

# Usage of Scientific References in MMR Vaccination Debates on Twitter

Aseel Addawood

*Illinois Informatics Institute & The iSchool, University of Illinois at Urbana-Champaign, IL, USA  
Information Systems Department, Imam Mohammad bin Saud University, Riyadh, Saudi Arabia  
aaddaw2@illinois.edu*

**Abstract**— This research analyzes scientific information sharing behaviors on Twitter. Over an eleven-month period, we collected tweets related to the controversy over the supposed linkage between the MMR vaccine and autism. We examined the usage pattern of scientific information resources by both sides of the ongoing debate. Then, we explored how each side uses scientific evidence in the vaccine debate. To achieve this goal, we analyzed the usage of scientific and non-scientific URLs by both polarized opinions. A domain network, which connects domains shared by the same user, was generated based on the URLs "tweeted" by users engaging in the debate in order to understand the nature of different domains and how they relate to each other. Our results showed that people with anti-vaccine attitudes linked many times to the same URL while people with pro-vaccine attitudes linked to fewer overall sources but from a wider range of resources, and that they provided fewer total links compared to people with anti-vaccine attitudes. Moreover, our results showed that vocal journalists have a huge impact on users' opinions. This study has the potential to improve understanding about how health information is disseminated via social media by showing how scientific evidence is referenced in discussions about controversial health issues. Monitoring scientific evidence usage on social media can reveal concerns and misconceptions related to the usage of these types of evidence.

**Keywords**— *Scientific sources; MMR vaccine; social media.*

## I. INTRODUCTION

Social media has revolutionized how people disclose personal health concerns and discuss public health issues. Social media provide unique platforms without time and location constraints for sharing health-related information [1, 2]. Social media have been found to be important tools for facilitating discussions on health information, especially in health crisis situations [3, 4] in which users share insights, opinions, and apprehensions while disseminating interpretations of health events outside of a public health context [4, 5].

When online users discuss topics on Twitter, they often include evidence to support their claims, including links to online sources, such as newspapers or blogs [6]. However, these sources may include unverified or even false information, which may amplify the perceived risks of these health issues [7, 8]. From an audience perspective, online health information offers a quick and useful reference, but its accuracy and credibility often falls into question [9].

Users of social media generally regard scientific sources, such as journal articles, to be credible. In this study, scientific sources are defined as sources that link to scholarly articles. Public opinion surveys from Europe and the US show that scientific institutions are trusted and are generally

considered to be more credible than non-scientific sources [10]. However, it is not clear from the previous literature how scientific evidence is deployed in discussions among Twitter users regarding health information. The problem of scientific research use in online, socially mediated discussions on health information is complicated by the controversies that surround certain health issues. These controversies can arise even when there is little to no credible evidence to support them. One significant controversy is the supposed relationship between the Measles, Mumps, and Rubella (MMR) vaccinations and autism.

During the 2014 holiday season, an outbreak of measles originated at the Disney theme parks in California. The outbreak generated extensive public discussion on some parents' resistance to childhood vaccinations. One reason for this outbreak is that, for some parents, concerns about the potential side effects of vaccines have overtaken concerns about the dangers of potentially deadly, vaccine-preventable diseases. "The Antivaccine Movement," a social movement composed of "antivaccine groups" and "antivaccine activists," is designated by scientists as the main cause of vaccine hesitancy or refusal [11-15].

Another challenge regarding health information controversies that presents itself in social media discourse is selective exposure to online information [16]. This phenomenon, which is due personalized web algorithms, happens when users find information that primarily supports their preconceptions and shields them from exposure to different ideas. Instances of selective exposure, including when like-minded people share their views with one another to reinforce their pre-conceived biases, are known as "echo chambers" [17]. Social media users' attitudes were confirmed through observations regarding the use of hashtags related to the vaccine controversy. One of the main uses of hashtags is to highlight users' sentiments towards the topic under discussion [18]. In order to conduct the observation, specific instances in the ongoing MMR vaccine debate from January 1, 2016 to November 28, 2016 served as a case study. These instances demonstrated how narrative elements are extracted for public debates regarding the vaccine issue.

While previous research has examined how scholars use social media, mainly Twitter, to request and offer assistance to others [19], to critique the work of other scholars [20], to contribute to conferences via hashtags [21, 22], to implement engaging pedagogies [23], and to share and comment upon preprint and published articles [24], no previous study found that scientific publications are referenced by online users to support their claims regarding vaccines. Moreover, previous research that has examined the credibility of information

shared via social media has not considered the use of scientific evidence by users of social media. Several studies have shown that the incorporation of URLs into social media posts is a means by which users attempt to confer credibility [25, 26]. However, no studies have been found that investigate the content or types of these URLs.

The goal of this study is to analyze scientific information sharing behaviors on Twitter regarding the controversy over the supposed linkage between MMR vaccine and autism. We examine the usage pattern of scientific information resources by both sides of the ongoing debate. Then, we explore how each side uses scientific evidence in the vaccine debate. To achieve this goal, we analyzed the usage of scientific and non-scientific URLs by both side of the debate. A domain network, which connects domains shared by the same user, was generated based on the URLs "tweeted" by other users in order to understand the nature of different domains and how they relate to each other. This study has the potential to improve understanding about the ways in which health information is disseminated via social media.

## II. BACKGROUND

### A. Social Media for Public Health

Studying the patterns and mechanisms of health-related communication via social media has the potential to give valuable insights into how health information shapes users' beliefs and attitudes. Salathé and Khandelwal [27] studied Twitter content to assess the levels of polarization between supporters and opponents of swine flu (H1N1) vaccination in the broader context of digital epidemiology [28]. They explored users' sentiments toward information shared via social media and users' following patterns. Their results show that people tend to follow other users who share the same sentiments about a topic. Radzikowski, Stefanidis, Jacobsen, Croitoru, Crooks and Delamater [29] analyzed Twitter narratives regarding MMR vaccination to identify key terms, connections among such terms, retweet patterns, the structure of the narrative, and connections to geographical space.

Social media data has also been used in outbreak detection. For example, Odlum and Yoon [30] studied the use of social media in the 2014 Ebola outbreak. They used a set of 42,236 tweets mentioning the word Ebola to assess the potential benefits of using social media as a real-time outbreak-tracking tool. Similarly, Lamos and Cristianini [31] and Culotta [32] correlated tweets mentioning influenza and related symptoms with historical data. Their results showed high correlations between Twitter statistics and Centers for Disease Control and Prevention (CDC) statistics in cases of influenza.

A closely related study was done by Love, Himelboim, Holton and Stewart [33], who conducted a content analysis of 2,580 reposted/shared vaccination Twitter posts to determine what vaccination information people share promote. Other researchers focused on one side of the debate: the anti-vaccine movement [34, 35]. This study's goal is to examine the use of scientific information sources from both sides of the debate with in-depth analysis of 6,112 tweets.

### B. Information Credibility in Social Media

Some studies on information credibility on Twitter focus on identifying sets of features that are indicative of credibility [25, 36, 37]. One of these features is the presence of links in the tweet text. Castillo et al. [25] use a complex set of features over tweets, re-tweets, the text of the posts, references to external sources, and users to predict the credibility of an event. Their results showed that having a URL tends to indicate that a tweet is credible. This was also confirmed by other studies [38, 39]. In this paper, we extend this research to investigating the use of a specific type of evidence, scientific evidence, in Twitter discussions about a controversial health issue.

## III. METHOD

### A. Data Collection

We collected a corpus that contained ground-truth, or gold standard, data, i.e., tweets that contain scientific versus non-scientific evidence on the topic of vaccines. Our corpus contained two main datasets. One dataset contained tweets that discussed the topic of MMR vaccines and their relation to autism, providing different types of supplementary evidence. The second dataset contained tweets that talked about vaccines and provided supplementary scientific evidence in the form of URLs linked to a scientific paper about vaccination. These two datasets are referred to as non-scientific and scientific, respectively.

For the non-scientific dataset, we collected data using Crimson Hexagon [40], a public social media analytics platform. We collected a sample of public posts made from January 1, 2016 to November 28, 2016. The sample only included tweets from accounts that set English as their language. The search criteria were: ("vaccinations" OR "vaccination" OR "vaccines" OR "vaccine" OR "measles-mumps-rubella" OR "MMR" OR "mmr" OR "#MMR") AND ("autism" OR "autistic disorder") AND NOT "RT:"

The total number of tweets retrieved was 45,320. To have more concise results, we removed all duplicate tweets (e.g. tweets repeated more than once in the dataset), which we believe is going to affect our final results. The total number of remaining tweets was 28,848. To collect more features related to each tweet, we ran these tweets' IDs through the Twitter API. Even though this process gave us more meta-data for each tweet, it reduced the usable number of tweets in the dataset to 27,816, since some tweets were deleted by the users or not found.

For the scientific dataset, we used PubMed to collect research articles related to vaccination and autism. PubMed is a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics. MEDLINE journals are selected by a technical advisory committee run by the U.S. National Institutes of Health [41]. The search query used in PubMed was: ("Vaccinations" OR "vaccination" OR "vaccines" OR "vaccine" OR "measles-mumps-rubella" OR "MMR" OR "mmr") AND ("autism" OR "autistic disorder"). This method resulted in a collection of 794 research papers.

For the next step, we chose Altmetric.com as the data source for social media and mainstream media counts, as it is the most comprehensive source of social media data associated with scientific papers [42]. Altmetric.com links an

identifier for each article that is provided by PubMed (i.e., its PMID). The Altmetric.com API then returns the social media reaction to a specific article that has been associated with a given PMID. Not all PMID queries produced results. For the 794 articles in our collection, Altmetric.com returned 346 results, i.e., the number of papers that had been referenced on social media. Since we chose Twitter as the social media platform for this study, we needed to further narrow these results to those that were Twitter-specific. Altmetric.com only provides the ID of a tweet, so once we eliminated non-Twitter results, we used the Twitter API to capture the actual tweet text by matching tweet IDs. The Twitter API returned 25,751 tweets. However, the Twitter API returned tweets in all languages, so languages other than English also had to be removed. To make our dataset consistent, all tweets that were retweets were removed. The final dataset contained 8,612 tweets, which we will refer to as the scientific dataset. This number is very small compared to the nonscientific dataset; this was expected since few online users refer to scientific sources.

We acknowledge the possibility of having scientific evidence in the 27,816 non-scientific dataset and vice versa. However, to mitigate this we crosschecked the scientific and non-scientific datasets to make sure no tweet appeared in both of them. We found 94 tweets that appeared in both datasets and removed them. The final combined datasets contained 36,428 tweets.

### B. Data Annotation

The next step was to annotate tweets for their stance towards vaccines. To accomplish this task, we utilized the hashtags present in each tweet. We followed previous work on hashtags use as indicators of users' common interests and opinions toward a health issue [43, 44]. As a first step, we identified all hashtags in the dataset; only 35% of the tweets contained hashtags (13,089 tweets). After that, two annotators hand-labeled all hashtags as either having a pro- or an anti-vaccine opinion. The inter coder reliability is 91.3%, with a 90.1% Cohen's Kappa. In total, there was 45 pro-vaccine hashtags and 94 anti-vaccine hashtags. The top hashtags identified in each category are shown in Table 1. To validate the selection of these hashtags a sample of 40 random tweets were chosen, two different annotators annotated the tweets for being either having a pro or an anti-vaccine opinion. Both annotators agreed on 39 cases out of 40 which also matched the hashtag opinion annotation.

### C. Hashtag Selection Discussion

To accurately identify a hashtag as either presenting an anti- or pro-vaccine attitude, we investigated the hashtag's usage on Twitter. For some tweets, it was easier to identify the opinion of the poster. For example, #vaccineswork clearly implies that the person believes in vaccines and their effectiveness. Similarly, #killingusslowly noticeably identifies that the person believes that vaccines can result in death. However, some hashtags are harder to identify since they require some understanding of the subject matter. For example, the hashtag #sb277 refers to the California Senate Bill 277, which is a law that removes personal belief exemptions to vaccination requirements for entry to schools in California, a state with relatively low vaccination levels in some schools [45]. This hashtag is used mostly by users who are against this bill and hold anti-vaccine attitudes.

TABLE 1. DISTRIBUTION FOR ANTI AND PRO-VACCINE ATTITUDES HASHTAGS

Anti-Vaccine	Count	Pro-Vaccine	Count
<b>vaxxed</b>	3408	vaccineswork	264
<b>cdcwhistleblower</b>	3132	vaccinesNOVA	98
<b>sb277</b>	322	vaxwithme	86
<b>cdcfraud</b>	195	vfvcall	61
<b>b1less</b>	182	whyivax	26
<b>vaccineinjury</b>	119	vaccinessavelives	17
<b>bigpharma</b>	100	antivaxxers	16
<b>coverup</b>	45	vaccinateyourkids	13
<b>learntherisk</b>	33	antivaxx	10
<b>vaccineskill</b>	30	teamvax	9

Another popular hashtag use is to reference people who hold the opposite opinion. For example, the hashtag #antivaxxers is used to refer to people who hold anti-vaccine attitudes. Similarly, #provaxxer is used to refer to people who hold pro-vaccine beliefs. Our assumption was people who hold anti-vaccine beliefs do not identify themselves with the #antivaxxers hashtag, and the same applies for people with pro-vaccine attitudes. This assumption was validated after closely reading a sample of tweets that use these hashtags. This investigation confirmed that users use these hashtags to refer to people who have the opposite belief than the one they have.

Some hashtags may be connected to the issue of vaccines but only provide information with no clear opinion on the issue, as in the hashtag #vaxxfacts. This hashtag clearly presents facts about vaccines without taking a side, but even it can be used by either side to claim that they are presenting facts. Moreover, some of the hashtags in the dataset did not demonstrate any clear opinion toward the issue. For example, the most used hashtag was #autism, which had 5447 occurrences. However, this hashtag does not show an opinion toward the issue. Other hashtags were out of the scope of the issue, such as #jewish, #dating and #sports.

## IV. FINDINGS

### A. Vaccine Attitudes

After identifying hashtags representing online users' vaccine attitudes on Twitter, we investigated the distribution of these attitudes in the data. To do so, we applied the selected hashtags to the data to identify tweets that signify pro- or anti- vaccine attitudes. We identified 6,112 tweets as having an opinion: 430 tweets with a pro-vaccine opinion and 5,682 tweets with an anti-vaccine opinion. We also found 215 tweets containing both anti- and pro-vaccine hashtags, which were removed from further analysis. These results show that there is a much higher number of tweets discussing anti-vaccination than pro-vaccination attitudes. This may indicate that people who hold anti-vaccine attitudes utilize Twitter as a venue for disseminating their opinions more than people who hold pro-vaccination beliefs. This may happen because social media is intensifying the reach and power of anti-vaccination messages, which may lead to negative reactions to vaccines being increasingly shared across online platforms [46].

## B. Usage of Scientific References

Another goal was to identify the use of scientific references in the discussion of vaccines on social media. To accomplish this, the number of pro- and anti-vaccine tweets with references to scientific and nonscientific evidence was recorded. Table 2 shows the distribution of vaccine attitudes and the use of scientific references. The ratio of pro-vaccine tweets containing links to non-scientific evidence compared to scientific evidence is 1:2.09, while the ratio of anti-vaccine tweets that contain a links to non-scientific evidence compared to scientific evidence is 1:5.01. These results show that people with both attitudes reference more non-scientific evidence compared to scientific evidence.

## C. Scientific Reference Analysis

To better understand the usage of scientific references in discussions of vaccines via social media, we did a thorough analysis of the URLs shared. Since many top domains are shortened URLs (e.g., bit.ly), we expanded them and extracted domain names. All tweets in our scientific dataset contained URLs. However, only 80% of the non-scientific dataset contained URLs. Table 3 shows the distribution of vaccine attitudes and the inclusion of URLs. The ratio of anti-vaccine tweets that contain URLs compared to tweets that do not is 5.31:1, while the ratio of pro-vaccine tweets with URLs compared with non-URLs is 3.53:1. This shows that users with anti-vaccine attitudes refer to external sources more often.

Online users can share more than one URL in their tweets. In our dataset, we found that users shared up to five URLs in their tweets. People with pro-vaccine attitudes shared up to two URLs, while people holding anti-vaccination views shared more. This result may indicate that people with anti-vaccine attitudes are trying to strengthen their arguments by sharing more links to external references.

At first, we investigated the top 15 URLs in pro-vaccine discussions. For users sharing non-scientific references, we found that these discussions mostly contained evidence showing that there is no link between the MMR vaccine and autism. This evidence mostly came from blogs or news websites such as npr.com. There were three references in the list that showed evidence of a positive linkage between the MMR vaccine and autism; however, these websites seemed untrustworthy based on a quick online search, such as "naturalnews.com" and "vaccines.news". Social media references, such as Twitter and YouTube, also appeared.

For users sharing scientific references, the top 15 URLs shared in pro-vaccine discussions show that this group shares references containing scientific evidence against the supposed link between the MMR vaccine and autism. The most shared URL linked to a paper published in the Vaccine journal with the title, "Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies," which confirmed that there is no link between MMR vaccine and autism.

People who are pro-vaccination frequently share scientific references citing articles published by Brian Deer,<sup>1</sup> a journalist. Deer did a series between 2004 and 2010 investigating the concerns over the MMR vaccine that arose after the 1998 publication of a research paper in the medical

TABLE 2. THE DISTRIBUTION OF VACCINE ATTITUDES AND THE USAGE OF SCIENTIFIC REFERENCES

	Pro-vaccine	Anti-vaccine
Scientific	139 (2.3%)	945 (15.5%)
Non-Scientific	291 (4.8%)	4,737 (77.5%)

TABLE 3. THE DISTRIBUTION OF VACCINE ATTITUDES AND THE INCLUSION OF URLs

	Pro-vaccine	Anti-vaccine
Contains a URL (one or more)	335 (5.5%)	4,782 (78.2%)
Does not contain a URL	95 (1.6%)	900 (14.7%)

journal, The Lancet, written by Andrew Wakefield and his colleagues [47], which was later retracted because of invalid research results. Two of Deer's articles appeared in our list: "How the case against the MMR vaccine was fixed" [48] and "How the vaccine crisis was meant to make money" [49]. In these articles, Deer shows how the results of this research were fraudulent.

The top 15 URLs shared in anti-vaccine discussions all claimed to show a link between the MMR vaccine and autism. We found that the domain "truthinmedia.com" was referenced the most in the list. We found that the website was no longer live. This website/project belongs to Ben Swann,<sup>2</sup> a journalist. He is best known for his investigation of the linkage between MMR vaccines and autism. In his fact-checking series, he argued that there is a link between MMR vaccines and autism. That same domain appeared in our top shared URLs four times. In our list, we also noticed many references to Ben Swann's Facebook page or tweets. All of these materials have been deleted. Other websites that appeared in the top shared URLs were "vaxxed.com" and "vaxxedthemovie.com," which are a movement and a movie with a conspiracy theory orientation that investigate the CDC's supposed destruction of a study linking autism to the MMR vaccine. The movie was directed by Andrew Wakefield, whose medical license was revoked after his paper was retracted.

We also found that these references linked to papers that claim to show a connection between MMR vaccines and autism. The most highly cited source was by Brian Hooker, which was titled "Measles-mumps-rubella vaccination timing and autism among young African American boys: a reanalysis of CDC data," which provides seemingly strong evidence of the linkage between MMR vaccine and autism. This paper was published in Translational Neurodegeneration journal in 2014 but was later retracted. Another highly referenced paper in our list was "Hepatitis B vaccination of male neonates and autism diagnosis, NHIS 1997-2002," [50] which showed evidence that Hepatitis B vaccination causes autism.

There were two articles that were shared in both the anti and pro-vaccine conversations. The first article was published in the journal of Immunologic Research in 2013 with the title "Aluminum in the central nervous system (CNS): toxicity in humans and animals, vaccine adjuvants, and autoimmunity" [51]. This paper linked aluminum used in vaccines to autism; both pro- and anti-vaccine posters

<sup>1</sup> <http://briandeer.com/>

<sup>2</sup> [https://en.wikipedia.org/wiki/Ben\\_Swann](https://en.wikipedia.org/wiki/Ben_Swann)

linked to this article. This finding may indicate that both groups of people have concerns regarding the content of vaccines and what they could do to young children, even though they have different attitudes toward the issue. The second article was published in the American Academy of Pediatrics in 2014 with the title “*Safety of Vaccines Used for Routine Immunization of US Children: A Systematic Review*” [52]. This paper had two main conclusions: “There is strong evidence that MMR vaccine is not associated with autism” and “We found evidence that some vaccines are associated with serious Adverse Events” (p. 334). While this paper confirms that there is strong evidence that MMR vaccines do not cause autism, it shows that in rare occasions rotavirus vaccines may be associated with intussusception, a different medical disorder. The citation of this paper by posters from both sides may indicate that people did not fully comprehend the results of the paper. All of these results show that on both sides there are influential people who people trust and reference as evidence for their beliefs.

When sharing non-scientific evidence, people mostly share links from social media websites such as Facebook, Twitter and YouTube. Sources such as news sites and personal blog posts are the next most commonly shared links. This may indicate that users on Twitter share other opinions manifested in tweets to support their own attitudes. Another observation was that the pro-vaccine community shared more diverse sources on Twitter than the anti-vaccine users. In the list of the top 15 URLs shared by the pro-vaccine community, all source domains were unique with no duplicates. The anti-vaccine top-shared URLs contained four pointers to the truthinmedia.com website and two references to Ben Swann, Truth in Media creator, including his social media webpages. Moreover, there were six different references to the movie *Vaxxed: From Cover-Up to Catastroph*, such as either a straight link to the website, a link to a website for a theatre for movie or a link to Periscope,<sup>3</sup> which shared a showing of the movie. This may

indicate that the anti-vaccine community has few sources to support their opinions.

The same results occurred in the sharing of scientific evidence on both sides of the controversy. Users holding anti-vaccine attitudes referenced the retracted paper by Brian Hooker five times on the list of top 15 URLs. Also, they referenced the paper linking Hepatitis B to vaccines twice. This may indicate that users holding anti-vaccine attitudes have little solid evidence to support their opinions. Meanwhile, users with pro-vaccine attitudes shared meta-analysis [53] and systemic review [52] papers, which are based on the evidence-based medicine pyramid considered to be as the highest and most trusted type of evidence [54]. These results indicate that it is mostly users holding anti-vaccine attitudes who rely on weak and redundant evidence compared with users holding pro-vaccine views.

#### D. Words Tell All: Unigram Analysis

Unigram analysis shows keywords for the overall narrative. Keywords reflect the topics that are considered relevant and important by the general public [29]. Given the design of the data collection process, all of the tweets in the data corpus for this analysis included one or more of the words in the search query used for the collection. All words that were used in the data collection were excluded from the analysis because their very high frequencies would make all other data points smaller. We also excluded stop words (i.e., articles, prepositions, and common verbs), as such words lack semantic significance.

In order to provide a general overview of the dominant narrative terms, Figure 1 shows a comparative word cloud visualization of the 100 most frequently encountered terms in each of the four data corpuses. The relative size of each word is proportional to its term frequency, where words in larger fonts are the ones more often encountered in the data corpus. Going from right to left, with the first being the non-scientific, pro-vaccine group’s top used words, we can see that the most common term encountered in these tweets is “cause.” Here is an example tweet from this dataset that shows how this word was used: @username I’ve said it before and I’ll say it again...VACCINES. DO. NOT. CAUSE. AUTISM. They do more good than harm. #vaccinateyourchildren

This reflects the fact that pro-vaccine advocates’ tweets demonstrate strong beliefs in a lack of linkage between vaccines and autism. The second most used word in this dataset is “link,” wherein users identify no link between vaccines and autism. The second word cloud set is for the dataset with anti-vaccine attitudes and the usage of non-scientific references. We can see that the most encountered term is “vaxxed,” followed up with “cdc” and “vaxxedthemovie.” These words are related to the prevalence of references to the movie in tweets. This result correlates with the previous results, which showed that users holding anti-vaccine attitudes share this movie widely using different outlets. The following tweet links to a podcast that discusses the importance of the movie and encourage others to watch it. #Vaxxed shows the fraud of #CDC and #BigPharma by revealing true link between #MMR #vaccine and #autism <link4>

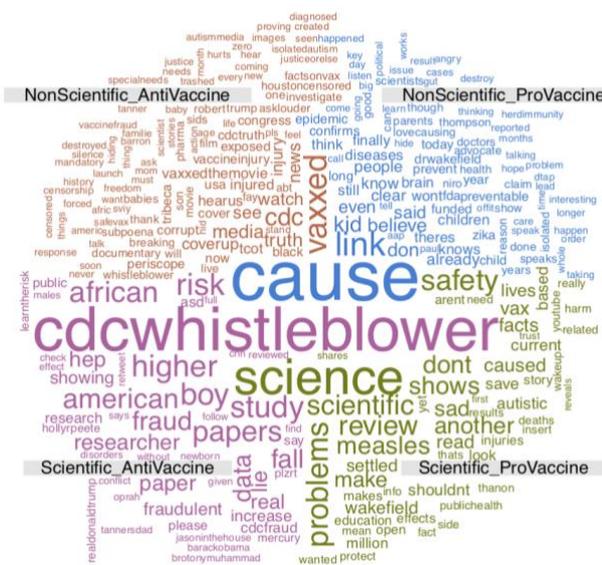


Figure 1. Comparative word cloud. Orange= Non Scientific\_AntiVaccine; Blue= Non Scientific\_ProVaccine; Pink= Scientific\_AntiVaccine; Green= Scientific\_ProVaccine.

<sup>3</sup> <https://www.pscp.tv/>

<sup>4</sup> <https://t.co/aOroYavCkr>

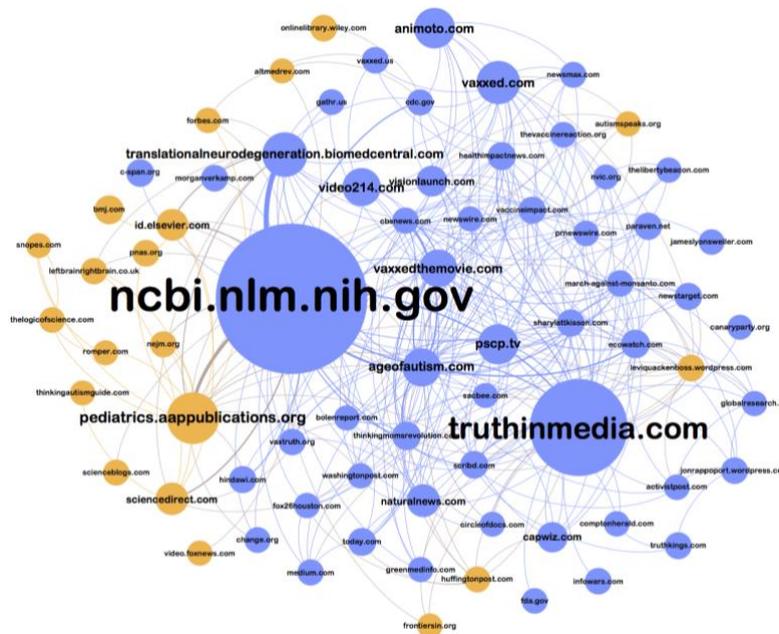


Figure 2. Domain Network Graph, colored by vaccine attitudes. Purple = Anti-vaccine; Orange = Pro-vaccine.

For the dataset discussing pro-vaccine attitudes and referencing scientific evidence, the most common terms encountered in these tweets are “science” and “scientific.”

The following tweets shows how these terms appeared in our dataset.

*@username science: vaccines aren't linked to autism. <link5> <link6>#vaccinesNOVA #hearthiswell*

*Reminder: An overwhelming body of scientific evidence shows that #VaccinesWork and don't cause #autism <link7>*

Finally, exploring the top word for anti-vaccine attitudes and the usage of scientific references, we can see the term “cdcwhistleblower,” which is a hashtag widely used by the antivaccination community in messages aligned with their views. This term did not originate from a formal organization, but instead is one that has emerged from an online advocacy community as a means to consolidate its views and promote its perspectives. Users using this hashtag claim that the CDC conceals evidence of the linkage between vaccine and autism, as shown in this tweet:

*@username Dr. Hooker's abstract removed with bogus reasoning. #CDCwhistleblower 's truth censored! <link8>*

In the same dataset of anti-vaccine attitudes and the usage of scientific references, we were puzzled by the word “fraud” in the top terms. When investigating the list of tweets including this term, we found that it was used to point to a study [55] showing that the child’s age when getting vaccines does not affect getting autism. The following tweet shows how the term was used.

*@TIME No integrity. The lead story should be this CDC study was called FRAUD by one of its authors. <link9> #CDCwhistleblower*

This tweet and some other similar ones claim that this study presented inaccurate results on the linkage between children ages when first administered with vaccines, and autism. Similar tweets claim the study was funded by the CDC to prove that there is no association between the required vaccines and autism. Taken together, these observations show how different opinion-holders express their attitudes toward the issue. These results show that users mostly discuss scientific studies to support their opinions. Others share these studies to reject them by accusing them of circulating fraudulent results.

### E. Domain Network

To understand how different internet domains are used as supports in online users’ discussions regarding the linkage between the MMR vaccine and autism, we created a network graph of domains connected through user activity, specifically the URLs shared in their tweets. Building on previous work on online rumors [56, 57] to create the graphs, we first identified every distinct domain that is linked to by a tweet in the set. 83.72% of the tweets contain one or more URLs (5,117 tweets), and together they reference 495 distinct domains. These domains became the initial nodes for the graph. We created the edges between the nodes by observing the tweet patterns of each user, connecting two nodes if the same user posted tweets referencing both domains [56].

Some domains were removed from the graphs for their high rates of connectivity to other sites as well as the different meanings encoded in those connections, i.e. they are used as tools, not for their content. These domains include social media services (e.g. Twitter.com, facebook.com, reddit.com) and all general link shortener services (e.g. bit.ly, t.co) which did not resolve to a URL. Finally, we trimmed the graph by removing domains that appeared fewer than five times in the set. In the graphs, nodes are sized proportionally to the total number of tweets that linked to the domain, and they are connected when an

<sup>5</sup> <http://www.sciencedirect.com/science/article/pii/S0264410X14006367>  
<sup>6</sup> <https://academic.oup.com/cid/article/48/4/456/284219/Vaccines-and-Autism-A-Tale-of-Shifting-Hypotheses>  
<sup>7</sup> <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0003140>  
<sup>8</sup> <https://translationalneurodegeneration.biomedcentral.com/articles/10.1186/2047-9158-3-16>

<sup>9</sup> <https://www.ncbi.nlm.nih.gov/pubmed/14754936>

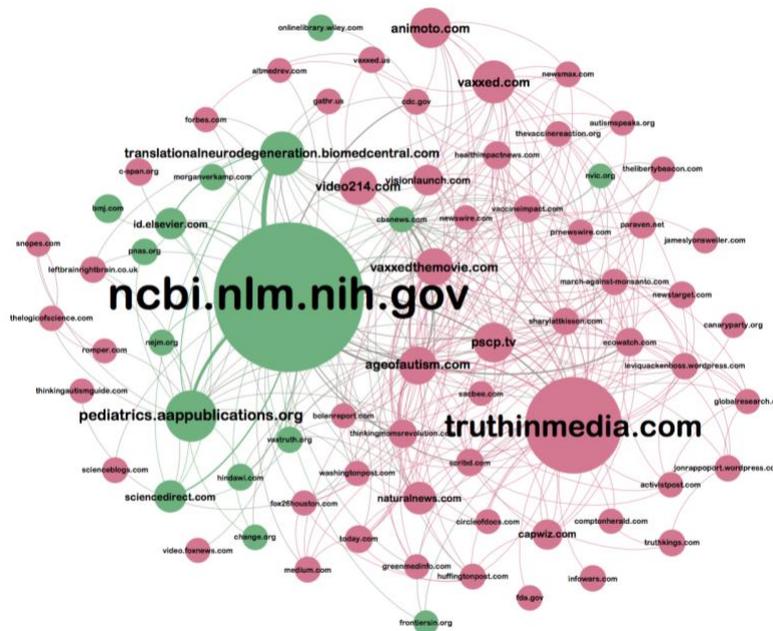


Figure 3. Domain Network Graph, colored by evidence type. Green = Scientific; Pink = Non-Scientific.

individual user wrote different tweets citing each domain. Furthermore, the strength of the edge grows proportionally to the number of users who shared tweets referencing both domains.

The resulting network graph represents how different domains are connected through the posting activity of Twitter users who contributed to the scientific evidence discourse surrounding the linkage of MMR vaccines to autism. We limited this analysis to the 74 nodes that are connected to the central graph. Figure 2 shows the domain network graph. In this graph, we distinguish domains by vaccine attitude, with anti-vaccine attitudes in Purple and pro-vaccine attitudes in Orange. To identify the attitude of each domain, we first identified all tweet attitudes where each domain was used. After that, we assigned each domain the attitude that was more strongly represented after normalization. 44 of 74 domains in our graph were classified as being used as a support for a tweet showing an anti-vaccine attitude, while 30 domains were classed as pro-vaccine. The network graph shows a tightly connected cluster of anti-vaccine domains, suggesting that many users cite multiple anti-vaccine sites as a support for their beliefs. Within that cluster, the three most-highly tweeted and most connected domains are ncbi.nlm.nih.gov, ageofautism.com, cbsnews.com, and truthinmedia.com.

Ncbi.nlm.nih.gov (the National Center for Biotechnology Information (NCBI)) is a branch of the National Institutes of Health. It houses a series of databases relevant to biotechnology and biomedicine. Online users use this domain as a reference for scientifically written and validated articles. However, some of these papers, which are frequently cited by the anti-vaccine community, have been retracted. Ageofautism.com is a site devoted to proving that autism is induced by the environment and that MMR vaccination is the main cause of autism. This domain is highly connected with domains that advocate for the same agenda, such as truthinmedia.com, vaxxedthemovie.com and vaxxed.com. CBSnews.com is a known source of daily news. This domain is the most highly connected domain (after

ncbi.nlm.nih.gov and ageofautism.com) that provides articles verifying and implying a causal link between vaccine and autism. Truthinmedia.com was the second most tweeted domain. Online users with anti-vaccine attitudes share different types of evidence. As an example, one user with apparently strong anti-vaccine attitudes used an article from cbsnews.com:

:( DOCTORS - FIRST! Do No HARM: UnSAFE Vax = Autism: COMPELLING 2011 Scientific Review!! CDC=UNTRUSTWORTHY!! #vaxxed <link10>

And another from ncbi.nlm.nih:

2009 PubMedC: Regressive Autism Due to Overuse of Vaccines? (7 YEARS AGO - 1 of SEVERAL STUDIES...) WTW? #Vaxxed <link11>

The use of diverse types of sources to support users' claims is intended to prove that their points of view are accurate and sound.

Figure 3 shows the domain network graph distinguished by the evidence type, with scientific evidence in Green and non-scientific evidence in Pink. To identify if a domain is classified as scientific or non-scientific, we first mapped the number of scientific and non-scientific tweets that used each domain and then assigned the class that was more strongly represented for the domain after normalizing based on class distribution. Only 15 out of the 74 domains were considered to provide scientific evidence; these domains were highly connected to each other and made their own cluster. Some domains were identified as scientific domains even though they did not actually represent a scientific source.

This may have happened because these websites were co-cited with scientific references, such as cbsnews.com, change.org, vaxtruth.org and morganverkamp.com. These domains include many references to scholarly articles, which may indicate that online users use non-scientific sources that

<sup>10</sup> <https://www.cbsnews.com/news/vaccines-and-autism-a-new-scientific-review/>

<sup>11</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3364648/>

cite scientific sources and regard them to be as sound as scientific references.

The domain of scientific evidence that was most heavily connected to non-scientific evidence domains was the National Center for Biotechnology Information (NCBI) website. This website, [cdc.gov](http://cdc.gov) and [fda.gov](http://fda.gov) were the only governmental websites used as references in our dataset, with the CDC and FDA referenced very minimally. This result is similar to previous research, which indicates that a small number of people have trust in government vaccine experts/officials [58]. The non-scientific evidence domain most connected to scientific evidence was [ageofautism.com](http://ageofautism.com) (described above). This domain is connected to 7 of the domains with scientific evidence. This may indicate that people consider this domain comparable with other higher credibility websites.

## V. DISCUSSION, CONCLUSION

In this study, our goal was to investigate the use of scientific and non-scientific evidence in social media when discussing a controversial health issue, such as the MMR vaccine debate. This study showcased emerging data analysis approaches. These approaches are inherently interdisciplinary, bringing together principles and practices from health informatics, data analytics, and network analysis. Our results show that online users with anti-vaccine attitudes share more content via Twitter than users with pro-vaccine attitudes, which correlate with previous research [33]. Moreover, they share more tweets linking external references and, specifically, non-scientific evidence. Furthermore, our results show that people with anti-vaccine attitudes share many sources but with low diversity, while people with pro-vaccine attitudes share a smaller but more diverse number of sources. Additionally, our results show that vocal journalists have a huge impact on users' opinions. Journalists often report on controversy by presenting claims both for and against an issue in a relatively 'balanced' fashion, which leads to more uncertainty on the part of their readers [59, 60].

The overall results of this study can help us make more accurate interpretations of people's attitudes and opinions regarding controversial health topics, such as the debate over vaccines. However, our work is limited in many ways. First, tweets gathered for the "non-scientific" dataset may contain references to scientific papers not found through PubMed. Second, even though previous work has indicated that hashtags can be an indicator of users' opinions [43, 44], some users use hashtags that indicate the opposite attitude to express the opinion that the other side of the debate is wrong or to voice sarcastic opinions regarding the other party. Moreover, given that we only examined one case, the vaccine debate, we are limited in the understanding and analysis of other sources of scientific information that users share online when discussing health issues. In future work, we would like to explore the strength of the attitudes held by each side of the debate and study if people with strong opinions differ in the usage of information sources from users with moderate or no opinions towards the debate. Eventually, we plan to expand our understanding of the use of scientific sources via social media by studying other health topics.

## REFERENCES

- [1] Lachlan, K.A., Spence, P.R. and Lin, X. 2014. Expressions of risk awareness and concern through twitter: On the utility of using the medium as an indication of audience needs. *Computers in Human Behavior*, 35554-559
- [2] Dunlap, J.C. and Lowenthal, P.R. 2009. Tweeting the night away: Using twitter to enhance social presence. *Journal of Information Systems Education*, 20,(2), 129
- [3] Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D. and Adini, B. 2014. Twitter in the cross fire—the use of social media in the westgate mall terror attack in kenya. *PLoS one*, 9,(8),
- [4] Sullivan, S.J., Schneiders, A.G., Cheang, C.-W., Kitto, E., Lee, H., Redhead, J., Ward, S., Ahmed, O.H. and McCrory, P.R. 2012. 'What's happening?' a content analysis of concussion-related traffic on twitter. *British Journal of Sports Medicine*, 46,(4), 258-263
- [5] Khan, A.S., Fleischauer, A., Casani, J. and Groseclose, S.L. 2010. The next public health revolution: Public health information fusion and social networks. *American Journal of Public Health*, 100,(7), 1237-1242
- [6] Addawood, A. and Bashir, M. 2016. "What is your evidence?" a study of controversial topics on social media. In *Proceedings of the 3rd Workshop on Argument Mining*. Association for Computational Linguistics
- [7] Fung, I.C.-H., Tse, Z.T.H., Cheung, C.-N., Miu, A.S. and Fu, K.-W. 2014. Ebola and the social media. *The Lancet*, 384,(9961), 2207
- [8] Kasperson, R.E., Renn, O., Slovic, P., Brown, H.S., Emel, J., Goble, R., Kasperson, J.X. and Ratick, S. 1988. The social amplification of risk: A conceptual framework. *Risk analysis*, 8,(2), 177-187
- [9] Metzger, M.J. and Flanagin, A.J. 2011. Using web 2.0 technologies to enhance evidence-based medical information. *Journal of health communication*, 16,(sup1), 45-58
- [10] Bucchi, M. and Trench, B. 2014. *Routledge handbook of public communication of science and technology* Routledge.
- [11] Ackermann, D., Chapman, S. and Leask, J. 2004. Media coverage of anthrax vaccination refusal by australian defence force personnel. *Vaccine*, 23,(3), 411-417
- [12] Betsch, C., Brewer, N.T., Brocard, P., Davies, P., Gaissmaier, W., Haase, N., Leask, J., Renkewitz, F., Renner, B. and Reyna, V.F. 2012. Opportunities and challenges of web 2.0 for vaccination decisions. *Vaccine*, 30,(25), 3727-3733
- [13] Betsch, C. 2011. Innovations in communication: The internet and the psychology of vaccination decisions. *Euro Surveill*, 16,(17), 1-6
- [14] Gangarosa, E.J., Galazka, A., Wolfe, C., Phillips, L., Miller, E., Chen, R. and Gangarosa, R. 1998. Impact of anti-vaccine movements on pertussis control: The untold story. *The Lancet*, 351,(9099), 356-361
- [15] Poland, G.A. and Jacobson, R.M. 2011. The age-old struggle against the antivaccinationists. *New England Journal of Medicine*, 364,(2), 97-99
- [16] Frey, D. 1986. Recent research on selective exposure to information. *Advances in experimental social psychology*, 1941-80
- [17] Sunstein, C.R. 2009. *Republic. Com 2.0* Princeton University Press.
- [18] Wang, X., Wei, F., Liu, X., Zhou, M. and Zhang, M. 2011. Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM
- [19] Veletsianos, G. 2012. Higher education scholars' participation and practices on twitter. *Journal of Computer Assisted Learning*, 28,(4), 336-349
- [20] Mandavilli, A. 2011. Trial by twitter. *Nature*, 469,(7330), 286
- [21] Li, J. and Greenhow, C. 2015. Scholars and social media: Tweeting in the conference backchannel for professional learning. *Educational Media International*, 52,(1), 1-14
- [22] Ross, C., Terras, M., Warwick, C. and Welsh, A. 2011. Enabled backchannel: Conference twitter use by digital humanists. *Journal of Documentation*, 67,(2), 214-237
- [23] Junco, R., Heiberger, G. and Loken, E. 2011. The effect of twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27,(2), 119-132
- [24] Eysenbach, G. 2011. Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, 13,(4),
- [25] Castillo, C., Mendoza, M. and Poblete, B. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM
- [26] Kinsella, S., Wang, M., Breslin, J. and Hayes, C. 2011. Improving categorisation in social media using hyperlinks to structured data sources. *The Semantic Web: Research and Applications*390-404

- [27] Salathé, M. and Khandelwal, S. 2011. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS computational biology*, 7,(10),
- [28] Salathé, M., Bengtsson, L., Bodnar, T.J., Brewer, D.D., Brownstein, J.S., Buckee, C., Campbell, E.M., Cattuto, C., Khandelwal, S. and Mabry, P.L. 2012. Digital epidemiology. *PLoS computational biology*, 8,(7),
- [29] Radzikowski, J., Stefanidis, A., Jacobsen, K.H., Croitoru, A., Crooks, A. and Delamater, P.L. 2016. The measles vaccination narrative in twitter: A quantitative analysis. *JMIR public health and surveillance*, 2,(1),
- [30] Odium, M. and Yoon, S. 2015. What can we learn about the ebola outbreak from tweets? *American journal of infection control*, 43,(6), 563-571
- [31] Lamos, V. and Cristianini, N. 2010. Tracking the flu pandemic by monitoring the social web. In *Proceedings of the 2nd International Workshop on Cognitive Information Processing (CIP)*. IEEE
- [32] Culotta, A. 2010. Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the First Workshop on Social Media Analytics*. ACM
- [33] Love, B., Himelboim, I., Holton, A. and Stewart, K. 2013. Twitter as a source of vaccination information: Content drivers and what they are saying. *American journal of infection control*, 41,(6), 568-570
- [34] 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,(25), 3778-3789
- [35] Tomeny, T.S., Vargo, C.J. and El-Toukhy, S. 2017. Geographic and demographic correlates of autism-related anti-vaccine beliefs on twitter, 2009-15. *Social science & medicine*, 191168-175
- [36] Gupta, A. and Kumaraguru, P. 2012. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*. ACM
- [37] Mendoza, M., Poblete, B. and Castillo, C. 2010. Twitter under crisis: Can we trust what we rt? In *Proceedings of the first workshop on social media analytics*. ACM
- [38] Morris, M.R., Counts, S., Roseway, A., Hoff, A. and Schwarz, J. 2012. Tweeting is believing?: Understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM
- [39] Sikdar, S.K., Kang, B., O'Donovan, J., Hollerer, T. and Adal, S. 2013. Cutting through the noise: Defining ground truth in information credibility on twitter. *Human*, 2,(3), 151-167
- [40] Etlinger, S. and Amand, W. 2012. Crimson hexagon [program documentation]. Retrieved September 15, 2016. From [http://www.crimsonhexagon.com/wp-content/uploads/2012/02/CrimsonHexagon\\_Altimeter\\_Webinar\\_111611.pdf](http://www.crimsonhexagon.com/wp-content/uploads/2012/02/CrimsonHexagon_Altimeter_Webinar_111611.pdf)
- [41] NLM. 2013. Medline fact sheet. Retrieved Oct 19 2017. From <http://www.nlm.nih.gov/pubs/>
- [42] Robinson-García, N., Torres-Salinas, D., Zahedi, Z. and Costas, R. 2014. New data, new possibilities: Exploring the insides of altmetric. *Com. El profesional de la información*, 23,(4), 1386-6710
- [43] Xu, W.W., Chiu, I.-H., Chen, Y. and Mukherjee, T. 2015. Twitter hashtags for health: Applying network and content analyses to understand the health knowledge sharing in a twitter-based community of practice. *Quality & Quantity*, 49,(4), 1361-1380
- [44] Fang, A., Ounis, I., Habel, P., Macdonald, C. and Limsopatham, N. 2015. Topic-centric classification of twitter user's political orientation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM
- [45] Aftab, A. 2015. California becomes a state divided after the controversial sb 277 vaccination bill is handed to governor brown. Retrieved October 14, 2017. From <http://www.independent.co.uk/news/world/americas/california-becomes-a-state-divided-after-the-controversial-sb-277-vaccination-bill-is-handed-to-10355888.html>
- [46] UNICEF. 2013. Tracking anti-vaccination sentiment in eastern european social media networks. *New York: UNICEF*
- [47] Wakefield, A.J., Murch, S.H., Anthony, A., Linnell, J., Casson, D., Malik, M., Berelowitz, M., Dhillon, A.P., Thomson, M.A. and Harvey, P. 1998. Retracted: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children
- [48] Deer, B. 2011. How the case against the mmr vaccine was fixed. *Bmj*, 342c5347
- [49] Deer, B. 2011. How the vaccine crisis was meant to make money. *Bmj*, 342c5258
- [50] Gallagher, C.M. and Goodman, M.S. 2010. Hepatitis b vaccination of male neonates and autism diagnosis, nhis 1997–2002. *Journal of Toxicology and Environmental Health, Part A*, 73,(24), 1665-1677
- [51] Shaw, C. and Tomljenovic, L. 2013. Aluminum in the central nervous system (cns): Toxicity in humans and animals, vaccine adjuvants, and autoimmunity. *Immunologic research*, 56,(2-3), 304-316
- [52] Maglione, M.A., Das, L., Raaen, L., Smith, A., Chari, R., Newberry, S., Shanman, R., Perry, T., Goetz, M.B. and Gidengil, C. 2014. Safety of vaccines used for routine immunization of us children: A systematic review. *Pediatrics*, 134,(2), 325-337
- [53] Taylor, L.E., Swerdfeger, A.L. and Eslick, G.D. 2014. Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine*, 32,(29), 3623-3629
- [54] Greenhalgh, T. 2014. *How to read a paper: The basics of evidence-based medicine* John Wiley & Sons.
- [55] DeStefano, F., Bhasin, T.K., Thompson, W.W., Yeargin-Allsopp, M. and Boyle, C. 2004. Age at first measles-mumps-rubella vaccination in children with autism and school-matched control subjects: A population-based study in metropolitan atlanta. *Pediatrics*, 113,(2), 259-266
- [56] Starbird, K. 2017. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. In *Proceedings of the International Conference on Web and Social Media*. Association for the Advancement of Artificial Intelligence
- [57] Maddock, J., Starbird, K., Al-Hassani, H.J., Sandoval, D.E., Orand, M. and Mason, R.M. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM
- [58] Freed, G.L., Clark, S.J., Butchart, A.T., Singer, D.C. and Davis, M.M. 2011. Sources and perceived credibility of vaccine-safety information for parents. *Pediatrics*, 127,(Supplement 1), S107-S112
- [59] Lewis, J. and Speers, T. 2003. Misleading media reporting? The mmr story. *Nature Reviews Immunology*, 3,(11), 913-918
- [60] Clarke, C.E. 2008. A question of balance: The autism-vaccine controversy in the british and american elite press. *Science Communication*, 30,(1), 77-107