS) coordinates that resolved to a micro/metro area or (2) if the user had a self-disclosed location in their profile that could be resolved to a micro/metro area. Retweets were geocoded to the area from which they were retweeted.

Less than one percent of the tweets had GPS coordinates (n = 511) because twitter users' tend to not opt-in to geotagging (Vargo & Hopp, 2015). A majority of tweets (n = 172,730, 63.3%) originated from user profiles that had self-disclosed locations, which were queried using Google Maps Places API (Google, 2016) to resolve an official city name, if possible. Geolocated tweets had a lower percentage of anti-vaccine beliefs compared to the full set of tweets (31.02% versus 49.56%). This is likely because most bots (i.e., software that automatically manages a Twitter account and performs tasks such as tweeting, retweeting, favoring) have been known to spread misinformation online and it has been noted that they tend not to report locations (Guo & Chen, 2014; Thomas et al., 2012) and were thus not included in the geotagged set of tweets.

CensusReporter's API (UserVoice, 2016) was then used to resolve city names to the correct micro or metro area. As an external validity check, a researcher examined 560 random user profiles and the resulting micro or metro area and found that 531 profiles (94.8%) were correctly identified. In all, 108,207 tweets (39.7% of all anti-vaccination tweets) were coded to 732 micro/metro areas. There were 61.52 users per micro/metro area on average (*SD* = 243.95) and each user sent an average of 45.49 anti-vaccination tweets (*SD* = 188.40). These standard deviations are comparable to other studies that have geo-resolved Twitter data by region (Vargo & Hopp, 2015).

Census data was obtained for all 732 areas using CensusReporter's API, which allows census variables to be downloaded for all micro/metro areas. For all census variables, the American Community Survey 2015 5-year data were used because 5-year collections are the most reliable and exhaustive estimates. Here, the 2015 dataset was selected because the 5-year aggregation period (January 1, 2011 to December 31, 2015) closely matched our Twitter data collection period.

Data analysis

To examine rates of anti-vaccine beliefs on Twitter over time, tweet volume was calculated at the state level because data were sparse at metro and micro areas across time. A total of 108,413 tweets were resolved from 47,236 unique users at the state level by year and month. Granger causality tests were performed to identify outlier states that generated anti-vaccine tweets that deviated from the national trend. A significant score was an indicator that any given state's monthly anti-vaccine tweet volume was predicted by the national average, given a one-day lag.

To identify demographic correlates of anti-vaccination tweet volume, we examined 188 census variables including percentage of population by age and sex, sex by marital status, women who had a birth by marital status and educational attainment, women who had a birth by marital status and receipt of public assistance income, women who had a birth by age, race, educational attainment, and household income in the past 12 months (See Supplementary Table 1). These census variables were chosen as predictors based on previous literature on vaccine-related beliefs and behavior (Freed et al., 2010; Smith et al., 2010; Song, 2014). We created a least absolute shrinkage and selection operator (LASSO) regression model in Python using the ElasticNet module in sklearn to determine which variables most uniquely correlated with anti-vaccine tweet volume (Tibshirani, 2011). LASSO is a regularized regression that assesses the combined effect of many correlated variables and is used when many possible correlated variables exist for its automatic selection of variables that explain a substantial proportion of the variance over those that explain little variance.

Results

Number of anti-vaccine tweets per month appears in Figure 1, which shows the monthly national average ($M = 23.77$, $SD = 40.35$) as well as the five states that most deviated from the national average: California, Connecticut, Massachusetts, New York, and Pennsylvania (all $p > 0.05$ on Granger causality test). Significant elevations (i.e., monthly averages greater than two standard deviations above the mean) were observed in anti-vaccine tweet volume nationally in August (5.15 *SDs*) and September (3.74 *SDs*) of 2014 and in January (2.47 *SDs*) and February (3.72 *SDs*) of 2015. Figure 2 contains the percentage of anti-vaccine tweets per micro/metro area from 2009 to 2015.

Six census-level demographic variables explained 67% of the variance in volume of anti-vaccine tweets (Table 1). We used the standard deviations for each variable to derive the unit change seen for each predictor variable (*x*) and tweet volume as a dependent variable (*y*). For every increase in population size by 5837 people, we observed an additional anti-vaccine tweet. For demographics, we observed an increase in anti-vaccine tweets by one standard deviation (n = 205) in association with a 1.26% increase in percentage of women who gave birth within the last 12 months, a 1.98% increase in percentage of households with an income equal to or more than $200,000, an 0.40% increase in percentage of males aged 40-44 years old, a 1.47% increase in percentage of males who attended college for one year but did not receive a college degree, and an 0.24% decrease in percentage of females ages 15-17 years old. Age, sex, and race did not explain unique variance in the model.

The aforementioned results included race as a possible census-level demographic predictor associated with areas that experienced heightened levels of anti-vaccine tweets. Entered as a predictor variable, the Asian-only race variable was significantly associated with anti-vaccine tweet volume whereas income was not. It is noteworthy that the Asian race variable was correlated with household income at both $150,000 and $200,000+ ($r > 0.5$, $p < .05$). This correlation was the strongest that any racial group had with other variables of interest. Because of the correlation between the proportions of those with Asian descent in a geographical area and higher household income, LASSO chose to include percent Asian in the model. Given the lack of literature suggesting a link between Asian descent and anti-vaccine beliefs and, conversely, evidence that suggests relations between higher socioeconomic status and anti-vaccine beliefs (Smith et al., 2010), we elected to include household income in our final model over the Asian-only race variable. More importantly, relations between census-level household income variables and anti-vaccine tweets were observed in the current dataset. Across income categories, the correlation between anti-vaccine beliefs grew in a positive direction as household income increased (Table 2). No other race variables explained unique variance in the model despite what others have found regarding vaccine-related beliefs among Latinos and African Americans compared to non-Hispanic Whites (Bazzano et al., 2012).

Discussion

This study documented trends of anti-vaccination beliefs on Twitter, their geographic distribution, and demographic correlates over a six-year period, 2009-2015. Despite evidence against a causal link between childhood vaccines and ASD (Centers for Disease Control and Prevention, n.d.; Institute of Medicine Board on Health Promotion and Disease Prevention, 2004; "Joint statement of the American Academy of Pediatrics (AAP) and the United States Public Health Service (USPHS)," 1999; Stratton et al., 2012; Taylor et al., 2014), this belief remains (Bazzano et al., 2012; Fischbach et al., 2016) and is promulgated on online forums (Kata, 2012). Our results show that the volume of anti-vaccine tweets remained steady from 2009 to 2014. In August/September of 2014 and again in January/February of 2015, spikes were observed. Dredze et al. (2016) documented comparable spikes in vaccine-related tweets during this same period, which could be attributed to media coverage of vaccine-related news, such as revelation of the famed “CDC Whistleblower” in August 2014 (Park, 2014) and a California measles outbreak that began in December 2014 (Zipprich et al., 2015). According to the agenda setting theory, media coverage increases issue salience among their audiences (Begg et al., 1998; Hackett, 2008; McCombs et al., 2014). The nature of the aforementioned events may have further instilled concerns associated with anti-vaccine beliefs such as vaccine safety and distrust in government (Kata, 2012; Larson et al., 2014).

Anti-vaccine tweets disproportionally originated from five states. During the data collection period, California and Pennsylvania granted both religious and philosophical exemptions from mandated vaccines whereas Connecticut, Massachusetts, and New York offered religious exemptions only (National Conference of State Legislatures, 2016). Research shows that religious and philosophical vaccine exemptions are associated with clusters of undervaccinated children (Omer et al., 2008; Omer et al., 2009; Richards et al., 2013). The ease of granting exemptions is also associated with a 5% annual increase in non-medical exemptions (Omer et al., 2006). It is noteworthy that California legislature banned nonmedical vaccine exemptions in the summer of 2015 (Firger, 2015). Vaccine rates appear to vary by location (Hill et al., 2015; Lieu et al., 2015; Smith et al., 2004). Rates for the combined seven-vaccine series were 71.9% in New York, 72.8% in Pennsylvania, 75.0% in California, 78.5% in Massachusetts, and 80.6% in Connecticut (Hill, 2016). Future research should examine state-level factors associated with online vaccination beliefs such as nonmedical exemptions, ease of obtaining exemptions, and vaccination rates.

Of census variables examined, the percentage of women who had recently given birth predicted anti-vaccine tweet volume. Other predictors included household incomes $200,000 and higher, percentage of men ages 40 to 44 years old, and percentage of men who received one year of college education. Finally, percentage of females between the ages of 15 and 17 was a negative predictor anti-vaccine volume. Our results align with previous studies that show that women were more likely to have concerns about vaccines (Freed et al., 2010), that high income levels were associated anti-vaccine beliefs and behaviors [i.e., delaying childhood vaccines; (Smith et al., 2010)], and that older individuals (over age 40) were concerned about vaccine safety (Gust et al., 2005). According to health behavior theories, beliefs are important predictors of vaccination behaviors (Gust et al., 2004). Consistently, studies show that decisions to delay/refuse childhood vaccination schedules were higher among married, Caucasian mothers with higher education, higher income, and older age (Freed et al., 2010; Smith et al., 2010). These results have implications for health disparities where an intentional delay/refusal of vaccines among these populations compromises the protection of herd immunity (Plotkin, 2011). This puts unvaccinated or undervaccinated children for reasons beyond choice (e.g., lack of health insurance, belonging to a racial/ethnic minority group, living below the federal poverty line) at risk (Chu et al., 2004; Fronstin, 2005; Hill, 2016).

Success of vaccines depends on public acceptance (Streefland et al., 1990). Social media sites like Twitter show promise for public health efforts (Avery & Lariscy, 2014; Broniatowski et al., 2013). "Social listening" (Cole-Lewis et al., 2015) allows for an examination of vaccination-related beliefs and can serve as an early indicator of shifts in public opinion that might not be captured in traditional surveys due to high costs (Dredze et al., 2016), and sampling (Call et al., 2011; Duggan et al., 2015) and social desirability biases (Krumpal, 2013). Further, online activity on internet search sites (e.g., Google) and social media (e.g., Twitter) has accurately mirrored health-related events (e.g., disease activity) in previous studies (Ginsberg et al., 2009; Polgreen et al., 2008; Signorini et al., 2011). Public health professionals should implement real-time interventions that, aided by computer-assisted content analysis software and machine learning algorithms, are designed to instantly detect anti-vaccine tweets and reply with counter messages using the twitter handle (i.e., @username) and/or hashtag of the original tweet. These intervention messages should also target cross postings (i.e., messages simultaneously posted on multiple social media sites). Previous interventions were successful in refuting misconceptions about the vaccine-ASD link (Nyhan et al., 2014). Research is needed to examine the effectiveness of real-time online interventions in curbing the spread of anti-vaccine beliefs (e.g., retweeting) and lowering their volume within defined geographic areas.

Additionally, monitoring social media for up-to-date anti-vaccine beliefs allows public health professionals to address such beliefs by targeting geographic areas where these beliefs are most prevalent and tailoring the approach to demographic characteristics of populations most correlated with these beliefs. Future research should further understand the qualitative variation in anti-vaccine beliefs by geographic distribution and demographic correlates. Further, future research should aim to better identify and track the wide spectrum of negative beliefs about vaccines (e.g., condemnation versus hesitancy) that are expressed on social media. This will afford fine-tuned targeting and tailoring activities of health interventions, which are important because localized geographic clusters of anti-vaccine beliefs and delay/refusal of vaccines compromise herd immunity even when the national and/or state-level coverage of vaccines is high (Hill, 2016; Plotkin, 2011; Washington State Department of Health, 2016).

Finally, our results on geographic clustering of anti-vaccine beliefs suggest that healthcare providers may be more or less likely to encounter parents who hold negative vaccine-related beliefs. Equipped with timely knowledge of anti-vaccine beliefs prevalent in their immediate communities, pediatricians can address such beliefs in their interactions with parents. This practice will help pediatricians comply with the recommendation of the American Academy of Pediatrics on Bioethics to continue interactions with parents who express anti-vaccine beliefs and/or refuse/delay vaccines (Diekema, 2005). This is particularly important because healthcare providers remain trusted sources for information on vaccines for many parents (Kennedy et al., 2011). Public health professionals should maintain an up-to-date interactive map with prevalent anti-vaccine beliefs that pediatricians and clinics/hospitals can constantly access and sign up for notifications of increases in anti-vaccine beliefs in their geographic locations (to investigate a specific area, see Figure 2, or visit: *website url provided upon acceptance* for an interactive version via MapBox (Zastrow, 2015). Public health professionals should also develop systems to provide clinicians with information about the nature of vaccine-related concerns being expressed in any given geographic areas. For example, algorithms that code and provide summaries of the content of anti-vaccine tweets may prove most helpful.

The study has several limitations. We dichotomized vaccine-related tweets into anti-vaccine versus other rather than a fine-tuned classification of the wide spectrum of anti-vaccine beliefs. Aside from basic self-disclosed information, little is known about the demographic characteristics (e.g., sex, age, income) of Twitter users. Therefore, our results only show that anti-vaccine beliefs were more prevalent in areas with certain demographic composition rather than among Twitter users. In other words, our results should be considered as *census conditions* in which an area (i.e., a city) experiences heightened levels of anti-vaccination social media chatter. Additionally, our data show that 2009 and 2010 yielded 9,565 geotagged tweets whereas 2011 alone yielded 15,732 tweets, which we attribute to the rise of Twitter in the United States from 8% in 2010 to 13% in 2011 (Smith, 2011; Smith & Rainie, 2010). Further, our dataset of ASD and vaccine tweets represents a sub-population of Twitter users who wanted their location to be broadcasted alongside their message. However, studies with similar limitations successfully predicted public health phenomena such as infectious disease transmission (Sadilek et al., 2012). It is well documented that anti-vaccination messages are spread by bots (Diresta & Lotan, 2015; Thomas et al., 2012), which were not included in our final geotagged sample. However, we believe analysis of geo-resolved messages is representative of public opinion.

Additionally, in accordance with Twitter privacy policy, Radian6 archives all tweets except those that have been deleted by users, tweets that belong to deleted accounts, or tweets from private Twitter accounts. This is unlikely to affect the contributions of this study because it is the public expression of anti-vaccine beliefs beyond close ties (e.g., friends) that is of interest here. Twitter users represent 23% of internet users (20% of the population) who are over proportionally urban (30% vs. 15% rural), younger (32% 18-29 years old, 29% 30-49 years old vs. 13% 50-64 years old and 6% 65+), educated (27% college+, 23% some college, 19% high school or less) individuals (Duggan, 2015). Thus, our results are not nationally representative of vaccine beliefs. Future research would benefit from a hybrid approach whereby survey participants report personal demographic information (e.g., age, race/ethnicity, insurance coverage) and provide consent for analysis of their social media posts. Finally, more research is needed to examine anti-vaccine beliefs among understudied minority groups (e.g., non-Hispanic Asians) and men.

In conclusion, vaccines are effective in preventing contagious diseases (Centers for Disease Control and Prevention, 1999; van Panhuis et al., 2013). Vaccination rates for children under the age of two remain high in the U.S. with four vaccines meeting the 90% coverage goal of *Healthy People 2020* (Hill, 2016). Further, public opinion of children vaccines remains favorable with 83% regarding vaccines as safe and 68% supporting mandatory children vaccines (Anderson, 2015). However, the volume of online anti-vaccine beliefs is alarming and may indicate shifts in public opinion, which can translate to lower vaccine coverage. We show that anti-vaccine tweets coincide with vaccine-related news events and cluster geographically in areas with high concentrations of women who recently gave birth, households with high income levels, 40-44-year-old men, and men with no college degree. Monitoring social media for anti-vaccine beliefs is beneficial for surveillance and intervention efforts to curtail anti-vaccine beliefs.

References

Adams, D., Thomas, K., Jajosky, R., & et al. (2016). Summary of notifiable infectious diseases and conditions -- United States 2014. *Morbidity and Mortality Weekly Report,* 63, 1-152.

Anderson, M. (2015). 5 facts about vaccines in the U.S.: Pew Research Center.

Avery, E.J., & Lariscy, R.W. (2014). Preventable disease practices among a lower SES, multicultural, nonurban, U.S. community: The roles of vaccination efficacy and personal constraints. *Health Communication,* 29, 826-836.

Bazzano, A., Zeldin, A., Schuster, E., Barrett, C., & Lehrer, D. (2012). Vaccine-related beliefs and practices of parents of children with autism spectrum disorders. *American Journal of Intellectual and Developmental Disabilities,* 117, 233-242.

Bean, S.J. (2011). Emerging and continuing trends in vaccine opposition website content. *Vaccine,* 29, 1874-1880.

Begg, N., Ramsay, M., White, J., & Bozoky, Z. (1998). Media dents confidence in MMR vaccine. *British Medical Journal,* 316, 561.

Betsch, C., Brewer, N.T., Brocard, P., Davies, P., Gaissmaier, W., Haase, N., et al. (2012). Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine,* 30, 3727-3733.

Betsch, C., Renkewitz, F., Betsch, T., & Ulshofer, C. (2010). The influence of vaccine-critical websites on perceiving vaccination risks. *Journal of Health Psychology,* 15, 446-455.

Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition,* 30, 1145-1159.

Brewer, N., Chapman, G., Gibbons, F., Gerrard, M., McCaul, K., & Weinstein, N. (2007). Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychology,* 26, 136-145.

Broniatowski, D.A., Paul, M.J., & Dredze, M. (2013). National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic. *PloS One,* 8, e83672.

Call, K.T., Davern, M., Boudreaux, M., Johnson, P.J., & Nelson, J. (2011). Bias in telephone surveys that do not sample cell phones. *Medical Care,* 49, 355-364.

Centers for Disease Control and Prevention (1999). Achievements in public health, 1900-1999 impacts of vaccines universally recommended for children -- United states, 1990-1998. *Morbidity and Mortality Weekly Report,* 48, 243-248.

Centers for Disease Control and Prevention (2004). Recommended childhood and adolescent immunization schedule--United States, January-June 2004. *Morbidity and Mortality Weekly Report,* 53, Q1.

Centers for Disease Control and Prevention. (2015). Vaccines Do Not Cause Autism.

Centers for Disease Control and Prevention. (n.d.). Science Summary: CDC Studies on Thimerosal in Vaccines.

Chen, R.T., & DeStefano, F. (1998). Vaccine adverse events: causal or coincidental? *Lancet,* 351, 611-612.

Chen, R.T., & Hibbs, B. (1998). Vaccine safety: current and future challenges. *Pediatric annals,* 27, 445-455.

Christensen, D., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J.M., et al. (2016). Prevalence and characteristics of autism spectrum disorder among children aged

8 years – autism and developmental disabilities monitoring networks, 11 sites, United States, 2012. *Morbidity and Mortality Weekly Report,* 65, 1-23.

Chu, S.Y., Barker, L.E., & Smith, P.J. (2004). Racial/ethnic disparities in preschool immunizations: United States, 1996-2001. *American Journal of Public Health,* 94, 973-977.

Cole-Lewis, H., Pugatch, J., Sanders, A., Varghese, A., Posada, S., Yun, C., et al. (2015). Social listening: A content analysis of e-cigarette discussions on Twitter. *Journal of Medical Internet Research,* 17, e243.

Diekema, D.S. (2005). Responding to parental refusals of immunization of children. *Pediatrics,* 115, 1428-1431.

Diresta, R., & Lotan, G. (2015). Anti-vaxxers are using Twitter to manipulate a vaccine bill. Wired.

Dredze, M., Broniatowski, D.A., Smith, M.C., & Hilyard, K.M. (2016). Understanding vaccine refusal why we need social media now. *American Journal of Preventive Medicine,* 50, 550-552.

Duggan, M. (2015). Mobile messaging and social media 2015. Pew Research Center.

Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social Media Update 2014. Pew Research Center.

Feikin, D.R., Lezotte, D.C., Hamman, R.F., Salmon, D.A., Chen, R.T., & Hoffman, R.E. (2000). Individual and community risks of measles and pertussis associated with personal exemptions to immunization. *Jama,* 284, 3145-3150.

Firger, J. (2015). California bans religious and philosophical vaccine exemptions for schoolchildren. Newsweek.

Fischbach, R.L., Harris, M.J., Ballan, M.S., Fischbach, G.D., & Link, B.G. (2016). Is there concordance in attitudes and beliefs between parents and scientists about autism spectrum disorder? *Autism,* 20, 353-363.

Freed, G.L., Clark, S.J., Butchart, A.T., Singer, D.C., & Davis, M.M. (2010). Parental vaccine safety concerns in 2009. *Pediatrics,* 125, 654-659.

Fronstin, P. (2005). Sources of health insurance and characteristics of the uninsured: analysis of the March 2005 current population survey. *EBRI Issue Brief*, 287.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature,* 457, 1012-1014.

Google. (2016). Google Places API. Mountain View, CA: Google.

Guo, D., & Chen, C. (2014). Detecting Non-personal and Spam Users on Geo-tagged Twitter Network. *Transactions in GIS,* 18, 370-384.

Gust, D., Strine, T., Maurice, E., Smith, P., Yusuf, H., Wilkinson, M., et al. (2004). Underimmunization among children: Effects of vaccine safety concerns on immunization status. *Pediatrics,* 114, e16-e22.

Gust, D.A., Brown, C.J., Sheedy, K., Hibbs, B., Weaver, D., & Nowak, G. (2005). Immunization attitudes and beliefs among parents: Beyond a dichotomous perspective. *American Journal of Health Behavior,* 29, 81-92.

Gust, D.A., Darling, N., Kennedy, A.M., & Schwartz, B. (2008). Parents with doubts about vaccines: Which vaccines and reasons why. *Pediatrics,* 122, 718-725.

Hackett, A.J. (2008). Risk, its perception and the media: the MMR controversy. *Community Pract,* 81, 22-25.

Hill, H.A. (2016). Vaccination coverage among children aged 19–35 months—United States, 2015. *Morbidity and Mortality Weekly Report,* 65.

Hill, H.A., Elam-Evans, L.D., Yankey, D., Singleton, J.A., & Kolasa, M. (2015). National, state, and selected local area vaccination coverage among children aged 19–35 months — United States, 2014. *Morbidity and Mortality Weekly Report,* 64, 889-896.

Hviid, A., Stellfeld, M., Wohlfahrt, J., & Melbye, M. (2003). Association between thimerosal-containing vaccine and autism. *Jama,* 290, 1763-1766.

Institute of Medicine Board on Health Promotion and Disease Prevention (2004). *Immunization Safety Review. Vaccines and Autism*. Washington D C: The National Academies Press.

Joint statement of the American Academy of Pediatrics (AAP) and the United States Public Health Service (USPHS). (1999). *Pediatrics,* 104, 568-569.

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm – An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine,* 30, 3778-3789.

Kennedy, A.M., Basket, M., & Sheedy, K. (2011). Vaccine attitudes, concerns, and information sources reported by parents of young children: Results from the 2009 HealthStyles Survey. *Pediatrics,* 127, S92-S99.

Kim, A.E., Hansen, H.M., Murphy, J., Richards, A.K., Duke, J., & Allen, J.A. (2013). Methodological considerations in analyzing Twitter data. *JNCI Monographs,* 47, 140-146.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity,* 47, 2025-2047.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

Larson, H.J., Caitlin, J., Eckersberger, E., Smith, D., M. D., & Paterson, P. (2014). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine,* 32, 2150-2159.

Lieu, T., Ray, G.T., Klein, N.P., Chung, C., & Kulldorff, M. (2015). Geographic clusters in underimmunization and vaccine refusal. *Pediatrics,* 135, 280-289.

Love, B., Himelboim, I., Holton, A., & Stewart, K. (2013). Twitter as a source of vaccination information: content drivers and what they are saying. *American Journal of Infection Control,* 41, 568-570.

Luman, E., Barker, L., & Shaw, K. (2005). Timeliness of childhood vaccinations in the United States: days undervaccinated and number of vaccines delayed. *Jama,* 293, 1204-1211.

Luman, E., Shaw, K., & Stokley, S. (2008). Compliance with vaccination recommendations for U.S. children. *American Journal of Preventive Medicine,* 34, 463-470.

Mayfield, E., & Rosé, C.P. (2013). LightSIDE: Open source machine learning for text accessible to non-experts. In M.D. Shermis, & J. Burstein (Eds.), Handbook of Automated Essay Evaluation pp. 124-135). New York, NY: Routledge.

McCombs, M.E., Shaw, D.L., & Weaver, D.H. (2014). New directions in agenda-setting theory and research. *Mass Communication and Society,* 17, 781-802.

National Conference of State Legislatures. (2016). States with religious and philosophical exemptions from school immunization requirements.

Nyhan, B., Reifler, J., Richey, S., & Freed, G.L. (2014). Effective messages in vaccine promotion: a randomized trial. *Pediatrics,* 133, e835-e842.

Offit, P. (2008). *Autism's false prophets: Bad science, risky medicine, and the search for a cure.* New York: Columbia University Press.

Omer, S.B., Enger, K.S., Moulton, L.H., Halsey, N.A., Stokley, S., & Salmon, D.A. (2008). Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. *American Journal of Epidemiology,* 168, 1389-1396.

Omer, S.B., Pan, W.K.Y., Halsey, N.A., Stokley, S., Moulton, L.H., Navar, A.M., et al. (2006). Nonmedical exemptions to school immunization requirements: Secular trends and association of state policies with pertussis incidence. *Jama,* 296, 1757-1763.

Omer, S.B., Salmon, D.A., Orenstein, W.A., deHart, M.P., & Halsey, N.A. (2009). Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. *New England Journal of Medicine,* 360, 1981-1988.

Park, A. (2014). Whistleblower claims CDC covered up data showing vaccine-autism link. Time: Time Inc.

Phadke, V.K., Bednarczyk, R.A., Salmon, D.A., & Omer, S.B. (2016). Association between vaccine refusal and vaccine-preventable diseases in the United States: A review of measles and pertussis. *Jama,* 315, 1149-1158.

Plotkin, S. (2011). "Herd immunity:" a rough guide. *Clinical Infectious Diseases,* 52, 911-916.

Polgreen, P.M., Chen, Y., Pennock, D.M., & Nelson, F.D. (2008). Using internet searches for influenza surveillance. *Clinical Infectious Diseases,* 47, 1443-1448.

Richards, J.L., Wagenaar, B.H., Van Otterloo, J., Gondalia, R., Atwell, J.E., Kleinbaum, D.G., et al. (2013). Nonmedical exemptions to immunization requirements in California: A 16-year longitudinal analysis of trends and associated community factors. *Vaccine,* 31, 3009-3013.

Sadilek, A., Kautz, H.A., & Silenzio, V. (2012). Modeling Spread of Disease from Social Interactions. ICWSM pp. 322-329).

Salmon, D.A., Haber, M., Gangarosa, E.J., Phillips, L., Smith, N.J., & Chen, R.T. (1999). Health consequences of religious and philosophical exemptions from immunization laws: Individual and societal risk of measles. *Jama,* 282, 47-53.

Salmon, D.A., Moulton, L.H., Omer, S.B., deHart, P., Stokley, S., & Halsey, N.A. (2005). Factors associated with refusal of childhood vaccines among parents of school-aged children: A case-control study. *Archives of Pediatrics and Adolescent Medicine,* 159, 470-476.

Scanfield, D., Scanfield, V., & Larson, E.L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control,* 38, 182-188.

Signorini, A., Segre, A.M., & Polgreen, P.M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS One,* 6, e19467.

Smailbegovic, M.S., Laing, G.J., & Bedford, H. (2003). Why do parents decide against immunization? The effect of health beliefs and health professionals. *Child: Care, Health, and Development,* 29, 303-311.

Smith, A. (2011). 13% of online adults use Twitter, and half of Twitter users access the service on a cell phone. Pew Research Center.

Smith, A., & Rainie, L. (2010). 8% of online Americans use Twitter. Pew Research Center.

Smith, P.J., Chu, S.Y., & Barker, L.E. (2004). Children who have received no vaccines: Who are they and where do they live? *Pediatrics,* 114, 187-195.

Smith, P.J., Humiston, S.G., Parnell, T., Vannice, K.S., & Salmon, D.A. (2010). The association between intentional delay of vaccine administration and timely childhood vaccination coverage. *Public health reports,* 125, 534-541.

Song, G. (2014). Understanding public perceptions of benefits and risks of childhood vaccinations in the United States. *Risk Analysis,* 34, 541-555.

Stavrakantonakis, I., Gagiu, A., Kasper, H., Toma, I., & Thalhammer, A. (2012). An approach for evaluation of social media monitoring tools. *Common Value Management,* 52, 52-64.

Stratton, K., Ford, A., Rusch, E., & Clayton, E.W. (Eds.) (2012). *Adverse effects of vaccines: evidence and causality*. Washington DC: The National Academies Press.

Streefland, P., Chowdhury, A.M.R., & Ramos-Jimenez, P. (1990). Patterns of vaccination acceptance. *Social Science and Medicine,* 49, 1705-1716.

Taylor, L.E., Swerdfeger, A.L., & Eslick, G.D. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine,* 32, 3623-3629.

Thomas, K., Grier, C., & Paxson, V. (2012). Adapting social spam infrastructure for political censorshi. Proceedings of the 5th USENIX conference on Large-Scale Exploits and Emergent Threats pp. 13-13). San Jose, CA: USENIX Association.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *J Roy Stat Soc B Met,* 73, 273-282.

U.S. Census Bureau. (2015). American Community Survey: 2011-2015 (5-Year Estimates).

UserVoice. (2016). Census Reporter. San Francisco, CA: UserVoice.

van Panhuis, W.G., Grefenstette, J., Jung, S.Y., Chok, N.S., Cross, A., Eng, H., et al. (2013). Contagious diseases in the United States from 1888 to the present. *New England Journal of Medicine,* 369, 2152-2158.

Vargo, C.J., & Hopp, T. (2015). Socioeconomic status, social capital, and partisan polarity as predictors of political incivility on Twitter: A congressional district-level analysis. *Social Science Computer Review*.

Washington State Department of Health. (2016). Washington State School Immunization Slide Set, 2015-2016 School Year.

Winter, K., Glaser, C., Watt, J., Harriman, K., Control, C.f.D., & Prevention (2014). Pertussis epidemic—California, 2014. *Morbidity and Mortality Weekly Report,* 63, 1129-1132.

Witteman, H.O., & Zikmund-Fisher, B.J. (2012). The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. *Vaccine,* 30, 3734-3740.

Wolfe, R.M., & Sharp, L.K. (2005). Vaccination or immunization? The impact of search terms on the internet. *Journal of Health Communication,* 10, 537-551.

World Health Organization. (n.d.). Six common misconceptions about immunization.

Zastrow, M. (2015). Science on the map. *Nature,* 519, 119.

Zipprich, J., Winter, K., Hacker, J., Xia, D., Watt, J., & Harriman, K. (2015). Measles outbreak —California, December 2014–February 2015. *Morbidity and Mortality Weekly Report,* 64, 153-154.

Figure Titles:

Figure 1. Number of monthly anti-vaccine tweets nationally and for top five states, 2009-2015

Figure 2. Percentage of anti-vaccination tweets per micro/metro area, 2009-2015. For a hi-resolution, interactive version of this map, please visit: (*url provided upon acceptance)*.

Table 1. LASSO Regression Model for Anti-Vaccination Beliefs

Predictor (Census Variable Code)	β
Raw count of general population (B01001001)	122.86
Percent of women who gave birth in the last 12 months (B13014002)	12.29
Percent with household income of $200,000 and above (B19001017)	1.18
Percent of males aged 40-44 (B01001014)	0.70
Percent of males with 1 year of college but no college degree (B15002013)	0.07
Percent of females aged 15 -17 (B01001030)	-0.34
Intercept	6.523
Training Data *RMSE*	186.05
Test Data *RMSE*	96.44
Training Data R^2	0.71
Test Data R^2	0.67

LASSO = Least Absolute Shrinkage and Selection Operator
β = linear regression coefficient
RMSE = Root Mean Squared Error
Training set (*n* = 117)
Test set (*n* = 584)
Results are 10-fold cross-validation averages.

Table 2. Correlations between Household Income and Anti-Vaccination Beliefs

Census-Level Household Income in Past 12 Months (Census Variable Code)	Anti-Vaccination Beliefs
Less than $10,000 (B19001002)	-.09*
$10,000-$14,999 (B19001003)	-.19***
$15,000-$19,999 (B19001004)	-.21***
$20,000-$24,999 (B19001005)	-.22***
$25,000-$29,999 (B19001006)	-.23***
$30,000-$34,999 (B19001007)	-.22***
$35,000-$39,999 (B19001008)	-.20***
$40,000-$44,999 (B19001009)	-.18***
$45,000-$49,999 (B19001010)	-.14***
$50,000-$59,999 (B19001011)	-.13***
$60,000-$74,999 (B19001012)	-.06
$75,000-$99,999 (B19001013)	-.06
$100,000-$124,999 (B19001014)	.21***
$125,000-$149,999 (B19001015)	.27***
$150,000-$199,999 (B19001016)	.35***
$200,000 or more (B19001017)	.41***

Note. Household income reflects the percentage of households in each income bracket per micropolitan/metropolitan area.

* $p < .05$. *** $p < .001$.