



Article

Studying Public Perception about Vaccination: A Sentiment Analysis of Tweets

Viju Raghupathi ¹, Jie Ren ² and Wullianallur Raghupathi ^{2,*}

¹ Koppelman School of Business, Brooklyn College of the City University of New York, Brooklyn, NY 11210, USA; VRaghupathi@brooklyn.cuny.edu

² Gabelli School of Business, Fordham University, New York, NY 10023, USA; JRen11@fordham.edu

* Correspondence: Raghupathi@fordham.edu; Tel.: +1-646-212-7230

Received: 17 March 2020; Accepted: 11 May 2020; Published: 15 May 2020



Abstract: Text analysis has been used by scholars to research attitudes toward vaccination and is particularly timely due to the rise of medical misinformation via social media. This study uses a sample of 9581 vaccine-related tweets in the period 1 January 2019 to 5 April 2019. The time period is of the essence because during this time, a measles outbreak was prevalent throughout the United States and a public debate was raging. Sentiment analysis is applied to the sample, clustering the data into topics using the term frequency–inverse document frequency (TF-IDF) technique. The analyses suggest that most (about 77%) of the tweets focused on the search for new/better vaccines for diseases such as the Ebola virus, human papillomavirus (HPV), and the flu. Of the remainder, about half concerned the recent measles outbreak in the United States, and about half were part of ongoing debates between supporters and opponents of vaccination against measles in particular. While these numbers currently suggest a relatively small role for vaccine misinformation, the concept of herd immunity puts that role in context. Nevertheless, going forward, health experts should consider the potential for the increasing spread of falsehoods that may get firmly entrenched in the public mind.

Keywords: healthcare; measles; sentiment analysis; text analysis; vaccination; vaccine; social media; Twitter

1. Introduction

In this research, we use Twitter data to discover and describe public sentiment regarding vaccination. This is timely since there was a large measles outbreak in the United States and other parts of the world in 2019 [1–3]. Additionally, approximately 142,000 people are estimated to have died from measles in 2017 [4]. This could have been prevented with timely vaccination [5]. By understanding the public sentiment about vaccination, government officials and public health policy makers can design more effective communication, education and policy implementation strategies to reach out to the public [6–8]. We explore the following research question in this study: what is the public perception regarding vaccination?

We apply text analysis techniques including word clouds and sentiment analysis on the corpus of tweets from Twitter as a representative of social media [8]. Data were clustered into topics using the term frequency–inverse document frequency (TF-IDF) technique.

The rest of the paper is organized as follows: Section 2 introduces key background information for the research and reviews previous literature; Section 3 discusses the methodology; Section 4 describes the results with a discussion; Section 5 provides the scope and limitations of the research; Section 6 provides an integrative discussion; and finally, Section 7 draws conclusions and discusses implications.

2. Background

The outbreak of infectious diseases, and the potential for vaccination to prevent them, is a rich topic in the study of public opinion and sentiment. There are various social media outlets (e.g., Twitter, blogs, Facebook, Snapchat, etc.) for the public to discuss and disseminate their opinions and experiences. Along these lines, we selected a corpus of text related to vaccine-related comments on Twitter, as a representative social media channel. We selected Twitter since it focuses on key words and allows people to post to a wider audience than other channels. Moreover, Twitter has been successfully used in prior research on blogs, images and textual information [9–15]. While we focused on tweets related to vaccines—and not necessarily to measles-related vaccines—it so happened that since the timeframe of our research coincided with the measles outbreak in the United States, most of the tweets naturally related to it. Therefore, in this section, we first lay the foundation for our research with a discussion on the incidence of measles and the status of vaccination for the same; we then cover the general topic of sentiment towards vaccination; next, we describe the dissemination of such sentiments through the channel of social media, highlighting the relevant paradox of information and misinformation; and finally we outline the potential and appropriateness of sentiment analysis as a technique for our research objective of analyzing public perception in healthcare.

2.1. Increasing Incidence of Measles and Low Vaccination Rates

The outbreak of measles as a viral infection is serious since it is especially dangerous for small children [4,16]. It can cause high fever and a spotty rash over the body. Prior to 1963, measles was responsible for about 400–500 deaths each year [17]. The measles vaccine, which was introduced after 1963, caused measles to be classified as a vaccine-preventable disease throughout the world [4,5,16]. Measles was virtually eliminated in the United States by the year 2000 [18]. The impact of measles in the post-elimination period is evaluated on the basis of the immunological cost, financial cost and the resulting strain on the healthcare system [19]. Unlike other diseases, measles causes post-infection immunosuppression, often making the person susceptible to contracting other bacterial and viral infections for a period of three years after the onset of measles. Additionally, this immunosuppression may also affect an individual's immunological response to other vaccines [19]. Therefore, the risk of mortality and morbidity is escalated. The financial cost of measles includes not just a direct treatment-related costs but also quarantine-related costs. There is a lot of strain on the healthcare system since a response to measles requires the diversion of human resources from other programs and functions. Considered together, these three factors highlight the importance of preventing the spread of measles in an economy.

In this aspect, the Centers for Disease Control and Prevention (CDC) proposes that measles is easily preventable by vaccination [17,20]. A person who carries the virus can spread it through talk, touch and cough. If unvaccinated people are in a room with a person who carries the measles virus (even if that person has left the room), the unvaccinated people in the room have high probability of catching droplets in the air. As a result, those individuals can get infected [21].

There has been a marked increase in the number of measles cases in the United States over the last 10 years. Figure 1 shows the number of measles cases from 2010 to 2019. From 2000 to 2009, fewer than 100 Americans were infected with measles each year. This number grew to 981 cases in May 2019 [22]. According to the CDC the number of patients diagnosed with measles in the U.S. continues to grow [22]. In the first five months of 2019, 981 individual cases of measles were confirmed in 26 states in the U.S [22,23]. Unfortunately, this number continued to grow through August 2019. In Figure 1, the number of cases over a five-month period was higher than in previous years from 2010 to 2018. Moreover, this is the most significant number in the U.S. since 2002. The number of reported cases grew to 1215 by 22 August 2019 [24].

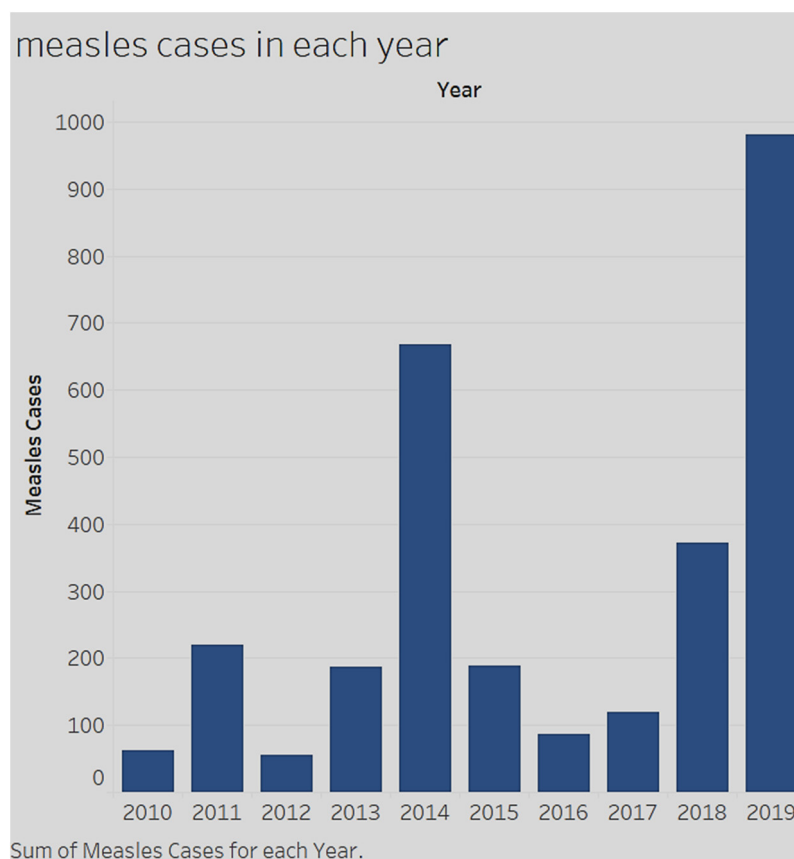


Figure 1. Measles cases in the United States (2010–2019) (<https://www.cdc.gov/measles/cases-outbreaks.html>).

There is heated discussion regarding the 2019 outbreak, especially in the United States [25,26].

Travel provides opportunities for an increase in cases. In the past two years, the number of measles cases increased sharply [25,27]. In 2018, the number increased by 300% globally. According to the World Health Organization (WHO), the countries most affected include Philippines, Madagascar, India, Pakistan, Ukraine, Brazil and Yemen. Travelers from these hotspot countries may carry the virus to the U.S, where its spread is related to vaccination rates [27]. Religious belief against vaccination for measles prevention is another factor in the outbreak [28–30]. Statistics show geographic clusters of the virus. In fact, the resurgence of measles is not widespread across the U.S. but occurs in local areas [31–34]. Seventy-five percent of recent measles cases occurred in close-knit communities in the states of California, Washington, New Jersey and New York [19,34–36]. Moreover, most of these cases (over 600) were localized in orthodox Jewish communities in New York City and its suburb, Rockland. In these communities, people share spaces at school and worship, creating ideal conditions for the virus to spread [19,37].

Another reason for a low vaccination rate is the public's lack of trust in government. A recent example from China shows how a vaccine scandal can shake the public's trust in government and reduce vaccination rates [38–40]. Vaccines have a booming market in China since pharmaceutical innovation is encouraged and prioritized by the government. The industry produces more than \$3 billion in revenue per year [41–43]. However, there are also issues concerning the violation of standards. For example, in one of three Chinese vaccine crises since 2010, Changchun Changsheng, one of the largest drug producers in China, violated standards in the manufacturing of at least 250,000 doses of a vaccine for diphtheria, tetanus and whooping cough. A series of recent vaccine scandals has deepened skepticism about the Chinese government's response; many parents in China have lost faith in vaccines produced in their own country [44,45]. Simultaneously and comparably in the U.S., regardless of the

reasons for the recurrence of the virus in the U.S., the most critical issue affecting its spread is the rate of vaccination in the population [46–48].

2.2. Sentiments towards Vaccination

A vaccine is a biological preparation that provides active acquired immunity to a disease [49–51]. It typically contains an agent which resembles a specific disease-causing microorganism. The agent is derived from weak or dead forms of the microbe, its toxins or one of its surface proteins [49–51]. The human body identifies the agent as a threat and stimulates the immune system to produce antibodies to destroy it [49]. Vaccines work because those antibodies will recognize and destroy the actual disease-causing microorganism if they encounter it in the future [49–51]. Evidence indicates that vaccines are a safe and effective approach to protecting individual and public health [49–51]. However, for a vaccine to work effectively at the population level, a good percentage of a community needs to get immunized [52–54] to achieve so-called “herd immunity” [55–57]. Herd immunity can prevent disease from spreading through populations even when some individuals remain unvaccinated [55–57]. Thus, a mostly vaccinated population protects those of its members who cannot be vaccinated, such as newborns, people with vaccine allergies and cancer patients [55–59].

However, vaccination rates (for diseases such as measles) in the U.S. (e.g., Connecticut) have continued to fall [24]. In fact, an antivaccine sentiment has been building in the U.S. for decades [60–62]. In 2019, the debate over the vaccination requirement in schools was reignited [63–65]. A majority of measles cases reported since the beginning of the outbreak in October 2018 involved unvaccinated children in Hasidic Jewish communities [66–68]. In response to the measles epidemic, in June 2019, New York state lawmakers approved a new law that banned religious exemptions to school vaccination requirements [69]. This law was opposed by thousands of anti-vaccination parents [70].

In addition to attitudes towards vaccination, the spread of misinformation is another prime factor in the resurgence of measles [71]. The anti-vaccination movement, led by a fringe group of parents who oppose the vaccine, believe that, contrary to scientific studies, the vaccine’s chemical makeup can cause autism [72]. It is easy for misinformation to be spread through social media or by word of mouth. A critical factor that contributed to the widespread dissemination of “fake news” was the publication of an article by a gastroenterologist Andrew Wakefield in the medical journal *Lancet* in 1988 [73–75]. The publication discussed 12 children who had pervasive developmental disorders associated with gastrointestinal symptoms [73–75]. According to retrospective accounts by their parents or physicians, eight of the children also had behavioral issues associated with the measles, mumps and rubella (MMR) vaccination [76]. Dr. Wakefield’s study has since been retracted because it was proven to be the result of serious professional misconduct, and most of the children in his original study were revealed to have symptoms that started either before, or much after, the MMR vaccination [73–75]. In addition to the problems with the original study, Dr. Wakefield was found to be engaging in unnecessary, undesirable and harmful procedures on children without any approval from the hospital ethics committee [73–75]. Moreover, he was found to be soliciting financial donations from the Legal Aid Board—a group pursuing legal action for children who were alleged to be vaccine-damaged [73,77–81]. In addition, on behalf of parents, he promoted vaccines to rival the established MMR (measles, mumps, and rubella) vaccine. The British General Medical Council finally revoked Dr. Wakefield’s medical license in the U.K in 2010. Even through the fraudulent results of the research have all been refuted by the scientific community [73–75], the study nevertheless has maintained its influence in social media [82,83]. Additionally, celebrities with “vaccine hesitancy” also got involved, sharing their opinions with their active audiences. This type of movement can be critical, especially when using social networking systems [84]. As a result, the antivaccination sentiment is a byproduct of misinformation on the Internet and remains a driving factor for reduced vaccination rates [85,86].

2.3. Social Media and the Dissemination of (Mis)information

The domain of healthcare, specifically, is characterized by high levels of public concern and therefore, misinformation can spread rampantly via social media. Erroneous information (or misinformation) related to vaccines and vaccine safety has been problematic for many years [87].

On one hand, great progress has been made in national immunization programs thanks to effective vaccines for children. In addition, there is general societal consensus about the rare side effects of vaccines and the view of vaccines as a public good. Throughout the 20th century, the Food and Drug Administration's (FDA) Center for Biologics Evaluation and Research (CBER) has been instrumental in safeguarding scientific discoveries such as vaccines. Furthermore, the vaccine industry became the most closely regulated industry in the U.S. because of severe vaccine misadventures in the 20th century [88]. On the other hand, immunization programs in the U.S and in Europe (U.K, Netherlands, Germany and Switzerland) [89–92] have dealt with many challenges, the predominant one being the dissemination of misinformation on vaccine safety [93,94]. In this regard, there is really very limited evidence-based information to support ideas related to the dangers of vaccines. Therefore, antivaccination activists have typically relied on the power of experience (or anecdotes) to influence large groups of parents with doubts and fears [5,86], through discussions on the Web and social media outlets [95]. In these arenas, accounts of perceived vaccine injury, together with access to Dr. Wakefield's now-retracted research study [75,86], have resulted in a large amount of vaccine hesitancy in parents, including vaccine refusal and delayed vaccine schedules [95]. Added to this is the unregulated nature of the Internet in giving people unlimited freedom of information content and disclosure [85,96]. Researchers have explored online anti-vaccination misinformation by analyzing arguments on anti-vaccination websites, studying the levels of misinformation, and examining discourses to support antivaccination claims [85]. Some studies pointed out that antivaccination websites have high ranks in search engines [97,98]. Moreover, some studies show that people who have strong negative attitudes and opinions are more active in posting information on social media and therefore exert a strong influence on people's attitudes [85,98,99]. In recent years, social media has become a prime channel for distributed global communication. As the demand for transparency in the experiences from vaccination had increased, social media has evolved to become a platform not only for patient engagement but also for empowerment. Social networks play a huge role in modulating individual health behaviors [100,101]. Twitter, for one, has been a platform for individuals to show their utterances to the public. Currently, 21% of U.S. adults use Twitter; forty-two percent of those users visit the Twitter platform on a daily basis [102]. Currently, there are over 69 million monthly active Twitter users in the U.S. [9]. Formal evaluations of vaccine studies can benefit from incorporating social media discussions of vaccine users as complementary evaluations [10]. As a result, there is rich potential for extracting health information and studying the dynamics of health behaviors on Twitter [103].

2.4. Sentiment Analysis

To this effect, sentiment analysis has been deployed as a natural language processing task at varying degrees of granularity. Natural language processing began with performing classification tasks on a document level [104,105]. Initially, analysis was handled at the sentence level [106,107], and later progressed to a phrase level [108]. In recent years, Agarwal et al. [109] brought natural language processing to a new level, using lexical scoring with n-gram analysis to model the effect of context in order to perform phrase-level polarity and n-gram analysis. Considering the advances in natural language processing and text analytics, sentiment analysis has increasingly become a beneficial tool to explore individuals' attitude patterns toward vaccines and vaccination in particular, and public health in general [109]. Several studies have applied sentiment analysis to explore sentiment trends about vaccination through social media like Twitter. For example, Du et al. [110] applied machine learning-based approaches to examine sentiment trends using Twitter data [110]. Surian et al. conducted topic modeling and community detection to identify negative sentiments about Human Papillomavirus (HPV) vaccines on Twitter [11]. Salathé and Khandelwal [8] defined a score

for sentiment on the H1N1 vaccination. Radzikowski et al. [12] applied a quantitative analysis to determine a measles vaccination narrative on Twitter [12].

This article augments this literature by employing the natural language toolkit (NLTK) to perform sentiment analysis to explore patterns and trends among collected Twitter data to understand the opinions of people towards measles vaccination. Insights from the study help formulate and shape public policy with regard to prevention of measles.

3. Methods

In this article, a corpus of vaccine-related text messages on Twitter (tweets) is examined. We selected Twitter as our social media platform since it focuses on key words and allows posting to a wider audience when compared to other platforms like Facebook. We did a random selection of six tweets as a pilot, in order to get initial insight into the various opinions on vaccines. We performed several analyses on vaccine-related tweets, including word count frequency, word cloud generation, word co-occurrence, TF-IDF clustering and sentiment scoring of tweets [111,112]. The TF-IDF technique searches through a corpus of documents to determine key words. The text data are then transformed into a TF-IDF matrix to provide the frequency of a word in the corpus. Collectively, these text analyses have been demonstrated to be comparable to traditional survey analytical methods in describing individual opinions on a vaccine [6–8,13,14,113,114]. Evaluation of research adopting sentiment analyses for vaccine-related tweets have created an awareness of the need for disseminating vaccine-related information.

This study uses a sample of 9581 vaccine-related tweets in the period 1 January 2019 to 5 April 2019. This time period is of the essence because during this time a measles outbreak was prevalent in the United States and public debate on vaccination was raging. We therefore decided to zoom in and capture public perception during this peak period. All the tweets with the keyword “vaccine” were scraped globally. Only tweets in English were included. They were coded in Python programming language to conduct vaccine-related sentiment analysis. The text data were put into data frames using Pandas packages in Python. Pandas package is an open source library that provides easy-to-use, high performance data structures and data analysis tools for Python language. The NLTK (Natural Language ToolKit), which is a suite of libraries and programs for natural language processing written in the Python programming language was used. In the next step, three stemmers (WordNet, Lancaster Stemmer and Porter Stemmer) in the NLTK were used to prune the text data. The three stemmers provided a comparison for improved performance. The Valence Aware Dictionary and Sentiment Reasoner (VADER) was used to handle sentiment analysis on the text data. In sentiment analysis, each sentence is analyzed to determine a sentiment score based on whether it conveys a positive, negative or neutral sentiment. VADER is based on lexicons of sentiment-related words. Each word will have a sentiment rating—for example a positive word (e.g., “good”) has a positive sentiment rating of 1.9. Words that are more positive words will have a higher sentiment rating (e.g., the word “great” will have a higher rating than “good”). The same rule applies to words that convey negative sentiments. The outcome metric has four parts: positive sentiment score; neutral sentiment score; negative sentiment score; and compound score. The compound score is the sum of the lexicon ratings, which are standardized values between -1 and 1 . The compound score was used as a classifier. A tweet with a compound score more than or equal to 0.05 is identified as a positive sentiment tweet. A tweet with a compound score less than or equal to -0.05 is identified as a negative sentiment tweet. Other tweets are identified as neutral sentiment tweets. In the final step, K-means was used to perform clustering on the text data. Skikit-learn, a software machine learning library for the Python programming language, was utilized to cluster documents by topics, using a bag-of-words approach. Initially, features were extracted using the TF-IDF method since some words were more frequently used within the large text corpus (e.g., “the”, “I”, “a”, “are”). These words carried little-to-no meaningful information on the actual content of the document. If one applies the direct count data to a classifier, the frequency of these terms would overshadow the frequencies of rarer, more interesting terms. Therefore,

this study aimed to reweigh the count features into the classifier to drill into the text data. As a result, TF-IDF was introduced as part of the methodology. TfidfVectorizer can convert a collection of raw documents into a matrix of TF-IDF features. Using an in-memory vocabulary to map the most frequent words to a features index, it computes a word occurrence frequency matrix. The word frequencies are reweighted using the IDF vector collected feature-wise over the corpus. Based on the TF-IDF sparse matrix, the authors used k-means for clustering. K-means is suitable for unsupervised data (or data without a label). The algorithm based on the extracted features works iteratively to assign each data point to a k-group. Overall, there were six attributes in the final dataset: (1) username; (2) content of the tweet; (3) number of replies; (4) number of retweets; (5) number of favorites; and (6) published date. The size of the final dataset was $6 \times 9581 = 57,486$. The final dataset was a comma-separated values (CSV) file. All words in the tweet content were changed to lower case. All “/n” were replaced by “.” The uniform resource locators (URLs) of websites in tweet content were removed because the analysis was based solely on the text data and most of the URLs were links to pictures. In the final step, the tweet content was split into individual words to transform them into vector form data that can be read by the computer.

4. Results

The six randomly selected tweets provided an initial insight into the various opinions on vaccines (see Figure 2). In this figure, there is a heated debate between antivaccine activists and vaccine supporters. From the tweets, it appears that there is some public perception that vaccination is linked to autism. Three of the six tweets are skeptical as to whether vaccines cause autism. Two hold the opposite opinion, stating that “vaccines cause autism”. Another supports antivaccination. Among the other three tweets, one utterance shows that people are not really against vaccination. Two of the tweets share news about vaccines.



Figure 2. Six randomly selected tweets in the dataset.

In the first step, the NLTK package was utilized to tokenize content data. While text words were retained, numbers were removed. Next, a frequency dictionary was generated based on tokens. The frequency list was sorted in descending order. Figure 3 shows the top 30 frequency words in the tokens. Most of the frequency words are stop words, that is, commonly used words. Only two nouns

("vaccine" and "vaccines") are among the top 10 keywords. With this result, there can be no insight gained from the analysis. As a result, stop words were routinely removed.

The NLTK stop words list was used to identify stop words (<https://nlp.stanford.edu/IR-book/html/htmledition/dropping-common-terms-stop-words-1.html>). After removing the stop words, the remaining words were added to a list named "word_nostopwords". A list was again generated. There are 30 frequency words in tokens without stop words (see Figure 4). The words "vaccine" and "vaccines" indicate the same content. In Figure 4, "measles," "children," and "autism" are the top three frequency words. However, words such as "kids" and "child" also frequently appear. As a result, the study found that "measles" is the most frequently used word. "Children" has the same meaning in all its appearances in the content. Therefore, these words were also stemmed.

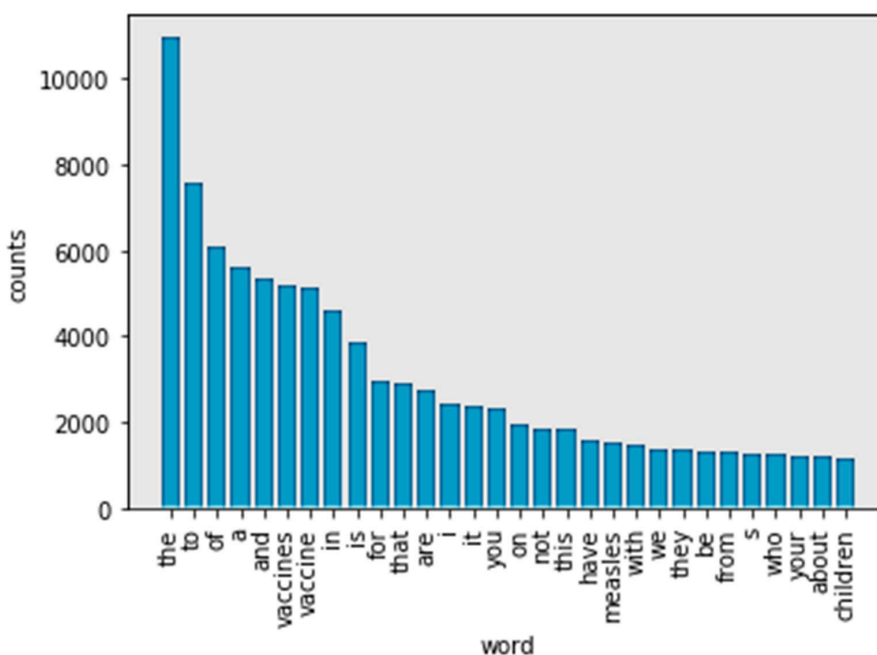


Figure 3. Top 30 frequency words in tokens.

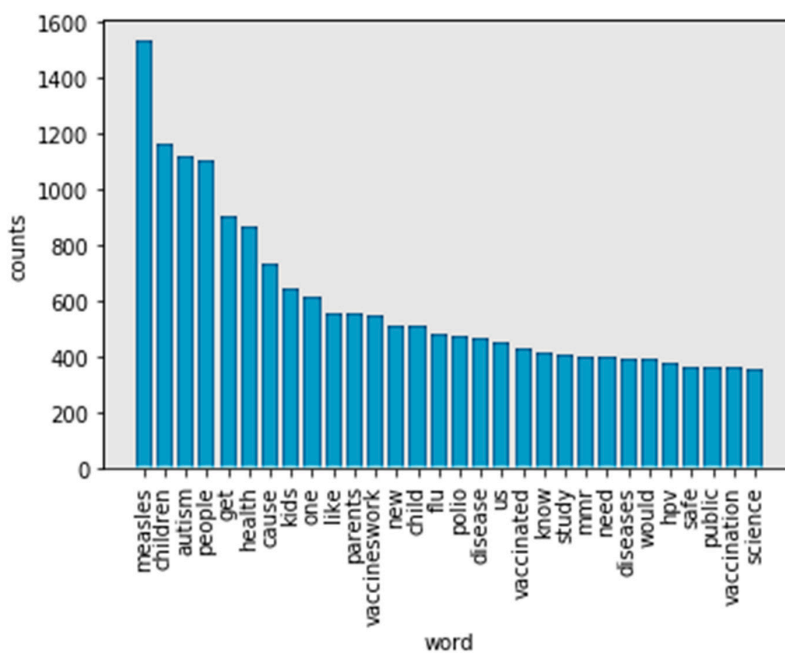


Figure 4. Top 30 frequency words in tokens without stop words.

Stemming refers to a simple, heuristic process that cuts off the ending of words in order to achieve a goal. Most often, this can include eliminating derivational affixes. Porter stemmer, Lancaster stemmer, and Wordnet are three techniques for stemming words. In Figure 5, the Porter stemmer did not solve the problem caused by words with a “similar” meaning. “Children” and “child” still appear separately in the frequency list.

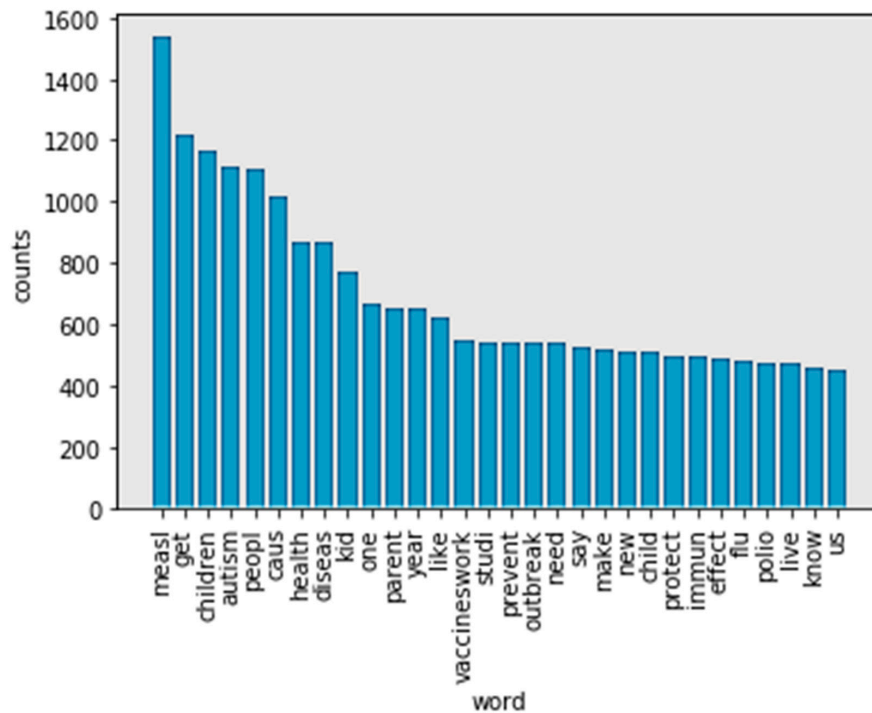


Figure 5. Top 30 frequency words in no-stop words tokens after processing by Porter stemmer.

This problem still appears in Figure 6. Using the Lancaster stemmer, the top 30 frequency words in the no-stop words tokens still show both “childr” and “child”.

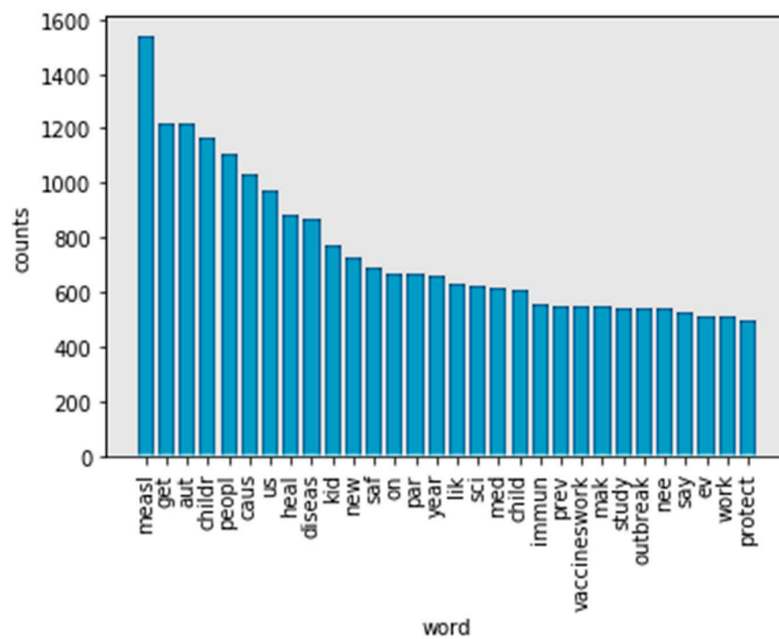


Figure 6. Top 30 frequency words in no-stop words tokens after processing by Lancaster stemmer.

Figure 7 (WordNet) has a different most frequent word compared to Figure 5 (Porter stemmer) and Figure 6 (Lancaster stemmer). For example, “child” is the most frequent word in tokens. This is more reasonable. In comparing Figure 5 (Porter stemmer) and Figure 6 (Lancaster stemmer), there are fewer words that have overlapping meanings using WordNet as a stemmer. In this case, WordNet achieved a better performance. The top 30 words in tokens mention “measles” (1534), “autism” (1117), “flu” (482) and “polio” (472), respectively. The words “child” (1681), “people” (1110), “kid” (769), and “parent” (631) are the most-mentioned figures, respectively. “Vaccineswork” is a frequently mentioned positive sentiment word in this study (549 mentions in 9581 tweets).

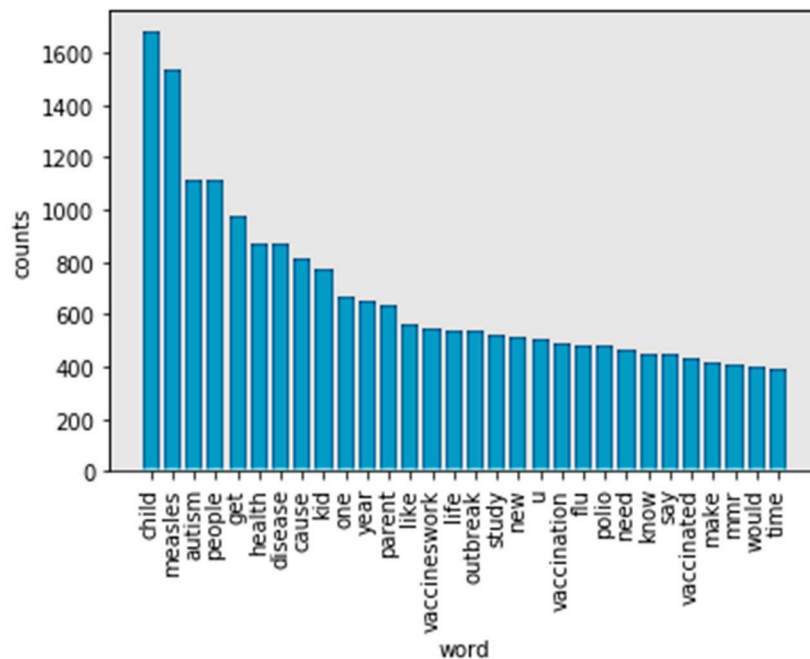


Figure 7. Top 30 frequency words in no-stop words tokens after processing by WordNet.

The word cloud in Figure 8 displays the words most frequently found in the corpus of tweets. The larger the word’s size in the cloud, the more often it occurs in the corpus. The figure shows that words like “measle,” “children,” “autism,” and “parents” are distinct in the case data set. “Measle” and “autism” are related to disease. “Children” and “parents” are the entities who use Twitter to talk about vaccines.

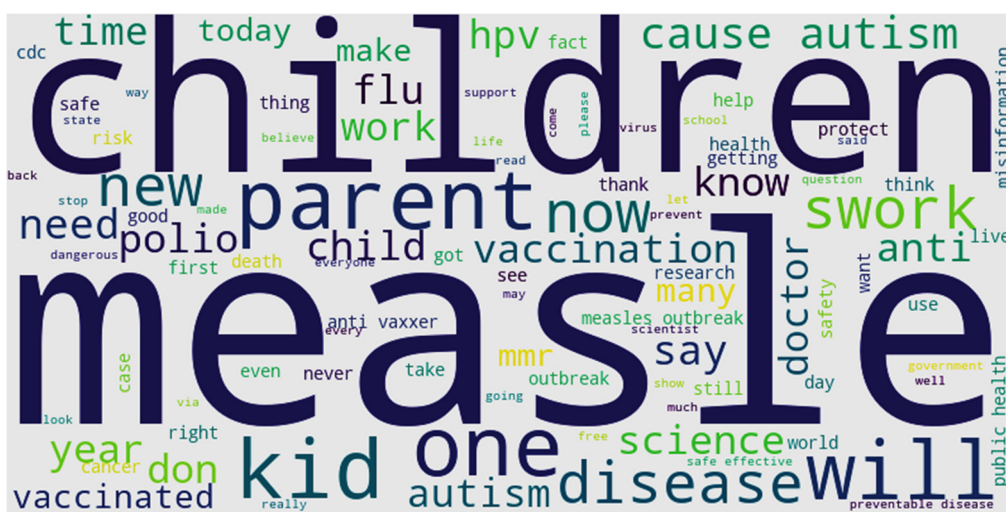


Figure 8. Word cloud of content.

To understand the relationships between frequency words, the study identified word pairs (bigram and trigram). It counted the number of times a word pair appeared. The bigram and trigram make (statistical) predictions about what is happening in a sentence. They help researchers understand vaccine-related data extracted from Twitter. For example, a particular word may appear or an element belonging to a particular word class may appear. A bigram makes a word prediction based on the prior word. A trigram makes a word prediction based on the two prior words. Figure 9 (bigram) and Figure 10 (trigram) show the top 10 bigram and trigram phrases and counts. The words “vaccine cause” (529), “cause autism” (444) and “measles vaccine” (307) are the top three terms in the bigram phrases.

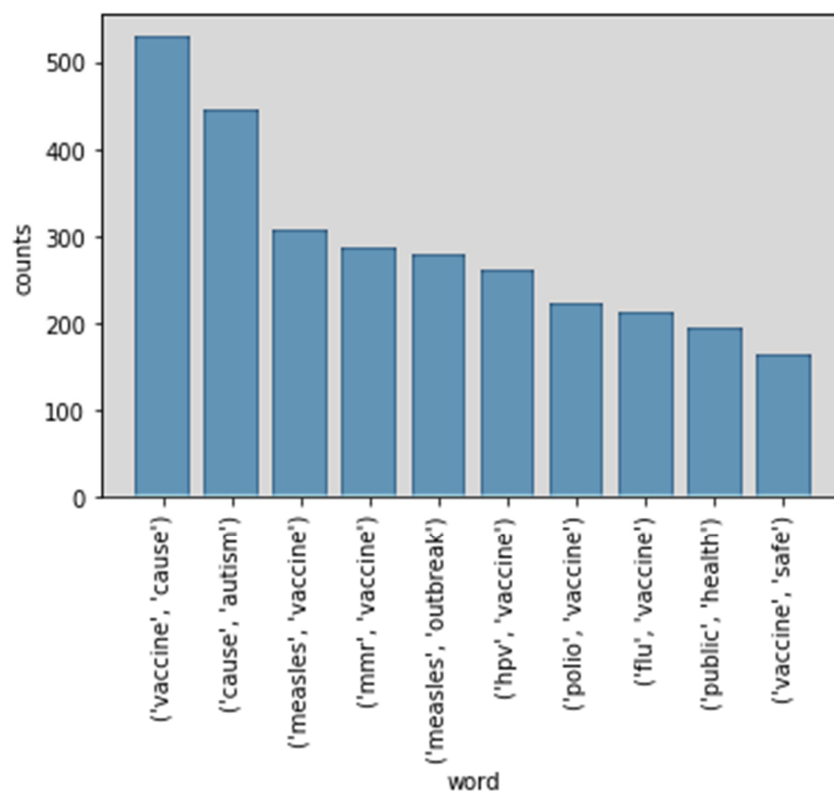


Figure 9. Top 10 bigram phrases.

Figure 10 shows the top 10 trigram phrases in vaccine-related tweets. “Vaccine cause autism” (380), “vaccine save life” (87), and “vaccine safe effective” (66) are the top three trigram phrases. There are three negative phrases among the top 10 trigram phrases. The first, “vaccine cause autism,” is also the most frequent word (380 mentions). The other two negative phrases are “link vaccine autism” (41) and “vaccine injured child” (30). Positive phrases in the top 10 trigram phrases are “vaccine save life” (87), “vaccine safe effective” (66) and “vaccine preventable disease” (34). However, the number of all the positive phrases is less than the number of “vaccine cause autism.”

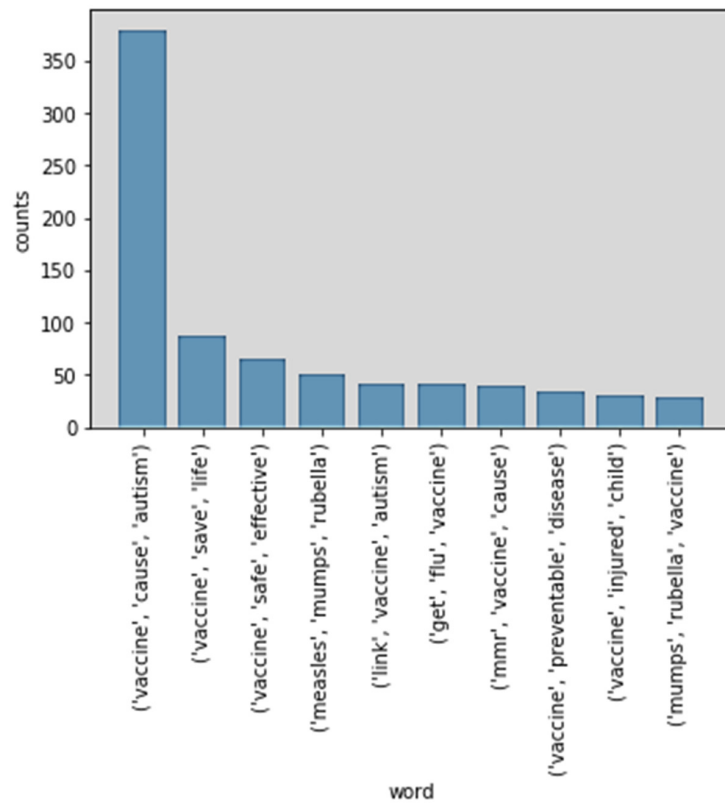


Figure 10. Top 10 trigram phrases.

Of the 9581 vaccine-related tweets in this dataset, 4151 were classified as negative sentiment tweets, 3869 tweets were positive sentiment tweets and 1561 tweets were neutral sentiment (see Figure 11). According to this pie chart, 43.3% were negative tweets, 40.4% were positive tweets, and 16.3% were neutral tweets. The number of negative tweets was slightly higher than positive tweets.

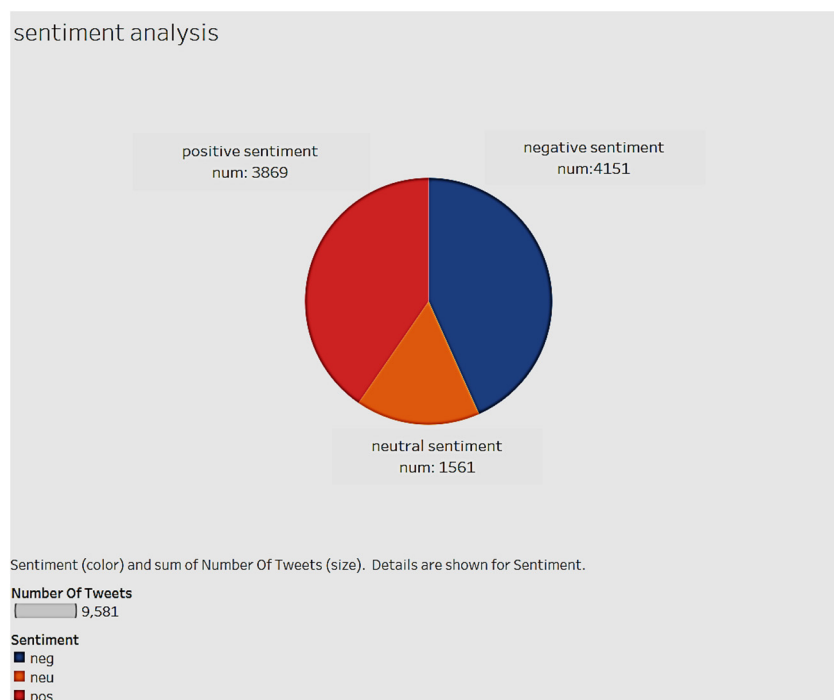


Figure 11. Positive, neutral and negative sentiment tweets in the data set.

The TF-IDF technique searches through a corpus of documents to determine which words are favorable for a query. This study transforms the text data into a TF-IDF matrix to provide the frequency of a word in the corpus.

Term Frequency (tf): This provides the frequency of words in each document in the corpus. It is the ratio of the number of times the word appears in a document compared to the total number of words in that document. TF increases as the number of occurrences of the word increases within the document. Each document has its own term frequency (Figure 12).

Inverse Data Frequency (idf): This measure is a calculation of the weight of rare words across all documents in the corpus. Words that rarely occur in the corpus have a high IDF score (Figure 13).

The product of TF and IDF gives the TF-IDF score for a word in a document in the corpus (Figure 14). A high TF-IDF indicates a strong relationship with the document in which it occurs.

```
tfBowA = computeTF(wordDictA, content2)
tfBowA

{'together.': 0.045454545454545456,
 '': 0.0,
 'they': 0.13636363636363635,
 'a': 0.0,
 'i': 0.0,
 'how': 0.09090909090909091,
 'even': 0.045454545454545456,
 '-': 0.0,
 '(official': 0.0,
 'multiple': 0.045454545454545456,
 'of': 0.09090909090909091,
 '@youtube': 0.0,
 'the': 0.045454545454545456,
 'study': 0.045454545454545456,
 'video': 0.0,
 'do': 0.09090909090909091,
 'playlist': 0.0,
 'quit': 0.0,
 'vaccines': 0.0,
 'effects': 0.045454545454545456,
 'babies': 0.045454545454545456,
 'video)': 0.0,
 'get': 0.045454545454545456,
 'tons': 0.045454545454545456,
 'react': 0.045454545454545456,
 'to': 0.0,
 "can't": 0.0,
 'added': 0.0,
 'know': 0.045454545454545456,
 'vaccines.': 0.09090909090909091}
```

Figure 12. Example of a term frequency score of Document 1 in the dataset.

```

idfs = computeIDF([wordDictA, wordDictB])
idfs
{'together.': 0.3010299956639812,
 ': 0.3010299956639812,
 'they': 0.3010299956639812,
 'a': 0.3010299956639812,
 'i': 0.3010299956639812,
 'how': 0.3010299956639812,
 'even': 0.3010299956639812,
 '-': 0.3010299956639812,
 '(official': 0.3010299956639812,
 'multiple': 0.3010299956639812,
 'of': 0.3010299956639812,
 '@youtube': 0.3010299956639812,
 'the': 0.0,
 'study': 0.3010299956639812,
 'video': 0.3010299956639812,
 'do': 0.3010299956639812,
 'playlist': 0.3010299956639812,
 'quit': 0.3010299956639812,
 'vaccines': 0.3010299956639812,
 'effects': 0.3010299956639812,
 'babies': 0.3010299956639812,
 'video)': 0.3010299956639812,
 'get': 0.3010299956639812,
 'tons': 0.3010299956639812,
 'react': 0.3010299956639812,
 'to': 0.3010299956639812,
 "can't": 0.3010299956639812,
 'added': 0.3010299956639812,
 'know': 0.3010299956639812,
 'vaccines.': 0.3010299956639812}

```

Figure 13. Example of inverse data frequency (IDF) scores in Document 1 and 2 of the data set.

```

tfidfBowA = computeTFIDF(tfBowA, idfs)
tfidfBowA
{'together.': 0.013683181621090055,
 ': 0.0,
 'they': 0.04104954486327016,
 'a': 0.0,
 'i': 0.0,
 'how': 0.02736636324218011,
 'even': 0.013683181621090055,
 '-': 0.0,
 '(official': 0.0,
 'multiple': 0.013683181621090055,
 'of': 0.02736636324218011,
 '@youtube': 0.0,
 'the': 0.0,
 'study': 0.013683181621090055,
 'video': 0.0,
 'do': 0.02736636324218011,
 'playlist': 0.0,
 'quit': 0.0,
 'vaccines': 0.0,
 'effects': 0.013683181621090055,
 'babies': 0.013683181621090055,
 'video)': 0.0,
 'get': 0.013683181621090055,
 'tons': 0.013683181621090055,
 'react': 0.013683181621090055,
 'to': 0.0,
 "can't": 0.0,
 'added': 0.0,
 'know': 0.013683181621090055,
 'vaccines.': 0.02736636324218011}

```

Figure 14. Example of the term frequency–inverse document frequency (TF-IDF) score from Document 1 in the data set.

Using the TF-IDF matrix, clustering algorithms can be run to better understand the hidden structure within a data set. In this study, k-means clustering was used. K-means initializes with a predetermined number of clusters (for example, three). In order to minimize the within-cluster sum of squares, each observation is assigned a cluster (or cluster assignment). Next, the mean of the clustered observations is calculated and used as the new cluster centroid. After this, observations get reassigned to clusters, and centroids get recalculated. This is done iteratively until the algorithm reaches convergence. Figure 15 shows the number of tweets per cluster. Cluster 1 has the greatest number of tweets (7361). Clusters 2 and 3 have 1076 tweets and 1145 tweets, respectively.

Figure 16 shows the word cloud of Cluster 1. Words like “study,” “research,” and “program” show that utterances in Cluster 1 discuss studies or innovations in the field of vaccines. Words like “mandatory” and “mandate” are attitudes surrounding vaccinations. Words such as “Ebola,” “HPV,” and “flu” show the diseases being worked on by researchers. Vaccine breakthroughs related to these diseases will draw the public’s attention. Twitter users in the first group discuss breakthroughs related to vaccines and patients. Tweets in Cluster 1 have the greatest count that do not share debatable information or misinformation about vaccines. Most utterances in the vaccine-related tweets discuss scientific issues.

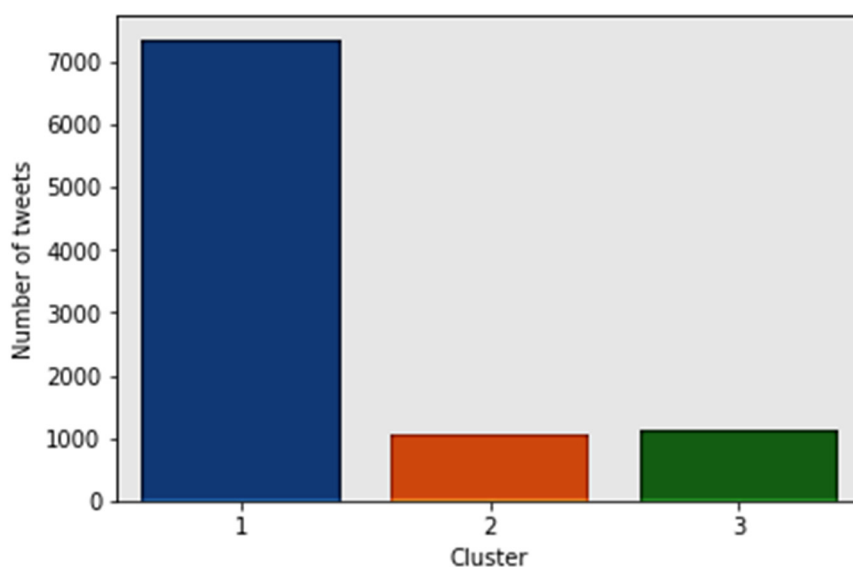


Figure 15. Tweets per cluster.

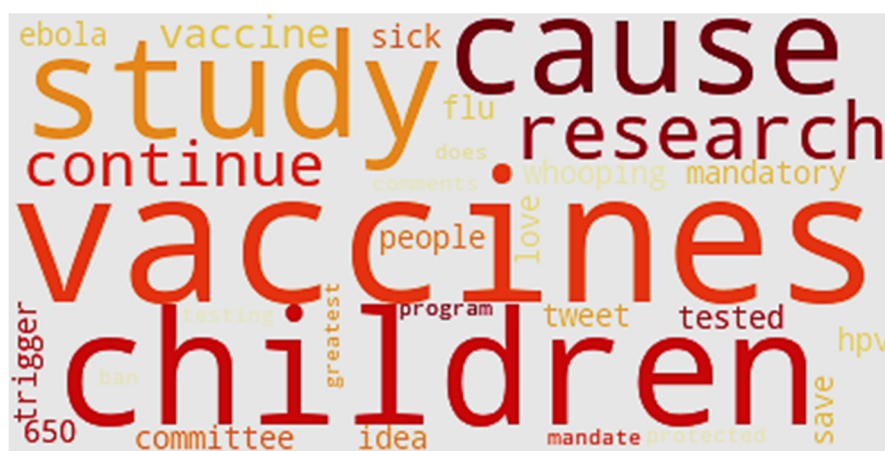


Figure 16. Word cloud of Cluster 1.

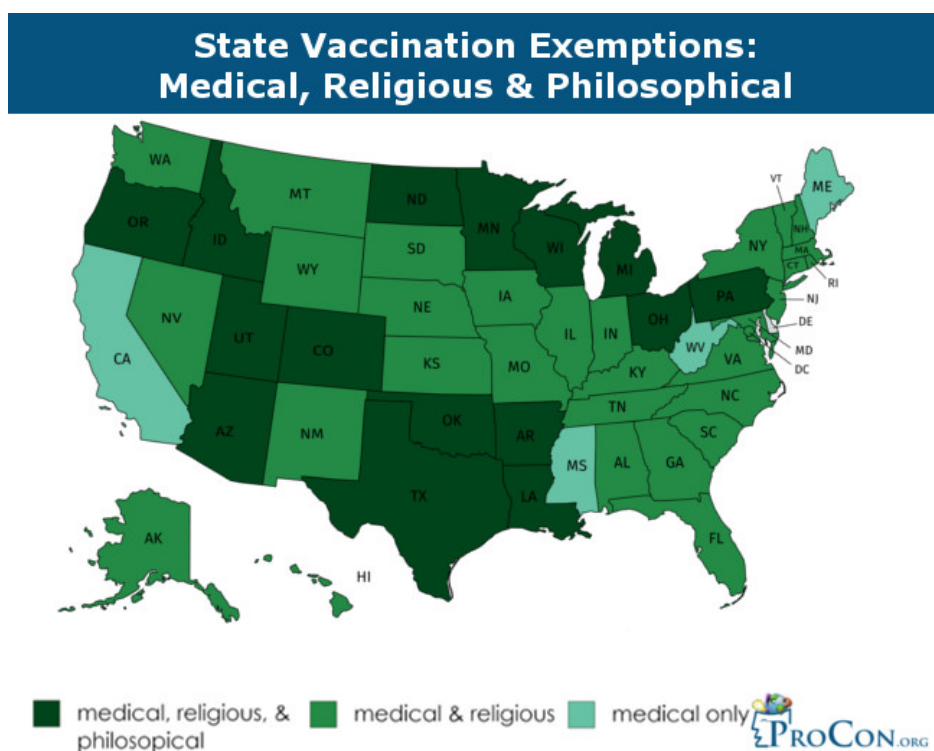


Figure 19. Vaccination exemptions by state (2018).

5. Discussion

This is one of the few studies to analyze online sentiment, word usage and attitudes on the measles vaccine using text mining and sentiment analysis on social media, specifically Twitter. The focus is on information related to vaccines and measles, which also includes health misinformation. The method can also be applied to other health domains and areas. Findings show that Twitter discussions focus on vaccine-related topics of measles, children, autism and parents, demonstrating public concern in these areas. The number of tweets with negative sentiments was only slightly higher than those with positive or neutral sentiments. The negative sentiments mostly centered on the link between vaccine and autism, the vaccine being a cause for autism, and the vaccine causing injury to children. The positive sentiments related to the existence of a vaccine for measles, the vaccine being effective and the vaccine actually saving lives. In this context, we need to highlight that all the tweets were analyzed, regardless of whether they originated from one or multiple users. The discussions converge in three holistic clusters: discussions on innovations in the arena of vaccines; discussions on outbreaks in the U.S; and frequency discussions on medical exemptions of vaccines by the states in the U.S. This depicts public concern on a range of issues related to disease and disease prevention, thus offering a lens into the level of awareness of public health.

It is interesting to explore factors that can contribute to the online posting of negative sentiments. In an empirical study of Facebook users, it was demonstrated that positive information gets disseminated fast but does not sustain as long as negative information [15,116–118]. In this context, future studies can investigate whether there is an optimal period in which information can be presented online to create a positive influence and keep it active in memory. It is also worth studying whether people post negative sentiments on vaccines just as an attention-seeking gesture of offering radically differing opinions. Particularly in healthcare, it is worth looking at a means to motivate people with positive sentiments to remain active and contribute more online. Positive emotions have been suggested to incite people to consider long-term benefits over short-term costs [119–121]. Lastly, considering the affinity of users in different age groups to certain platforms, future studies can incorporate hybrid methods involving multiple platforms to be able to compare sentiments across age groups and across platforms.

6. Scope and Limitations

This study is not without limitations. First, it is a snapshot in time. Public sentiment may change over time. Second, the sentiments expressed in the tweets may not be truthful and may introduce elements of bias into the data. Third, the subjectivity introduced in the use of various techniques such as stop word, stemming, etc., may impact the results. Long-term research should address generalizability and reproducibility of the research. Fourth, sample size is an issue. Future studies may include larger sample sizes. In the data analysis portion of the research, the study used bigram and trigram to study the data. Using the data set, the study concluded that people talked most about “vaccine cause autism.” However, the study cannot show that every tweet with “vaccine cause autism” is a negative sentiment. The six randomly chosen tweets from the data set showed that half of the tweets mentioned “vaccine cause autism.” However, two of the three tweets criticized the opinion, meaning that they sided with the belief that “vaccine save life.” Based on the generated clusters, the study performed a sentiment analysis with each cluster. The word cloud clusters were not persuasive. Additional statistical analysis may shed more light on this.

Nevertheless, the study surfaces the key reasons for vaccine resistance and its direct association with the spread of measles. Additionally, the study is one of the few to demonstrate the efficacy and usefulness of machine learning and text analytics in conducting rapid studies of large sets of social media data to gain insight into public opinion in the public health and infectious disease arenas.

7. Conclusions

Overall, the results in this study have promising implications. First, while social media chatter is more about the association between vaccination and autism (misinformation discussed previously), much of the public discussion is about how vaccination can save lives and is safe and effective. The sentiment is that vaccination prevents disease. We highlight this important revelation. At the same time, we urge scientists and public health policy officials to be on high alert regarding the potential for the rapid spread of misinformation, as well as challenging falsehoods, that tend to get firmly entrenched in the public mind (e.g., autism). Second, considering the current era of social media, it is imperative that stakeholders and government officials the world over—being informed through monitoring social media discussion and sentiment—engage the public aggressively and continually in risk communication and education. Third, we show that there is a positive trend in attitudes towards public health, as revealed in the discussion of the role of overall science and research in regard to vaccination, and the association between measles, contagion and a lack of vaccination. Fourth, as old infectious diseases surface again (e.g., Ebola) and new viruses emerge (e.g., COVID-19), and rapid progress is made in the development of new vaccines, we suggest that policy makers be fully mindful of the fact that health is in the context of society. Therefore, paying attention to what the public thinks can lead to informed decisions. Lastly, the use of advanced technology such as machine learning, natural language processing, sentiment analysis and text analysis can accelerate the maturing process of understanding public opinion in the context of the spread of infectious diseases.

Author Contributions: Both the authors contributed equally to the data analysis, design and development of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cousins, S. Measles: A global resurgence. *Lancet Infect. Dis.* **2019**, *19*, 362. [CrossRef]
2. Karasz, P.W.H.O. Warns of ‘Dramatic’ Rise in Measles in Europe. *The New York Times*. 30 August 2019. Available online: <https://www.nytimes.com/2019/08/29/world/europe/measles-uk-czech-greece-albania.html?searchResultPosition=1> (accessed on 10 January 2020).

3. Larson, K. WHO: Death Toll From Measles Outbreak in Congo Hits 6000. *Associated Press*. 7 January 2020. Available online: <https://www.usnews.com/news/world/articles/2020-01-07/who-death-toll-from-measles-outbreak-in-congo-hits-6-000> (accessed on 10 January 2020).
4. Kupferschmidt, K. Study pushes emergence of measles back to antiquity. *Science* **2020**, *367*, 11–12. [CrossRef]
5. Hussain, A.; Ali, S.; Ahmed, M.; Hussain, S. The anti-vaccination movement: A regression in modern medicine. *Cureus* **2018**, *10*, 1–8. [CrossRef]
6. Mitra, T.; Counts, S.; Pennebaker, J.W. Understanding anti-vaccination attitudes in social media. In Proceedings of the Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.
7. Numnark, S.; Ingsriswang, S.; Wichadakul, D. VaccineWatch: A monitoring system of vaccine messages from social media data. In Proceedings of the 8th International Conference on Systems Biology (ISB), Qingdao, China, 24–27 October 2014.
8. Salathé, M.; Khandelwal, S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Comput. Biol.* **2011**, *7*, e1002199.
9. The Statistics Portal. Twitter: Number of Monthly Active U.S. Users 2010–2018. Available online: <https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/> (accessed on 20 February 2020).
10. Sewalk, K.C.; Tuli, G.; Hswen, Y.; Brownstein, J.S.; Hawkins, J.B. Using Twitter to examine Web-based patient experience sentiments in the United States: Longitudinal study. *J. Med. Internet Res.* **2018**, *20*, e10043. [CrossRef] [PubMed]
11. Surian, D.; Nguyen, D.Q.; Kennedy, G.; Johnson, M.; Coiera, E.; Dunn, A.G. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *J. Med. Internet Res.* **2016**, *18*, e232. [CrossRef] [PubMed]
12. Radzikowski, J.; Stefanidis, A.; Jacobsen, K.H.; Croitoru, A.; Crooks, A.; Delamater, P.L. The measles vaccination narrative in Twitter: A quantitative analysis. *Jmir Public Health Surveill.* **2016**, *2*, e1. [CrossRef] [PubMed]
13. Chen, T.; Dredze, M. Vaccine images on twitter: Analysis of what images are shared. *J. Med. Internet Res.* **2018**, *20*, e130. [CrossRef]
14. Zhou, X.; Coiera, E.; Tsafnat, G.; Arachi, D.; Ong, M.S.; Dunn, A.G. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Stud. Health Technol. Inf.* **2015**, *216*, 761–765.
15. Dredze, M.; Wood-Doughty, Z.; Quinn, S.C.; Broniatowski, D.A. Vaccine opponents' use of Twitter during the 2016 US presidential election: Implications for practice and policy. *Vaccine* **2017**, *35*, 4670–4672. [CrossRef]
16. Thakkar, N.; Gilani SS, A.; Hasan, Q.; McCarthy, K.A. Decreasing measles burden by optimizing campaign timing. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11069–11073. [CrossRef]
17. Centers for Disease Control and Prevention. Chapter 7: Measles. Available online: <https://www.cdc.gov/vaccines/pubs/surv-manual/chpt07-measles.html> (accessed on 25 April 2020).
18. Brown, D. In 2000, Measles Had Been Officially 'Eliminated' in the U.S. Will That Change? *The Washington Post*. 28 September 2019. Available online: https://www.washingtonpost.com/health/in-2000-measles-had-been-officially-eliminated-in-the-us-will-that-change/2019/09/27/56ac487a-deed-11e9-b199-f638bf2c340f_story.html (accessed on 10 January 2020).
19. Sundaram, M.E.; Guterman, L.B.; Omer, S.B. The true cost of measles outbreaks during the post elimination era. *JAMA* **2019**, *321*, 1155. [CrossRef]
20. National Institute of Health. Decline in Measles Vaccination is Causing a Preventable Global Resurgence of the Disease. Available online: <https://www.nih.gov/news-events/news-releases/decline-measles-vaccination-causing-preventable-global-resurgence-disease> (accessed on 25 April 2020).
21. Centers for Disease Control and Prevention. Transmission of Measles. Available online: <https://www.cdc.gov/measles/transmission.html> (accessed on 25 April 2020).
22. Centers for Disease Control and Prevention. Measles Cases and Outbreaks. Available online: <https://www.cdc.gov/measles/cases-outbreaks.html> (accessed on 25 April 2020).
23. Centers for Disease Control and Prevention. National Updates on Measles Cases and Outbreaks—United States, January 1–October 1, 2019. Available online: <https://www.cdc.gov/mmwr/volumes/68/wr/mm6840e2.htm> (accessed on 25 April 2020).

24. Avila, J.D. Connecticut's Measles Vaccination Rates Keep Falling. *The Wall Street Journal*. 29 August 2019. A9A. Available online: <https://www.wsj.com/articles/connecticuts-measles-vaccination-rates-keep-falling-11567109137> (accessed on 10 January 2020).
25. Morbidity and Mortality Weekly Report (2 May 2019). Increase in Measles Cases—United States, 1 January–26 April 2019. Available online: <https://www.cdc.gov/mmwr/volumes/68/wr/mm6817e1.htm> (accessed on 25 April 2020).
26. Gonzales, R. CDC Reports Largest U.S. Measles Outbreak Since Year 2000. Available online: <https://www.npr.org/2019/04/24/716953746/cdc-reports-largest-u-s-measles-outbreak-since-year-2000> (accessed on 25 April 2020).
27. Jost, M.; Luzi, D.; Metzler, S.; Miran, B.; Mutsch, M. Measles associated with international travel in the region of the Americas, Australia and Europe, 2001–2013: A systematic review. *Travel Med. Infect. Dis.* **2015**, *13*, 10–18. [[CrossRef](#)]
28. CDC Notes from the field: Measles outbreak among members of a religious community—Brooklyn, New York, March–June 2013. *MMWR* **2013**, *62*, 752–753.
29. Salmon, D.A.; Haber, M.; Gangarosa, E.J.; Phillips, L.; Smith, N.J.; Chen, R.T. Health consequences of religious and philosophical exemptions from immunization laws: Individual and societal risk of measles. *JAMA* **1999**, *282*, 47–53. [[CrossRef](#)]
30. Wombwell, E.; Fangman, M.T.; Yoder, A.K. Religious Barriers to Measles Vaccination. *J. Commun. Health* **2015**, *40*, 597–604. [[CrossRef](#)]
31. Fiebelkorn, P.A.; Redd, S.B.; Gallagher, K.; Rota, P.A.; Rota, J.; Bellini, W.; Seward, J. Measles in the United States during the post elimination era. *J. Infect. Dis.* **2010**, *202*, 1520–1528. [[CrossRef](#)]
32. Opel, D.J.; Omer, S.B. Measles, mandates, and making vaccination the default option. *JAMA Pediatrics* **2015**, *169*, 303–304. [[CrossRef](#)]
33. Phadke, V.K.; Bednarczyk, R.A.; Salmon, D.A.; Omer, S.B. Association between vaccine refusal and vaccine-preventable diseases in the United States: A review of measles and pertussis. *JAMA* **2016**, *315*, 1149–1158. [[CrossRef](#)]
34. Stanley-Becker, I. Officials in anti-vaccination 'hotspot' near Portland declare an emergency over measles outbreak. *The Washington Post*, 23 January 2019.
35. Jackson, M.A.; Harrison, C. On the Brink: Why the US is in Danger of Losing Measles Elimination Status. *Mol. Med.* **2019**, *116*, 260.
36. Papania, M.J.; Wallace, G.S.; Rota, P.A.; Icenogle, J.P.; Fiebelkorn, A.P.; Armstrong, G.L.; Hao, L. Elimination of endemic measles, rubella, and congenital rubella syndrome from the Western hemisphere: The US experience. *JAMA Pediatrics* **2014**, *168*, 148–155. [[CrossRef](#)]
37. Patel, M.; Lee, A.D.; Clemmons, N.S.; Redd, S.B.; Poser, S.; Blog, D.; Arciuolo, R.J. National update on measles cases and outbreaks—United States, January 1–October 1, 2019. *Morb. Mortal. Wkly. Rep.* **2019**, *68*, 893. [[CrossRef](#)]
38. Cao, L.; Zheng, J.; Cao, L.; Cui, J.; Xiao, Q. Evaluation of the impact of Shandong illegal vaccine sales incident on immunizations in China. *Hum. Vaccines Immunother.* **2018**, *14*, 1672–1678. [[CrossRef](#)]
39. Wang, L.D.; Lam, W.W.; Wu, J.T.; Liao, Q.; Fielding, R. Chinese immigrant parents' vaccination decision making for children: A qualitative analysis. *BMC Public Health* **2014**, *14*, 133. [[CrossRef](#)]
40. Zhou, M.; Qu, S.; Zhao, L.; Kong, N.; Campy, K.S.; Wang, S. Trust collapse caused by the Changsheng vaccine crisis in China. *Vaccine* **2019**, *37*, 3419. [[CrossRef](#)]
41. Hendriks, J.; Liang, Y.; Zeng, B. China's emerging vaccine industry. *Hum. Vaccines* **2010**, *6*, 602–607. [[CrossRef](#)]
42. Kaddar, M.; Milstien, J.; Schmitt, S. Impact of BRICS? investment in vaccine development on the global vaccine market. *Bull. World Health Organ.* **2014**, *92*, 436–446. [[CrossRef](#)]
43. Levin, C.E.; Sharma, M.; Olson, Z.; Verguet, S.; Shi, J.F.; Wang, S.M.; Kim, J.J. An extended cost-effectiveness analysis of publicly financed HPV vaccination to prevent cervical cancer in China. *Vaccine* **2015**, *33*, 2830–2841. [[CrossRef](#)]
44. Yiming, L.; Zaiping, J. Lessons from the Chinese defective vaccine case. *Lancet Infect. Dis.* **2019**, *19*, 245. [[CrossRef](#)]
45. Yuan, X. China's vaccine production scare. *Lancet* **2018**, *392*, 371. [[CrossRef](#)]
46. Clemmons, N.S.; Gastanaduy, P.A.; Fiebelkorn, A.P.; Redd, S.B.; Wallace, G.S. Measles—United States, 4 January–2 April 2015. *Mmwr. Morb. Mortal. Wkly. Rep.* **2015**, *64*, 373.

47. Majumder, M.S.; Cohn, E.L.; Mekaru, S.R.; Huston, J.E.; Brownstein, J.S. Substandard vaccination compliance and the 2015 measles outbreak. *JAMA Pediatrics* **2015**, *169*, 494–495. [CrossRef]
48. Olive, J.K.; Hotez, P.J.; Damania, A.; Nolan, M.S. The state of the antivaccine movement in the United States: A focused examination of nonmedical exemptions in states and counties. *PLoS Med.* **2018**, *15*, 1–10.
49. Lambert, P.H.; Siegrist, C.A. Science, medicine, and the future: Vaccines and vaccination. *Br. Med. J.* **1997**, *315*, 1595–1598. [CrossRef]
50. Leo, O.; Cunningham, A.; Stern, P.L. Vaccine immunology. *Perspect. Vaccinol.* **2011**, *1*, 25–59. [CrossRef]
51. Siegrist, C.A. Vaccine immunology. *Vaccines* **2008**, *5*, 17–36.
52. Bart, K.J.; Orenstein, W.A.; Hinman, A.R. The current status of immunization principles: Recommendations for use and adverse reactions. *J. Allergy Clin. Immunol.* **1987**, *79*, 296–315. [CrossRef]
53. Edsall, G. Principles of active immunization. *Annu. Rev. Med.* **1966**, *17*, 39. [CrossRef]
54. Grabenstein, J.D.; Nevin, R.L. Mass immunization programs: Principles and standards. In *Mass Vaccination: Global Aspects—Progress and Obstacles*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 31–51.
55. Biss, E. Sentimental Medicine: Why We Still Fear Vaccines. *Harper's Magazine*. January 2013. Available online: <https://harpers.org/archive/2013/01/sentimental-medicine/> (accessed on 20 January 2020).
56. Gidengil, C.; Chen, C.; Parker, A.M.; Nowak, S.; Matthews, L. Beliefs around childhood vaccines in the United States: A systematic review. *Vaccine* **2019**, *37*, 6793–6802. [CrossRef]
57. Sobo, E.J. What is herd immunity, and how does it relate to pediatric vaccination uptake? US parent perspectives. *Soc. Sci. Med.* **2016**, *165*, 187–195. [CrossRef]
58. Bechtol, Z. Launching a community-wide flu vaccination plan. *Fam. Pract. Manag.* **2008**, *15*, 19. [PubMed]
59. Ruderfer, D.; Krilov, L.R. Vaccine-preventable outbreaks: Still with us after all these years. *Pediatric Ann.* **2015**, *44*, e76–e81. [CrossRef] [PubMed]
60. Dube, E.; Vivion, M.; MacDonald, N.E. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: Influence, impact and implications. *Expert Rev. Vaccines* **2015**, *14*, 99–117. [CrossRef] [PubMed]
61. Jacobson, R.M.; Targonski, P.V.; Poland, G.A. A taxonomy of reasoning flaws in the anti-vaccine movement. *Vaccine* **2007**, *25*, 3146–3152. [CrossRef]
62. Poland, G.A.; Jacobson, R.M. Understanding those who do not understand: A brief review of the anti-vaccine movement. *Vaccine* **2001**, *19*, 2440–2445. [CrossRef]
63. Calvert, N.; Cutts, F.; Miller, E.; Brown, D.; Munro, J. Measles in secondary school children: Implications for vaccination policy. *Commun. Dis. Rep. Cdr Rev.* **1990**, *4*, R70–R73.
64. Hodge, J.G., Jr.; Gostin, L.O. School vaccination requirements: Historical, social, and legal perspectives. *Ky. Law J.* **2001**, *90*, 831.
65. Hoffman, J. How anti-vaccine sentiment took hold in U.S. *The New York Times*, 24 September 2019; A1.
66. Coleman-Brueckheimer, K.; Dein, S. Health care behaviors and beliefs in Hasidic Jewish populations: A systematic review of the literature. *J. Relig. Health* **2011**, *50*, 422–436. [CrossRef]
67. Schmidt, K. Measles and Vaccination: A Resurrected Disease, A Conflicted Response. *J. Christ. Nurs.* **2019**, *36*, 214–221. [CrossRef]
68. Tanne, J.H. US county bars unvaccinated children from public spaces amid measles emergency. *BMJ Br. Med. J.* **2019**, 364. [CrossRef] [PubMed]
69. Vielkind, J. Vaccination foes move to stop law. *The Wall Street Journal*, 15 August 2019; A8A.
70. Otterman, S. Thousands of anti-vaccine parents face an ultimatum. *The New York Times*, 4 September 2019; A19.
71. Pluviano, S.; Watt, C.; Della Sala, S. Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLoS ONE* **2017**, *12*, 1–15. [CrossRef] [PubMed]
72. Reuters. US Measles Outbreak Now Worst since 1994 after 60 New Cases Reported. *The Guardian*. 27 May 2019. Available online: <https://www.theguardian.com/us-news/2019/may/27/us-measles-outbreak-worst-since-1994> (accessed on 10 January 2020).
73. Eggertson, L. Lancet retracts 12-year-old article linking autism to MMR vaccines. *Can. Med. Assoc. J.* **2010**, *182*, E199. [CrossRef] [PubMed]
74. Flaherty, D.K. The vaccine-autism connection: A public health crisis caused by unethical medical practices and fraudulent science. *Ann. Pharmacother.* **2011**, *45*, 1302–1304. [CrossRef] [PubMed]
75. Godlee, F.; Smith, J.; Marcovitch, H. Wakefield's article linking MMR vaccine and autism was fraudulent. *Br. Med. J.* **2011**. [CrossRef]

76. Wakefield, A.J.; Murch, S.H.; Anthony, A.; Linnell, J.; Casson, D.M.; Malik, M.; Valentine, A. Retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* **1998**, *351*, 637–641. [CrossRef]
77. Deer, B. Wakefield's "autistic enterocolitis" under the microscope. *Br. Med. J.* **2010**, *340*, c1127. [CrossRef]
78. Deer, B. How the case against the MMR vaccine was fixed. *Br. Med. J.* **2011**, *342*, c5347. [CrossRef]
79. Deer, B. How the vaccine crisis was meant to make money. *Br. Med. J.* **2011**, *342*, c5258. [CrossRef]
80. Deer, B. The Lancet's two days to bury bad news. *Br. Med. J.* **2011**, *342*, c7001. [CrossRef]
81. Wakefield, A.J.; Murch, S.H.; Anthony, A. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children (Retraction of 351, 637, 1998). *Lancet* **2010**, *375*, 445.
82. Horton, R. A statement by the editors of The Lancet. *Lancet* **2004**, *363*, 820–821. [CrossRef]
83. Murch, S.H.; Anthony, A.; Casson, D.H.; Malik, M.; Berelowitz, M.; Dhillon, A.P.; Walker-Smith, J.A. Retraction of an interpretation. *Lancet* **2004**, *363*, 750. [CrossRef]
84. Kylstra, C. Celebrities and 'vaccine hesitancy'. *The New York Times*, 24 August 2019; A23.
85. Kata, A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine* **2010**, *28*, 1709–1716. [CrossRef]
86. Tafuri, S.; Gallone, M.S.; Cappelli, M.G.; Martinelli, D.; Prato, R.; Germinario, C. Addressing the anti-vaccination movement and the role of HCWs. *Vaccine* **2014**, *32*, 4860–4865. [CrossRef]
87. Myers, M.; Pineda, D. Misinformation about vaccines. In *Vaccines for Biodefense and Emerging and Neglected Diseases*; Barrett, A.D.T., Stanberry, L.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; pp. 255–270.
88. Zoon, K.C. *Science and the Regulation of Biological Products*; U.S. Food and Drug Administration: Washington, DC, USA, 2008. Available online: <https://www.fda.gov/about-fda/histories-product-regulation/science-and-regulation-biological-products> (accessed on 15 January 2020).
89. Jansen, V.A.; Stollenwerk, N.; Jensen, H.J.; Ramsay, M.E.; Edmunds, W.J.; Rhodes, C.J. Measles outbreaks in a population with declining vaccine uptake. *Science* **2003**, *301*, 804. [CrossRef]
90. Lernout, T.; Kissling, E.; Hutse, V.; Top, G. Clusters of measles cases in Jewish orthodox communities in Antwerp, epidemiologically linked to the United Kingdom: A preliminary report. *Wkly. Releases* **2007**, *12*, 3308. [CrossRef]
91. Richard, J.L.; Spicher, V.M. Ongoing measles outbreak in Switzerland: Results from November 2006 to July 2007. *Wkly. Releases* **2007**, *12*, 3241. [CrossRef]
92. Wichmann, O.; Hellenbrand, W.; Sagebiel, D.; Santibanez, S.; Ahlemeyer, G.; Vogt, G.; van Treeck, U. Large measles outbreak at a German public school, 2006. *Pediatric Infect. Dis. J.* **2007**, *26*, 782–786. [CrossRef] [PubMed]
93. Gust, D.A.; Strine, T.W.; Maurice, E.; Smith, P.; Yusuf, H.; Wilkinson, M.; Schwartz, B. Under immunization among children: Effects of vaccine safety concerns on immunization status. *Pediatrics* **2004**, *114*, e16–e22. [CrossRef]
94. Gust, D.A.; Kennedy, A.; Shui, I.; Smith, P.J.; Nowak, G.; Pickering, L.K. Parent attitudes toward immunizations and healthcare providers: The role of information. *Am. J. Prev. Med.* **2005**, *29*, 105–112. [CrossRef] [PubMed]
95. Shelby, A.; Ernst, K. Story and science: How providers and parents can utilize storytelling to combat anti-vaccine misinformation. *Hum. Vaccines Immunother.* **2013**, *9*, 1795–1801. [CrossRef] [PubMed]
96. Mayer, M.; Till, J.E. The Internet: A modern Pandora's box? *Qual. Life Res.* **1996**, *5*, 568–571. [CrossRef] [PubMed]
97. Bean, S.J. Emerging and continuing trends in vaccine opposition website content. *Vaccine* **1996**, *29*, 1874–1880. [CrossRef]
98. 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. [CrossRef]
99. Dunn, A.G.; Leask, J.; Zhou, X.; Mandl, K.D.; Coiera, E. 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. [CrossRef]
100. Christakis, N.A.; Fowler, J.H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **2007**, *357*, 370–379. [CrossRef]
101. Christakis, N.A.; Fowler, J.H. The collective dynamics of smoking in a large social network. *N. Engl. J. Med.* **2008**, *358*, 2249–2258. [CrossRef]

102. Greenwood, S.; Perrin, A.; Duggan, M. *Social Media Update 2016*; Pew Research Center: Washington, DC, USA, 2016; Available online: www.pewinternet.org/2016/11/11/social-media-update-2016/ (accessed on 20 January 2020).
103. Centola, D. The spread of behavior in an online social network experiment. *Science* **2010**, *329*, 1194–1197. [[CrossRef](#)] [[PubMed](#)]
104. Pang, B.; Lee, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; p. 271.
105. Turney, P.D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 417–424.
106. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 168–177.
107. Kim, S.M.; Hovy, E. Determining the sentiment of opinions. In Proceedings of the 20th International Conference on Computational Linguistics, Stroudsburg, PA, USA, 23–27 August 2004; p. 1367.
108. Wilson, T.; Hoffmann, P.; Somasundaran, S.; Kessler, J.; Wiebe, J.; Choi, Y.; Patwardhan, S. OpinionFinder: A system for subjectivity analysis. In Proceedings of the HLT/EMNLP 2005 Interactive Demonstrations, Vancouver, BC, Canada, 7 October 2005; pp. 34–35.
109. Agarwal, A.; Biadys, F.; Mckeown, K.R. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Athens, Greece, 30 March–2 April 2009; pp. 24–32.
110. Du, J.; Xu, J.; Song, H.Y.; Tao, C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Med. Inf. Decis. Mak.* **2017**, *17*, 69. [[CrossRef](#)] [[PubMed](#)]
111. Raghupathi, V.; Zhou, Y.; Raghupathi, W. Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases. *IEEE Access.* **2018**, *6*, 41518–41528. [[CrossRef](#)]
112. Raghupathi, V.; Zhou, Y.; Raghupathi, W. Exploring Big Data Analytic Approach to Social Media Cancer Blog Analysis. *Int. J. Healthc. Inf. Syst. Inform.* **2019**, *14*, 1–20. [[CrossRef](#)]
113. Glanz, J.M.; Wagner, N.M.; Narwaney, K.J.; Kraus, C.R.; Shoup, J.A.; Xu, S.; Daley, M.F. Web-based social media intervention to increase vaccine acceptance: A randomized controlled trial. *Pediatrics* **2017**, *140*, e20171117. [[CrossRef](#)]
114. Gu, Z.; Badger, P.; Su, J.; Zhang, E.; Li, X.; Zhang, L. A vaccine crisis in the era of social media. *Natl. Sci. Rev.* **2018**, *5*, 8–10. [[CrossRef](#)]
115. Schiff, A. The Letter to Amazon CEO Regarding Anti-Vaccine Misinformation. Available online: <https://schiff.house.gov/news/press-releases/schiff-sends-letter-to-amazon-ceo-regarding-anti-vaccine-misinformation> (accessed on 12 January 2020).
116. Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Quattrociocchi, W. The spreading of misinformation online. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 554–559. [[CrossRef](#)]
117. Dredze, M.; Broniatowski, D.A.; Hilyard, K.M. Zika vaccine misconception: A social media analysis. *Vaccine* **2016**, *34*, 3441. [[CrossRef](#)]
118. Dredze, M.; Broniatowski, D.A.; Smith, M.C.; Hilyard, K.M. Understanding vaccine refusal: Why we need social media now. *Am. J. Prev. Med.* **2016**, *50*, 550–552. [[CrossRef](#)]
119. Kang, G.J.; Ewing-Nelson, S.R.; Mackey, L.; Schlitt, J.T.; Marathe, A.; Abbas, K.M.; Swarap, S. Semantic network analysis of vaccine sentiment in online social media. *Vaccine* **2017**, *35*, 3621–3638. [[CrossRef](#)]
120. Raghunathan, R.; Trope, Y. Walking the tightrope between feeling good and being accurate: Mood as a resource in processing persuasive messages. *J. Personal. Soc. Psychol.* **2017**, *83*, 510–525. [[CrossRef](#)]
121. Xu, Z.; Guo, H. Using Text Mining to Compare Online Pro- and Anti-Vaccine Headlines: Word Usage, Sentiments, and Online Popularity. *Commun. Stud.* **2017**, *69*, 103–122. [[CrossRef](#)]

