Comparison of Voluntary versus Mandatory Vaccine Discussions in Online Health Communities: A Text Analytics Approach

Massara Alazazi
The University of Texas Rio Grande Valley

Bin Wang
The University of Texas Rio Grande Valley

Recommended Citation

Comparison of Voluntary versus Mandatory Vaccine Discussions in Online Health Communities: A Text Analytics Approach

Massara Alazazi
University of Texas Rio Grande Valley
massara.alazazi01@utrgv.edu

Bin Wang
University of Texas Rio Grande Valley
bin.wang@utrgv.edu

Abstract

Vaccines are vital health interventions. However, they are controversial and some people support them while others reject them. Social media discussion and big data are a rich source to understand people's insights about different vaccines and the related topics that concern most of them. This study aims to explore the online discussions about mandatory and voluntary vaccines using text analysis techniques. Reddit social platform is popular in online health discussion and thus data from Reddit is analyzed. The results show that different aspects are discussed for different types of vaccines. The discussion of mandatory vaccines is more interactive and is focused on the risks associated with them. Voluntary vaccines' discussion is focused on their effectiveness and whether to get them or not. The study have important implications for health agencies and researchers as well as for healthcare providers and caregivers.

1. Introduction

Vaccination is an essential part of individuals' health interventions. They are found to contribute to the reduction of mortality rates [1]. However, many individuals either refuse vaccinations or doubt their effectiveness and efficiency [2]. Statistics show that between 10 to 20% of parents refuse or postpone at least one vaccine for their children [3]. Because some vaccinations are controversial, many people turn to social media outlets to get information about vaccines [3, 4] or to influence others by disseminating their beliefs and theories [5]. Online blogs, microblogs, discussion boards, videos and their content can impact the decision of vaccination for many individuals [4]. Particularly, new parents could be affected by their social networks regarding their children's vaccination due to the lack of experience, hesitancy, and others [6].

Many debates are available on social media between supporters and rejecters of vaccines. It is vital to explore this content in order to understand the accuracy of information, the beliefs and attitudes of people, their concerns, and the topics that dominate the online discussion. Vaccines differ in their deliberative nature. Some vaccines have more agreement by individuals and science regarding their necessity and thus are mandatory. However, other vaccines such as the flu vaccine are more controversial. In addition, due to governmental restrictions and policies, some vaccines are mandatory and others are voluntary. When a vaccine is voluntary, people have less motivation to be vaccinated [7].

This study aims to compare and contrast online discussions about mandatory versus voluntary vaccines. Previous research has focused on studying one type of vaccines or studying vaccines in general. However, knowing people's insights and opinions about different types of vaccines help in shaping health policies and determining the need to convert some vaccines from voluntary to mandatory if necessary. In addition, our data analysis is based on a social online platform – Reddit – that is not adequately explored, unlike Facebook and Twitter. Exploring various online platforms' content can help in providing us with a better understanding of people's conversations and their associated feelings, attitudes, and topics of concern. Our study uses text analysis to analyze the terms that are correlated with each vaccine discussion, the affects and sentiment expressed, and topics discussed for each type of vaccine. The research question that this research addresses is: What are the differences between mandatory and voluntary vaccine online discussions in terms of the topics discussed, the sentiment expressed, and the terms associated with each of them?

Our study finds that flu vaccine online discussion has more emotional expressions compared with the MMR vaccine discussion. In addition, the flu vaccine discussion has some dominated topic such as the effectiveness of the vaccine and getting the vaccine annually. The topics that dominate the MMR vaccine discussion include the debate about the MMR vaccine

and the risks associated with it in addition to the need for evidence about these risks.

2. Literature Review

2.1 Online health Communities

Social media is leveraged by many patients, physicians, and caregivers to generate health related-content [8]. Online health communities are widely used due to the convenience, anonymity and support provided [9]. Many health issues are discussed on social media and their related content is studied including eating disorder [10], smoking [11], drugs [12], breast cancer [13], mental health [14], chronic diseases [15] and many others. Participation in health communities and forums is motivated by many factors including reciprocity, altruism, and homophily [9, 16]. Patients refuge to the online environment to fulfill their needs such as sharing their experience, managing their illness, learning from similar others, or getting others support, opinions, or advice [8, 17]. Social networking sites and blogs give users the ease of asking health-related questions, sharing their thoughts, watching educational videos related to their health condition, and reading peers' and physicians' posts and replies [8]. Communication with physicians on these social sites may lead to both better self-health management and better patient-physician relationship [15]. Even though online health communities and forums are frequented by many individuals, the credibility of health and medical information generated by users is a concern [12]. If patients take serious health-actions based on the information they receive from online content, the outcome may not be desirable if the information is inaccurate.

2.2 Mandatory vs Voluntary Vaccines

Vaccination is a vital tool to prevent many illnesses [18]. The occurrence of many diseases has declined after the implementation of their vaccines such as the measles, mumps, and rubella MMR. However, there is a wide spectrum of vaccines available. Some vaccines are mandatory and others are voluntary in the US. Some vaccines are required before a child enters school. Other vaccines are voluntary and depend on the individual or on the parents' preference. Voluntary vaccines decision depends on personal social norms [19], religious issues [20], beliefs, risk of the illness, or the vaccine effectiveness and side effects [7, 21]. Previous research proposed a risk assessment model to vaccination decision-making by comparing the payoff

and the risk. By modeling a vaccination game, the study found that social networks of individuals play a major role in the decision of vaccination [7]. Another study focused on the controversial debate about mandatory vs voluntary healthcare workers vaccination [20]. There is a need to protect them and reduce the spread of infectious diseases. On the other hand, health ethics and individuals' autonomy need to be considered. The study concludes that there is a moral need for vaccination of healthcare workers and enforcing vaccination would be more fruitful [20].

Online health communities play a key role in discussing vaccine controversies. A recent study explored their role in terms of knowledge delivery for pro and anti-vaccine movements [1]. The study found that both knowledge delivery practices online and offline impact each other while differ in their configurations. The study compared the format of the knowledge, the delivery of the knowledge, and the availability of the knowledge. Another study examined the influence of social media on individuals' decision of vaccination [22]. An online experimental design targeted parents of girls to test the impact of online information about HPV vaccine on them from the lens of the health belief model. The study used a novel exploratory network analysis and found that perceived efficacy is influential while cues to action is not. Moreover, a previous study applied the semantic network analysis on articles shared on Twitter about vaccines and analyzed the resulting networks. The study also identified positive, negative, and neutral sentiment in the text of the shared articles [23]. In their study the researchers used manual sentiment coding of the articles.

2.3 Text Mining

Big data and the enormous amount of unstructured texts available on social media encouraged researchers to analyze these texts, extract knowledge and patterns, and infer results by applying techniques such as classification, clustering, machine learning, and social network analysis [24]. Previous research focused on Twitter and Facebook extracted-text and how different text analytics techniques are used to provide textual orders and thus provide key themes in data [25]. Sentiment analysis is used to extract sentiment-related words, emotions and find their polarity; either positive or negative. Previous research used sentiment analysis using both supervised and unsupervised machine learning techniques to analyze tweets from Twitter about a particular product or movie [26]. The analysis is performed based on the sentence-level sentiment identification and used SentiWordNet software for sentiment analysis. The study found that tweeters use

hashtags to express their emotions. Another study used emotional text mining to analyze customer profiling for brand management of a sportswear from Twitter. Cluster analysis with a bisecting k-means algorithm and correspondence analysis are used to analyze the data. The study identified Twitter users' symbolic categories, measured their sentiment, and found their representations of the sportswear brand [26]. In addition, another study analyzed news channels online textual data from Facebook posts using the RapidMiner tool. The results of the study indicate that the most covered topics on these news channels pages were US election news. CNN had the most shared posts about the topic [27]. Another research paper proposed a deception detection mechanism for crowdfunding projects by considering both static project information and dynamic communication between funders and fundraiser for classification [28]. Cues extracted from the text such as content-based cues and linguistic cues to detect fraudulent crowdfunding projects are used to detect fraud using machine learning algorithms.

Text mining for health-related text on social media has also been applied in previous studies [29]. One study analyzed flu tweets on Twitter using network analysis. The study found that effective information about flu are generated by accounts found in the important Twitter accounts and these tweets would stay active a longer period compared with other individual accounts' tweets [30]. Another study investigated opinions of Twitter users about influenza vaccine using natural language classifiers to identify vaccine attitudes and behaviors [31]. Using MedHelp.org health platform, a study performed text mining to identify main stakeholders participating in lung cancer, breast cancer and diabetes forums. Patients and caregivers were the main population of participants while specialists formed the minority of them [32]. The discussed topics included symptoms, drugs, procedures, examinations, and complications. Furthermore, sentiment analysis was performed using sentiment lexicon software SentiWordNet. Cluster techniques such as topic and probabilistic clustering were applied as well as keyword extraction and topic identification. An interesting study explored the number of messages on different social media platforms during the measles outbreak in Netherlands in 2013 and compared them with both the number of related-online news and the number of reported measles cases [33]. Classification, text mining techniques, and manual sentiment analysis were used.

3. Institutional Background

As one of the largest online communities [34], Reddit is a social networking site that encompasses a variety of online forums and communities and was dubbed "the front page of the internet" [34]. In 2018, Reddit was ranked as the fourth most visited social website in the United States and the sixth worldwide with 234 million users [35]. In 2020, its user base has reached 430 million active monthly users and 130,000 active communities according to statistics on its website. Reddit is broken down into communities called "subreddits" that focus on different topics such as news, politics, gaming, and videos. The platform provides interactive features to engage the users. For instance, users can signal their support of a post by clicking on the upvote arrow or click on the downvote arrow to indicate their disapproval. The total score of the post is shown to indicate the number of upvotes minus the number of downvotes.

Academic research has explored the content of Reddit discussion. For example, two studies investigated the text polarity, age, geographic distribution of users, and product acquisition related to discussion on e-cigarette use [36, 37]. Another study explored the questions posted on Reddit about gout illness and classified them into 13 categories such as symptom uncertainty and diagnosis [38]. Another research examined user posting behavior on Reddit based on network structure and revealed that most users participate in one Reddit community [39]. Another study on the weight loss subreddit examined how online interactions affect weight loss in regards to the number of votes and replies received and used topic modeling and hierarchical clustering algorithm to identify global topics and local clusters [40].

4. Sample

We collected publicly available data about one mandatory vaccine – the MMR vaccine – and one voluntary vaccine – the flu vaccine – from Reddit. The MMR vaccine is a vaccine against three viruses: Measles, Mumps, and Rubella. We chose the MMR vaccine as the mandatory vaccine of our research for two reasons. First, the three diseases that the MMR vaccine protects humans against can cause serious health complications including bronchopneumonia, brain damage, mental retardation, fetal anomalies, and parotitis [41]. Hence, the MMR vaccine is vital. Second, the MMR vaccine had an unproven linkage with autism that created a controversy [41] and there has been debate on whether children should receive the vaccine. As a result, it is important to study this vaccine, due to its importance but possible side effect.

We chose the flu vaccine as the voluntary vaccine to examine. While it is optional to take the flu vaccine, it has many benefits according to the Centers for Disease Control and Prevention (CDC) including the reduction of risks associated with the flu illness and reducing flu-related doctor visits by 40-60% [42]. On the other hand, not all people agree with these benefits. Many people claim getting the flu after receiving the vaccine and the vaccine make them sick [43]. In addition, there are potential sources of bias in studies related to flu vaccine effectiveness, particularly among the elderly [44]. In addition, mercury in thimerosal found in some flu vaccines could be decomposed to toxic compounds, making it of high concern for many parents and individuals [45]. Hence, the benefits as well as potential side effects of both vaccines allow us to examine how online discussion differs on mandatory versus optional vaccines.

Posts that were available on Reddit in May 2020 with "MMR vaccine" or "measles, mumps, and rubella vaccine" in the title were collected for the MMR vaccine and posts containing "flu vaccine", "influenza vaccine" in the title were collected for the flu vaccine. We obtained 11,176 posts on the MMR vaccine and 10,152 posts on the flu vaccine. The descriptive statistics for the two types of vaccine posts and comments are shown in Table 1. There were more comments on MMR vaccine-related posts compared with the flu vaccine discussion. The post score is a rating Reddit gave to posts based on factors including up votes and down votes. Up votes are given by users to posts that they think that posts contribute to a conversation, and down votes are given to posts that user thinks they do not contribute to the conversation in the subreddit. Posts about MMR vaccines have higher average score than flu vaccine posts.

Table 1. Sample descriptive statistics

Variable	Min	Max	Mean	Std. Dev.
Flu Vaccine Posts (N_1=10,152)				
# Comments on a Post	1	1,982	71.27	246.53
Post Score	0	49,292	1,146.84	5,552.27
Up Vote (%)	0.14	1	0.81	0.18
MMR Vaccine Posts (N_2=11,176)				
# Comments on a Post	1	3,468	61.21	279.31
Post Score	0	77,348	881.47	6,159.56
Up Vote (%)	0.18	1	0.83	0.17

5. Text Analytics Results

Our text mining process includes data collection, data preprocessing and cleansing, and text analysis. After collecting the posts, we first performed text preprocessing to improve the effectiveness of text analytics by removing unnecessary text from the analysis that may overwhelm the analysis [46]. This process included converting the text to lower case, removing stop words that do not provide useful information [47], and reducing the words to their stems [48]. In addition, we removed punctuations and numbers and stripped extra white space in the text. To keep the amount of text we analyze manageable, we only retained terms that appeared in at least 2.5% of the posts. We generated our document term matrix using these remaining words and performed our subsequent term association and topic modeling analyses.

5.1 Term Association Analysis

For the term association analysis on the flu vaccine-related posts, we started with the term 'flu' and identified 18 first-level terms that had at least a 0.20 correlation with flu. Then we identified 76 second-level terms that had at least a 0.20 correlation with one of the first-level terms. Due to space constraint, we summarize only the first-level terms and their correlations with the term 'flu' in Table 2. Next, we plot the resulting first- and second-level term associations as edges in a term network in Figure 1. The size of each vertex is proportional to its eigenvector centrality in the network. That is, the larger the node representing the term, the more influential the term is in the network. The network graph reveals that the most important terms in flu vaccine-related discussions were vaccine, flu, year, get, virus, influenza, strain, risk, effect, people, even and got. These term association results reveal that the flu vaccine-related discussions focused on the requirement of annual vaccination, different strains of the flu viruses, protection offered by the flu vaccines, the impact of the vaccine on the immune system, and the risks involved in getting the flu shots.

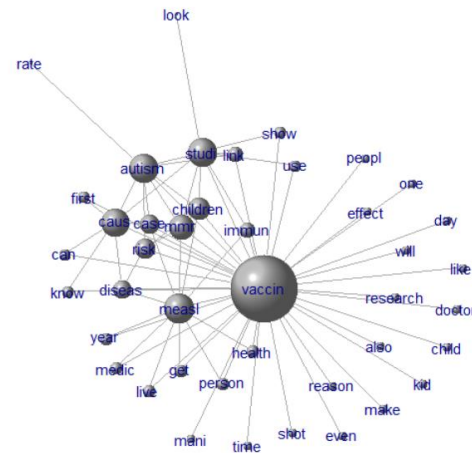We performed a similar term association analysis for MMR vaccine-related posts using 'mmr' as the starting term and identified six first-level terms with at least a 0.20 correlation with 'mmr' and 78 second-level terms with at least a 0.20 correlation with one of the first-level terms. We summarize the first-level terms and their correlations with 'mmr' in Table 3. Figure 2 shows the term association network graph for

the MMR vaccine-related posts. These results suggest that the discussions on the MMR vaccines focused on parents vaccinating their children and the risks associated with the vaccine. The linkage between the MMR vaccine and autism and the need for studies and evidence to support or negates this issue were also discussed. The term network diagram for the MMR vaccine discussion is denser.

Table 2. Flu vaccine posts term association analysis results

Term 1	Term 2	Correlation
flu	shot	0.53
flu	get	0.46
flu	vaccin	0.42
flu	year	0.37
flu	peopl	0.28
flu	everi	0.27
flu	strain	0.26
flu	got	0.25
flu	immun	0.24
flu	season	0.23
flu	week	0.22
flu	effect	0.21
flu	help	0.21
flu	still	0.21
flu	virus	0.21
flu	even	0.20
flu	like	0.20
flu	risk	0.20

Figure 1. Term association network for flu vaccine-related posts

Table 1. MMR vaccine posts term association analysis results.

Term 1	Term 2	Correlation
mmr	vaccin	0.34
mmr	autism	0.23
mmr	risk	0.22
mmr	measl	0.22
mmr	studi	0.22
mmr	caus	0.20

Figure 2. Term association network for MMR vaccine-related posts.

We summarize the comparison of the term association network analysis results for the discussions on the two vaccines in Table 4. We note that flu vaccine discussions mentioned more terms or vertices with more term associations or edges higher than 0.20 than the MMR vaccine-related discussions. However, the term association network density is higher for the MMR vaccine-related discussions. These results suggest that the discussions on the flu vaccines are more sparse compared with the MMR vaccine-related discussions.

Table 4. Comparison of term association network analysis results

Characteristic	Flu Vaccine Posts	MMR Vaccine Posts
# Vertices	131	97
# Edges	94	84
Network Density	0.013	0.019

5.2 Sentiment Analysis Results

We next performed sentiment analysis [49]. We analyzed the emotions expressed in each vaccine discussion using the Linguistic Inquiry and Word Count (LIWC) 2015 software [50]. The LIWC has an internal dictionary that defines the words to be counted in the required text file [51]. The LIWC2015

dictionary encompasses 6400 word stems, words, and emoticons. Sub dictionaries and word categories are defined using words entries included in dictionaries. Scales of categories are defined by a list of dictionary words. Some categories follow a hierarchical arrangement, which means that some categories are included in a broader category. For example, sadness words are also included in the broader negative emotions category. The dictionary also captures stems of the words in order to group words with the same stem together. There are main steps in the creation of the main LIWC2015 dictionary including word collection, judge rating, base rate analysis, candidate word list generation, psychometric evaluation, refinement, and addition of summary variables including analytical thinking, clout, authenticity, and emotional tone [52].

Examples of flu-vaccine discussion with positive sentiment are: "Thank you! Vaccines are my passion and I'm working hard in staying in this field for the rest of my career" and "I got my flu shot! ... It took less than five minutes for it all and I'm SO glad I got it. I can't wait to not get the flu!" Examples of flu-vaccine discussion with negative sentiment are:

"This is bad information. A healthy adult still needs the flu shot…" and "people who won't get the flu shot (or take proper precautions) make me so angry."

Examples of MMR vaccine discussion with positive sentiment are: "Today my daughter was in the lucky 5% of people who developed a mild rash and fever as a result of the vaccine. Lucky because I'd much rather her have a mild fever and some polkadots than fucking measles" and "[I]t's great to see Vaccine rates have increased during this measles outbreak we need herd immunity!" Examples of MMR vaccine discussion with negative sentiment are: "… I'm up to date on all my vaccines, but my titres for MMR show I keep losing immunity ... I'm angry that we're not doing more to stop this pro-plague idiocy. Get vaccinated!" and "[f]riend of mine's kid got roseola a few days after the MMR. Blames all vaccines, again. Purple made me angry."

Table 5 summarizes the LIWC sentiment analysis results including positive emotions, negative emotions, anxiety, anger, and sadness. Anxiety, anger and sadness are considered subcategories of negative emotions. Overall, the results suggest that discussions on flu and MMR vaccines were not statistically different in terms of positive emotions but the discussions on the flu vaccine were more negative compared with those on the MMR vaccine. A more in-depth analysis on the negative emotions shows that there were also more negative emotions of anxiety and anger in the flu vaccine-related posts than the MMR vaccine-related ones. However, discussions on the two types of vaccines do not differ significantly in the negative emotion of sadness.

Table 5. Comparison of Sentiment Analysis Results

Senti-ment	Mean of Flu Vaccine Posts	Mean of MMR Vaccine Posts	t-Stat on Difference in Means
Positive emotions	3.570 (0.082)	3.372 (0.078)	1.570
Negative emotions	2.712 (0.056)	2.415 (0.050)	4.372***
Anxiety	0.428 (0.017)	0.317 (0.017)	4.196***
Anger	1.125 (0.039)	0.962 (0.033)	3.501***
Sadness	0.306 (0.016)	0.285 (0.017)	0.888

Note: Standard deviations in parentheses. *** p<0.01; ** p<0.05; * p<0.10.

5.3 Topic Modeling Analysis Results

We performed topic modeling of the posts using Latent Dirichlet Allocation (LDA). LDA is a text mining technique that helps in finding hidden relationships among text documents and discover topics among them [53]. The use of topic modeling for social media analytics research is gaining traction [53] as it helps understand discussions and reactions of individuals participating in different social media sites. As a probabilistic topic modeling technique, the main idea behind the LDA is that "documents are represented as random mixtures over latent topics" and topics have a distribution over words [53]. We specified six, ten, and fifteen topics in our LDA analysis and obtained similar results. Table 6 summarizes the results based on six topics with the terms associated with each topic generated by the LDA for the flu vaccine-related discussions and MMR vaccine-related discussions, respectively.

For the flu vaccine-related discussions, we interpreted the topics as the following based on the terms associated with each topic: (1) flu vaccine effectiveness and risks, (2) need for annual flu vaccines and side effects, (3) flu vaccine strains and cancer risk, (4) general discussions on flu vaccines, (5) flu vaccines and children, and (6) getting answers to questions from healthcare professionals. Terms associated with the first topic show that the discussions focused on the effectiveness of the vaccine

against flu viruses and their risks in causing infection. Topic 2 is concerned with need to get the flu vaccine every year and the side effect of getting sick after the vaccine. Topic 3 focused on the capability for the flu vaccine to protect against different strains of the flu viruses and cancer risk. Topic 4 included general discussions. Topic 5 focused on flu vaccines and children, and how being around children without the flu vaccine may get them sick. Topic 6 included discussions on getting answers to questions from healthcare professionals such as nurses.

Table 6. LDA topic modeling results.

Topic	Terms									
	1	2	3	4	5	6	7	8	9	10
Flu vaccine-related discussions										
1. Flu vaccine effectiveness and risks	vaccin	effect	influenza	virus	studi	risk	caus	season	infect	protect
2. Need for annual flu vaccines and side effects	flu	get	shot	year	got	one	time	sick	day	never
3. Flu vaccine strains and cancer risk	can	immun	system	use	differ	will	virus	make	cancer	strain
4. General discussions	just	like	know	think	peopl	say	thing	Don't	realli	Something
5. Flu vaccines and children	peopl	work	make	can	will	even	take	also	kid	now
6. Getting answers to questions from healthcare professionals	good	want	see	remov	question	person	comment	medic	thank	nurs
MMR vaccine-related discussions										
1. Debate about MMR vaccines	like	peopl	just	say	think	thing	believ	antivaxx	bad	good
2. Effects and risks on children	vaccin	measl	mmr	immun	diseas	children	effect	risk	year	case
3. Timing of MMR vaccines	get	just	kid	know	time	one	even	need	still	now
4. Research on MMR vaccines	studi	link	one	research	scienc	claim	evidence	use	read	find
5. General discussions	peopl	will	make	want	right	person	can	post	comment	question
6. MMR vaccines and autism	autism	caus	actual	like	differ	use	much	also	first	autist

For MMR vaccine-related discussions, we interpreted the discussions as the following based on the terms associated with each topic: (1) debate about MMR vaccines, (2) effects and risks of MMR vaccines on children, (3) timing of MMR vaccine, (4) research and evidence on MMR vaccines, (5) general discussions, and (6) MMR vaccines and autism. Topic 1 focused on the debate about MMR vaccine and some people who are against the vaccine. Topic 2 included discussions on the effectiveness of the MMR vaccine and its risks on children. Topic 3 related to the discussions on the timing of the MMR vaccine at one year of age. Topic 4 focused on research and scientific studies on MMR vaccines. Topic 5 included general discussions on the MMR vaccines. Topic 6 is concerned with the relationship between the MMR vaccine and autism. Some people argue that the vaccine could cause autism, while other people call for a scientific evidence and research to proof this linkage.

6. Discussion

6.1 Results Summary

We analyzed online discussions about flu and MMR vaccine using different types of text analytics techniques. Term association analysis of Reddit threads reveals the terms that are highly associated with the flu vaccine including get, every, year, season, strain, and people. The terms that are highly associated with the MMR vaccine include vaccine, risk, autism,

cause, and disease. The results imply that discussions about voluntary vaccines such as the flu vaccine is concerned with encouraging or discouraging people to take the vaccine since it is their choice to take the vaccine. However, the mandatory vaccine discussion is more focused on risks claimed to be related to the vaccination such as autism. In addition, since the MMR vaccine is given to children, the most discussed age group is the children. The flu vaccine is given to different age groups so that the people term is associated with it in the online discussion.

The sentiment analysis shows that sentiment and affect expressed for the two vaccine are different. Flu vaccine discussions are shown to be more emotional positively and negatively. Emotions such as anger, sadness, and anxiety are expressed in the discussion.

Topics discussed in regards to the two vaccines varied as well. The effectiveness of the flu vaccine is one of the main topics that are common and diffused. The MMR vaccine risks and the need for evidence and credible information from popular health protection agencies such as the Center for Disease Control (CDC) are the most common topics discussed.

6.2 Theoretical Contribution

We examined how two different types of vaccines are discussed on online communities. Our research has the following contribution to theory. First, our study extends the online health communities literature by examining their content which may reveal latent variables that could be hard to be found by other means. Second, our research is one of the first to compare and contrast the online discussion of two types of vaccines, one is mandatory and the other is voluntary using text mining including topic modeling and sentiment analysis. Our results enable researchers to get a better insight about vaccine acceptance and rejection and the factors that impact them. And thus researchers and psychologist can explore these factors sufficiently.

Third, our sentiment analysis results reveal the importance of emotions expressed in online discussion, which could also help in identifying anti-vaccine movement and improve the knowledge about the factors behind vaccine rejection such as individual's personality treat, bad experience with vaccines, or misinformation. Misinterpretations of causality between MMR vaccine and autism is one example that many people debate. Evidence and research can help to negate these doubts and misinformation [1].

6.2 Practical Implications

Our results highlight the importance of using online health communities and social networks to discuss various health-related issues such as vaccines. The online content helps health agencies to improve vaccine communication and address the concerns associated with vaccines to enhance vaccine confidence and eliminate vaccines hesitancy [23]. Public healthcare strategies could leverage online content and big data power to be more effective and control infectious diseases spread and prevent relate misinformation dissemination [6]. Administrative agencies could use the suitable awareness campaigns for different types of vaccines to target the most concerned people such as new parents for vaccines required for newborns and infants [7]. Moreover, our content analysis reveals the importance of social media to disseminate required information from credible agencies so that misinformation could be prevented. The topic analysis shows the need for evidence of vaccine effectiveness and related risks. Health agencies could fulfill individuals' needs by providing them with this information and thus encourage them to take the vaccines.

7. Conclusion, Limitations, and Future Research Directions

The study reveals that different types of vaccines have different online discussions. Mandatory vaccines have more online participation. Voluntary vaccines have more controversial discussion. Text mining of vaccines online discussion can help in understanding the concerns and beliefs related to vaccination. Our work has some limitations. First, we analyzed publicly available data about posts from Reddit. Other data such as the posts structure and who replies to whom could be helpful to understand the nature of these social networks and their social influence. Future research could address this limitation. Second, our study is focused on flu and MMR vaccine. These two vaccines are highly recommended to achieve personal and public protection, particularly for healthcare workers [54], however, online conversations concerning other essential vaccines could be examined such as hepatitis B vaccine. Finally, our study is limited to Reddit online communities. Future research could examine other online health communities and compare the content from different platforms.

8. References

[1] G. Aghili and L. Lapointe, "The Role of Online Communities in Vaccine Controversies," 2019.

[2] J. M. Glanz *et al.*, "Web-based social media intervention to increase vaccine acceptance: a randomized controlled trial," *Pediatrics,* vol. 140, no. 6, p. e20171117, 2017.

[3] M. F. Daley, K. J. Narwaney, J. A. Shoup, N. M. Wagner, and J. M. Glanz, "Addressing parents' vaccine concerns: a randomized trial of a social media intervention," *American journal of preventive medicine,* vol. 55, no. 1, pp. 44-54, 2018.

[4] B. Love, I. Himelboim, A. Holton, and K. Stewart, "Twitter as a source of vaccination information: content drivers and what they are saying," *American journal of infection control,* vol. 41, no. 6, pp. 568-570, 2013.

[5] G. Bello-Orgaz, J. Hernandez-Castro, and D. Camacho, "Detecting discussion communities on vaccination in twitter," *Future Generation Computer Systems,* vol. 66, pp. 125-136, 2017.

[6] E. K. Brunson, "The impact of social networks on parents' vaccination decisions," *Pediatrics,* vol. 131, no. 5, pp. e1397-e1404, 2013.

[7] E. Fukuda, S. Kokubo, J. Tanimoto, Z. Wang, A. Hagishima, and N. Ikegaya, "Risk assessment for infectious disease and its impact on voluntary vaccination behavior in social networks," *Chaos, Solitons & Fractals,* vol. 68, pp. 1-9, 2014.

[8] M. N. Hajli, J. Sims, M. Featherman, and P. E. Love, "Credibility of information in online communities," *Journal of Strategic Marketing,* vol. 23, no. 3, pp. 238-253, 2015.

[9] M. Alazazi and E. W. Ayaburi, "Effect of Stigmatization and Privacy Concerns on Engagement in Virtual Health Communities," 2019.

[10] S. J. Sowles *et al.*, "A content analysis of an online pro-eating disorder community on Reddit," *Body image,* vol. 24, pp. 137-144, 2018.

[11] S. Myneni, K. Fujimoto, N. Cobb, and T. Cohen, "Content-driven analysis of an online community for smoking cessation: integration of qualitative techniques, automated text analysis, and affiliation networks," *American journal of public health,* vol. 105, no. 6, pp. 1206-1212, 2015.

[12] S. Mukherjee, G. Weikum, and C. Danescu-Niculescu-Mizil, "People on drugs: credibility of user statements in health communities," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 65-74.

[13] N. Elhadad, S. Zhang, P. Driscoll, and S. Brody, "Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions," in *AMIA Annual Symposium Proceedings*, 2014, vol. 2014: American Medical Informatics Association, p. 516.

[14] A. Shepherd, C. Sanders, M. Doyle, and J. Shaw, "Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation," *BMC psychiatry,* vol. 15, no. 1, p. 29, 2015.

[15] Q. B. Liu, X. Liu, and X. Guo, "The Effects of Participating in a Physician-Driven Online Health Community in Managing Chronic Disease: Evidence from Two Natural Experiments," *MIS Quarterly,* vol. 44, no. 1, 2020.

[16] X. Zhang, S. Liu, Z. Deng, and X. Chen, "Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users," *Computers in Human Behavior,* vol. 75, pp. 797-810, 2017.

[17] J. Huh, D. W. McDonald, A. Hartzler, and W. Pratt, "Patient moderator interaction in online health communities," in *AMIA Annual Symposium Proceedings*, 2013, vol. 2013: American Medical Informatics Association, p. 627.

[18] M. Haverkate *et al.*, "Mandatory and recommended vaccination in the EU, Iceland and Norway: results of the VENICE 2010 survey on the ways of implementing national vaccination programmes," *Eurosurveillance,* vol. 17, no. 22, p. 20183, 2012.

[19] A. Gesser-Edelsburg, N. Walter, Y. Shir-Raz, and M. S. Green, "Voluntary or mandatory? The valence framing effect of attitudes regarding HPV vaccination," *Journal of health communication,* vol. 20, no. 11, pp. 1287-1293, 2015.

[20] E. Galanakis, A. Jansen, P. L. Lopalco, and J. Giesecke, "Ethics of mandatory vaccination for healthcare workers," *Eurosurveillance,* vol. 18, no. 45, p. 20627, 2013.

[21] F. H. Chen, "A susceptible-infected epidemic model with voluntary vaccinations," *Journal of mathematical biology,* vol. 53, no. 2, pp. 253-272, 2006.

[22] D. J. Langley, R. Wijn, S. Epskamp, and R. Van Bork, "Should I get that jab? Exploring influence to encourage vaccination via online social media," 2015.

[23] G. J. Kang *et al.*, "Semantic network analysis of vaccine sentiment in online social media," *Vaccine,* vol. 35, no. 29, pp. 3621-3638, 2017.

[24] S. A. Salloum, M. Al-Emran, and K. Shaalan, "Mining social media text: extracting knowledge from Facebook," *International Journal of Computing and Digital Systems,* vol. 6, no. 02, pp. 73-81, 2017.

[25] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A survey of text mining in social media: facebook and twitter perspectives," *Adv. Sci. Technol. Eng. Syst. J,* vol. 2, no. 1, pp. 127-133, 2017.

[26] J. K. Rout, K.-K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," *Electronic Commerce Research,* vol. 18, no. 1, pp. 181-199, 2018.

[27] C. Mhamdi, M. Al-Emran, and S. A. Salloum, "Text mining and analytics: A case study from news channels posts on Facebook," in *Intelligent Natural Language Processing: Trends and Applications*: Springer, 2018, pp. 399-415.

[28] M. Siering, J.-A. Koch, and A. V. Deokar, "Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts," *Journal of Management Information Systems,* vol. 33, no. 2, pp. 421-455, 2016.

[29] H. Wimmer, V. Y. Yoon, and V. Sugumaran, "A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy," *Decision Support Systems,* vol. 88, pp. 51-66, 2016.

[30] G. W. Yun *et al.*, "Social media and flu: Media Twitter accounts as agenda setters," *International journal of medical informatics,* vol. 91, pp. 67-73, 2016.

[31] X. Huang *et al.*, "Examining patterns of influenza vaccination in social media," in *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[32] Y. Lu, Y. Wu, J. Liu, J. Li, and P. Zhang, "Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community," *Journal of medical Internet research,* vol. 19, no. 4, p. e109, 2017.

[33] L. Mollema *et al.*, "Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013," *Journal of medical Internet research,* vol. 17, no. 5, p. e128, 2015.

[34] P. Singer, F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier, "Evolution of reddit: from the front page of the internet to a self-referential community?," in *Proceedings of the 23rd international conference on world wide web*, 2014, pp. 517-522.

[35] G. Delnevo *et al.*, "Patients reactions to non-invasive and invasive prenatal tests: a machine-based analysis from reddit posts," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018: IEEE, pp. 980-987.

[36] Y. Zhan, Z. Zhang, J. M. Okamoto, D. D. Zeng, and S. J. Leischow, "Underage juul use patterns: Content analysis of reddit messages," *Journal of medical Internet research,* vol. 21, no. 9, p. e13038, 2019.

[37] E. I. Brett *et al.*, "A content analysis of JUUL discussions on social media: Using Reddit to understand patterns and perceptions of JUUL use," *Drug and alcohol dependence,* vol. 194, pp. 358-362, 2019.

[38] C. Derksen, A. Serlachius, K. J. Petrie, and N. Dalbeth, ""What say ye gout experts?" A content analysis of questions about gout posted on the social news website Reddit," *BMC musculoskeletal disorders,* vol. 18, no. 1, p. 488, 2017.

[39] C. Buntain and J. Golbeck, "Identifying social roles in reddit using network structure," in *Proceedings of the 23rd international conference on world wide web*, 2014, pp. 615-620.

[40] Y. Liu and Z. Yin, "Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques," *Journal of Medical Internet Research,* vol. 22, no. 6, p. e13745, 2020.

[41] J. C. Watson, S. C. Hadler, C. A. Dykewicz, S. Reef, and L. Phillips, "Measles, Mumps, and Rubella-Vaccine Use and Strategies for Elimination of Measles, Rubella, and Congenital Rubella Syndrome and Control of Mumps: Recommendations of the Advisory Committee on Immunization Practices (ACIP). Vol. 47/No. RR-8," CENTERS FOR DISEASE CONTROL ATLANTA GA, 1998.

[42] "What are the benefits of flu vaccination?" Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD). https://www.cdc.gov/flu/prevent/vaccine-benefits.htm (accessed 2020).

[43] B. Nyhan and J. Reifler, "Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information," *Vaccine,* vol. 33, no. 3, pp. 459-464, 2015.

[44] C. Trucchi, C. Paganino, A. Orsi, D. De Florentiis, and F. Ansaldi, "Influenza vaccination in the elderly: why are the overall benefits still hotly debated?," *Journal of preventive medicine and hygiene,* vol. 56, no. 1, p. E37, 2015.

[45] D. A. Drum, "Are toxic biometals destroying your children's future?," *Biometals,* vol. 22, no. 5, pp. 697-700, 2009.

[46] R. Igawa, A. Almeida, and B. Zarpelão, "Recognition of compromised accounts on twitter," in *Anais Principais do XI Simpósio Brasileiro de Sistemas de Informação*, 2015: SBC, pp. 9-14.

[47] N. Kühl, M. Mühlthaler, and M. Goutier, "Automatically quantifying customer need tweets: Towards a supervised machine learning approach," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[48] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining-an overview," *International Journal of Computer Science & Communication Networks,* vol. 5, no. 1, pp. 7-16, 2015.

[49] B. Biswas, A. Mukhopadhyay, and G. Gupta, ""Leadership in Action: How Top Hackers Behave" A Big-Data Approach with Text-Mining and Sentiment Analysis," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[50] J. W. Pennebaker, R. J. Booth, R. L. Boyd, and M. E. Francis, "Linguistic inquiry and word count: Liwc 2015 [computer software]. pennebaker conglomerates," ed: Inc, 2015.

[51] B. Biswas, P. Sengupta, and D. Chatterjee, "Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques," *Decision Support Systems,* p. 113324, 2020.

[52] R. L. B. James W. Pennebaker and a. K. B. Kayla Jordan, "The development and psychometric properties of LIWC2015," *Austin, TX: University of Texas at Austin,* 2015.

[53] H. Jelodar *et al.*, "Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey," *Multimedia Tools and Applications,* vol. 78, no. 11, pp. 15169-15211, 2019.

[54] K. Little *et al.*, "Occupational vaccination of health care workers: uptake, attitudes and potential solutions," *Public Health,* vol. 129, no. 6, pp. 755-762, 2015.