

Wright State University

CORE Scholar

---

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

---

2022

## The Impact of Sentiment and Misinformation Cycling Through the Social Media Platform, Twitter, During the Initial Phase of the COVID-19 Vaccine Rollout

Emily Grace Burwell  
*Wright State University*

Follow this and additional works at: [https://corescholar.libraries.wright.edu/etd\\_all](https://corescholar.libraries.wright.edu/etd_all)



Part of the [Biology Commons](#)

---

### Repository Citation

Burwell, Emily Grace, "The Impact of Sentiment and Misinformation Cycling Through the Social Media Platform, Twitter, During the Initial Phase of the COVID-19 Vaccine Rollout" (2022). *Browse all Theses and Dissertations*. 2592.

[https://corescholar.libraries.wright.edu/etd\\_all/2592](https://corescholar.libraries.wright.edu/etd_all/2592)

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).

THE IMPACT OF SENTIMENT AND MISINFORMATION CYCLING THROUGH  
THE SOCIAL MEDIA PLATFORM, TWITTER, DURING THE INITIAL PHASE OF  
THE COVID-19 VACCINE ROLLOUT

A thesis submitted in partial fulfilment of the  
Requirements for the degree of  
Master of Science

By

EMILY GRACE BURWELL  
B.S., Ohio University, 2020

2022  
Wright State University

COPYRIGHT BY  
EMILY GRACE BURWELL  
2022

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

April 26, 2022

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Emily Grace Burwell ENTITLED The Impact of Sentiment and Misinformation Cycling Through the Social Media Platform, Twitter, During the Initial Phase of the COVID-19 Vaccine Rollout BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

---

William Romine, Ph.D.

Thesis Director

---

Scott Baird, Ph.D.

Chair, Department of Biological Sciences

Committee on Final Examination:

---

William Romine, Ph.D.

---

Paula Bubulya, Ph.D.

---

Jeffrey Peters, Ph.D.

---

Barry Milligan, Ph.D.

Vice Provost for Academic Affairs  
Dean of the Graduate School

## ABSTRACT

**Burwell, Emily Grace. M.S. Department of Biological Sciences, Wright State University, 2022. The Impact of Sentiment and Misinformation Cycling Through the Social Media Platform, Twitter, During the Initial Phase of the COVID-19 Vaccine Rollout**

This study assesses the underlying topics, sentiment, and types of information regarding COVID-19 vaccines on Twitter during the initiation of the vaccine rollout. Tweets about the COVID-19 vaccine were collected and the relevant tweets were then filtered out using a relevancy classifier. Latent Dirichlet Allocation (LDA) was used to uncover topics of discussion within the relevant tweets. The NRC lexicon was used to assess positive and negative sentiment within tweets. The type of information (information, misinformation, opinion, or question) in tweets was evaluated.

The relevancy classifier resulted in a dataset of 210,657 relevant tweets. Eight topics provided the best representation of the relevant tweets. Tweets with negative sentiment were associated with a higher percentage of misinformation. Tweets with positive sentiment showed a higher percentage of information. The proliferation of information and misinformation on social media platforms are associated with building trust and mitigating negative sentiment associated with COVID-19 vaccines.

## TABLE OF CONTENTS

	Page
I. COVID-19 INTRODUCTION .....	1
Coronaviruses .....	1
COVID-19 Symptoms .....	3
Global Crisis .....	3
COVID-19 Vaccination .....	4
Vaccine Sentiment .....	6
Misinformation on Social Media Platforms .....	8
Twitter .....	10
II. INTRODUCTION TO STUDY AND PURPOSE .....	11
Limitations of Prior Studies .....	11
Importance and Purpose .....	11
Research Questions .....	12
Experimental Design .....	13
III. METHODS .....	16
Data Preparation .....	16
Data Collection .....	16
Data Annotation .....	17
Relevancy Classification .....	17
Topic Modelling .....	18

	Sentiment and Emotion Analyses . . . . .	18
	Statistical Analysis . . . . .	19
	Misinformation Analysis . . . . .	19
IV.	RESULTS . . . . .	21
	Relevancy Classification . . . . .	21
	Relevancy and Irrelevant Tweets . . . . .	22
	Topic Modelling . . . . .	24
	Optimal Number of Topics . . . . .	24
	Naming Topics . . . . .	24
	Sentiment Analysis . . . . .	38
	NRC Lexicon Analysis . . . . .	38
	Statistical Analysis. . . . .	52
	Kruskal-Wallis Tests . . . . .	52
	Post-hoc Mann-Whitney Tests . . . . .	52
	Positive and Negative Sentiment Medians . . . . .	53
	Sentiment Word Associations . . . . .	54
	Emotional Analysis . . . . .	58
	Individual Topic Means . . . . .	59
	Misinformation Analysis . . . . .	60
	Quantifying the Type of Information . . . . .	63
	Categorizing the Type of Information . . . . .	63

	The Distribution of Information and Misinformation . . . . .	65
V.	DISCUSSION . . . . .	68
	Limitation and Future Studies . . . . .	73
	Implications for Improving Public Health . . . . .	74
VI.	REFERENCES . . . . .	77



## LIST OF FIGURES

Figure	Page
1. Coherence Score Plot . . . . .	25
2. Topic Percent Contribution Distribution and Means of Individual Topics . . . . .	26
3. Median Scores of Positive and Negative Sentiment Compares Among 8 Topics . . . . .	54
4. Word Clouds for Positive and Negative Sentiments in Each Topic . . . . .	57
5. Comprehensive Mean Association Scores Compared . . . . .	60
6. The Distribution of Information in Positive and Negative Sentiment . . . . .	64
7. The Distribution of Information Between Topics in Positive and Negative Sentiment . . . . .	67

## LIST OF TABLES

Table	Page
1. Results of the Relevancy Classifier . . . . .	22
2. Relevant VS Irrelevant Representative Tweets. . . . .	22
3. Representative Tweets per Topic . . . . .	26
4. Representative Tweets for Sentiment . . . . .	38
5. Kruskal-Wallis Test Results . . . . .	52
6. Mean Association Scores from Emotion Analysis . . . . .	59
7. Representative Tweets for Types of Information . . . . .	61
8. Categorization of Representative Tweets . . . . .	64

## **ACKNOWLEDGEMENTS**

I would first like to express my gratitude to Dr. Romine for giving me the amazing opportunity to study in his lab and for being my thesis advisor. Thank you for guiding me through each step of this process. I would like to thank you for always being patient with me and giving me opportunities to work alongside you on additional projects. I feel that my time in your lab has truly allowed me to find my passion. Thank you for spending your time to help me learn and become a better research candidate. Next, I would like to thank my committee members, Dr. Bubuyla and Dr. Peters, for encouraging me and always providing helpful feedback on my work. I would also like to thank Ankita Agarwal who worked with me in Dr. Romine's lab. Thank you for taking the time to help teach me methods and help mentor me through my graduate experience. I would additionally like to thank all my professors at Wright State University for always inspiring me to learn.

I would like to also thank my fiancé, family, and friends for being so supportive of my education. Jacob, thank you for always believing in me. Thank you for showing me love when I needed to vent. Celebrating my success with you has given me so much joy. Mom, George, and Mary, thank you for always demonstrating a strong work ethic and being role models to me. Finally, thank you to my sweet pets, Marvin, Bowie, and Yoko, who provided companionship during the good and stressful times of my experience.

## **I. INTRODUCTION TO COVID-19 CORONAVIRUSES**

Coronavirus disease (COVID-19) is a respiratory illness caused by the severe acute respiratory syndrome coronavirus 2, also termed SARS-CoV-2. The first known case of COVID-19 was reported in Wuhan, China in December of 2019 (Shereen et al., 2020). The first reported case of COVID-19 reported in the United States was on January 20, 2020 (Holshue et al., 2020). Within only a matter of months from the first reported case in Wuhan, COVID-19 surpassed a disease outbreak and became a global pandemic. On January 30, 2020 COVID-19 became a public health emergency of international concern and on March 11, 2020 the World Health Organization (WHO) announced the start of the COVID-19 pandemic (Hiscott et al., 2020).

COVID-19 is a disease caused by one of the is one of several known coronaviruses, SARS-CoV-2. In fact, COVID-19 is not the first coronavirus to cause a global outbreak. In 2002, the SARS-CoV outbreak was established and had a fatality rate of 9.8% (Petersen et al., 2020). The Middle Eastern Respiratory Syndrome-CoV (MERS-CoV) was identified in 2012 and led to an epidemic with a fatality rate of 34% (Memish et al., 2020). MERS-CoV, SARS-CoV and SARS-CoV-2 are genetically related, and all are in the *Nidovirales* family meaning they use their messenger RNA (mRNA) to replicate (Konda et al., 2020). While COVID-19 has the lowest fatality rate (2.2%) in comparison to SARS-CoV and MERS-CoV, it has killed more people due to the high transmissibility and large number of the population that was infected with COVID-19

(Petersen et al., 2020). While SARS-CoV, SARS-CoV-2, and MERS-CoV are all related, there are notable fundamental differences that ultimately have contributed to the COVID-19 pandemic prospering and the virus being significantly more susceptible and lethal (Rossi et al. 2020).

Viruses require specific surface proteins called spike proteins in order to infect the host cell (Rossi et al. 2020). Coronaviruses encode between four and five structural proteins in their genome: spike, membrane, envelope, nucleocapsid, and hemagglutinin-esterase proteins (Konda et al., 2020). Viruses first recognize and bind to a specific receptor that is located on the cell's surface (Huang et al., 2020). It is important to emphasize that viruses and receptor proteins mimic a lock and key model. Therefore, a virus can only bind to a specific protein. Once the virus becomes bound to the receptor, it can enter the cell by penetration, membrane fusion, or endocytosis (Huang et al., 2020). The virus then begins invading cells and replicating itself (Huang et al., 2020).

MERS-CoV recognizes receptor dipeptidyl peptidase-4 (DPP4) which is highly expressed in the respiratory system while SARS-CoV and SARS-CoV-2 both specifically recognize angiotensin-converting enzyme 2 (ACE2) (Huang et al., 2020; Konda et al., 2020). ACE2 originally functions to degrade Angiotensin II, which can reduce the amount of oxygen in the body by increasing apoptosis in alveoli cells (Huang et al., 2020). When ACE2 is bound with either SARS-CoV or SARS-CoV-2, it cannot properly degrade Angiotensin II (Huang et al., 2020). This leads to the respiratory symptoms associated with both viruses (Huang et al., 2020). While both viruses exploit ACE2, they

have a key difference in their receptor-binding domain (RBD) (Huang et al., 2020). In SARS-CoV-2, the spike protein has a furin-like cleavage site unlike that observed in SARS-CoV (Konda et al., 2020). It is hypothesized that this difference increases the efficiency of viral entry because it primes the spike proteins allowing more receptors for binding and therefore causes the increased transmission rates associated with SARS-CoV-2 (Konda et al., 2020).

### **COVID-19 Symptoms**

Symptoms of COVID-19 vary substantially. Most commonly people present with flu-like symptoms and loss of smell and taste; however, more severe symptoms, such as trouble breathing, can arise (Jayaweera et al., 2020). On the other side of the spectrum, COVID-19 can present as asymptomatic. Asymptomatic individuals pose a troublesome situation because they can still transmit the virus, thus creating a challenging health concern (Chowdhury et al., 2021). COVID-19 is primarily transmitted by the exchange of respiratory droplets (from coughing or talking) (Jayaweera et al., 2020).

### **Global Crisis**

The incubation period of the virus is another consistent problem regarding transmission rates. The incubation period can be up to 14 days (WHO, 2020). The average incubation period for patients on average of 60 years old is 12.5 days (Quesada et al., 2021). This increases transmission rates because an individual can infect another before they show symptoms. This is a public health crisis being that over 80 million cases of COVID-19 in the United States have been reported and over 981,000 COVID-related

deaths have been documented as of April 2022 according to the Centers for Disease Control and Prevention (CDC) (“CDC Covid Tracker Data”, 2021). Worldwide, there have been over 258 million cases reported as of March 2022 (Ritchie et al. 2020).

Due to the global crisis, health and science organizations, such as the WHO and CDC, advise individuals to take preventative measures to reduce the transmission of COVID-19. This includes wearing a face mask, social distancing at least 6 feet, avoiding large gatherings, and practicing effective hygiene to avoid spreading respiratory droplets (“CDC Covid Tracker Data”, 2021).

### **COVID-19 Vaccination**

Even though preventative measures such as hand washing, wearing masks, and vaccination, have demonstrated effectiveness at reducing the transmission rates of COVID-19, there is a significant portion of the United States’ population that does not wish to comply with these practices, thus promoting the continuation of the COVID-19 pandemic. Currently, there are several highly effective vaccines in the United States that are available to help prevent severe COVID-19 cases and transmission. These vaccines include Pfizer-BioNTech, Moderna-NIAID, and Johnson & Johnson-Janssens. Pfizer-BioNTech and Moderna-NIAID use mRNA technologies while Johnson & Johnson-Janssens uses viral vector technology (Haynes 2021). mRNA is used to target spike proteins because it can act as a guide to generate spike proteins in the cells (Branswell, 2021). Spike proteins function to assist the virus in entering host cells (Shulla and Gallagher 2009). Therefore, when a vaccine has SARS-CoV-2 mRNA, the mRNA can

instruct the immune systems to build spike proteins (Hendaus & Jomha, 2021). Once the spike proteins are generated, the body begins producing antibodies for SARS-CoV-2 that attach to the spike proteins to prevent infection (Hendaus & Jomha, 2021). mRNA vaccines differ from traditional vaccines. For example, the mRNA vaccines do not use adenoviruses like the Johnson and Johnson-Janssens vaccine (Kaur and Gupta 2020). Because the mRNA vaccines do not contain live virus, there is no risk of contracting COVID-19 from the vaccine (Hendaus & Jomha, 2021).

Worldwide, there are other vaccines being used to prevent COVID-19. For example, Oxford-Astrazeneca, Novavax, and Covishield are not available in the United States, but all show similar efficacies to the vaccines widely used in the United States (Pormohammad et al. 2021).

According to the Centers for Disease Control and Prevention, 77% of the United States population has had one dose of the COVID-19 vaccine, whereas only 66% of the population is fully vaccinated as of April 2022 (“CDC Covid Data Tracker”, 2022). It is important to note that the term “vaccine” in this document is comprehensive of all the COVID-19 vaccines available unless a specific vaccine is specified. In comparison to a fully vaccinated individual, an unvaccinated individual has a 6.1 times greater risk of testing positive for COVID-19 and an 11.3 times greater risk of dying from COVID-19 (“CDC Covid Data Tracker”, 2022). Booster vaccines are now available for those who are eligible. The booster vaccine helps to reduce severe COVID-19 infections as the original dosage begins to lose effectiveness (Gabutti et al. 2014). As of March of 2022,



over 96 million booster vaccines have been administered in the United States (“CDC Covid Data Tracker”, 2022). Herd immunity refers to the majority of a population being immune to a targeted pathogen, therefore reducing the incidence rate (John and Samuel 2000). It is estimated that approximately 85% of the population would need to be fully vaccinated for this to be achieved assuming the virus does not mutate (Loomba et al. 2021).

### **Vaccine Sentiment**

Sentiment describes a method used to analyze feelings or emotions (Devika et al., 2016). Studying vaccine sentiment on Twitter can help describe the public’s perception and feelings towards vaccines. Raghupathi et al. (2020) conducted a vaccine sentiment study from January 1, 2019 to April 5, 2019. A total of 9581 tweets were determined to be negative, neutral, or positive. A negative tweet was determined by relating to anti-vaccination sentiment, whereas a positive tweet related to pro-vaccination sentiment (Raghupathi et al., 2020). Of all the tweets total, their data showed 43.3% tweets with negative sentiment, 40.4% with positive sentiment, and 16.3% with neutral sentiment (Raghupathi et al., 2020). Vaccine sentiment surrounding the COVID-19 vaccine differs slightly, but still breaks down into mostly negative and positive sentiment (Saleh et al., 2021). Saleh et al. (2021) conducted a COVID-19 vaccine sentiment study from February 1, 2020 until December 11, 2020. This study had a total of 2,356,285 tweets and analyzed the tweets by using a sentiment reasoner (VADER) (Saleh et al., 2021). VADER scored the tweets according to positive, negative, and neutral sentiment. They then created a

standardized threshold for each sentiment and scaled the scores accordingly. Positive tweets were equal or greater than -0.05, neutral tweets were defined by a scale of -0.05 to 0.05, and negative tweets were less than or equal to -0.05 (Saleh et al., 2021). Saleh et al. (2021) distinguished different emotions such as anger, disgust, fear, trust, anticipation, joy, sadness, and surprise, as either negative or positive. These were compared over time to determine if sentiment regarding the COVID-19 vaccine changed over time. Fear was the most prevalent emotion (Saleh et al., 2021). Fear was found in 40% of tweets initially but decreased to about 20% by the end of the study (Saleh et al., 2021). In contrast, trust increased from 20% initially to 40% by the end of the study (Saleh et al., 2021). Tweets showing anticipation sentiment maintained a 25%-30% prevalence rating throughout the duration of the study (Saleh et al., 2021). All other emotions examined were present in less than 10% of tweets during the study (Saleh et al., 2021).

Understanding people's perceptions of vaccinations can be fostered by using social media platforms such as Twitter to obtain data. Additionally, information regarding the mechanisms of vaccines and how they work can also be promoted through social media like Twitter. When the COVID-19 pandemic began, Twitter became a popular information outlet for people all over the world. Verified accounts, such as the CDC (@CDCgov) and WHO (@WHO), can relay information to the public regarding preventative measures, symptoms, vaccination sites, vaccine information, and debunking myths regarding COVID-19. While the CDC and WHO provide information, which reflects current scientific consensus, Twitter's platform encourages all users to share

content regardless of whether or not it is factual. Twitter has been used to amplify both factual information and misinformation regarding the COVID-19 vaccine. In fact, during the COVID-19 pandemic, from October 2019 to June 2020, vaccine opposition on Twitter has increased 80% (Bonnevie et al., 2021). This is likely due to the rapid spread of misinformation and subsequent distrust in health officials (Bonnevie et al., 2021).

### **Misinformation on Social Media Platforms**

The abundance of misinformation/misconceptions and the lack of agreement regarding COVID-19 vaccination and precautions is contributing to a complex public health challenge. Ultimately, this has allowed the COVID-19 pandemic to be accompanied by an infodemic (Cinelli et al., 2020). An infodemic describes the generation of misinformation during a disease outbreak (Cinelli et al., 2020). Prior studies have indicated that misinformation about the COVID-19 vaccination can increase hesitancy in receiving the vaccine for oneself or accepting it for a loved one (Loomba et al., 2021). Declining the vaccine can affect herd immunity being achieved. As of November of 2021, only 59% of the United States was fully vaccinated according to the CDC's vaccine tracker ("CDC Covid Data Tracker", 2022). However, looking in May of 2022, the percent of fully vaccinated people in the United States has increased to 66% ("CDC Covid Data Tracker", 2022).

Social media platforms pride themselves on providing a space that promotes a user's uncensored thoughts. This creates opportunities for users to be influential while remaining relatively unknown. Chowdhury et al. (2021) sought to analyze the

information being shared on 11 different social media platforms such as Twitter, Reddit, Youtube, Gab and Facebook. Overall, this study found that Youtube and Reddit reduced unreliable posts, while platforms such as Gab, exacerbated unreliable posts (Chowdhury et al. 2021). This study distinguished information to be reliable by using MEDLINE as a reference database. MEDLINE is the National Library of Medicine's (NLM) database (MEDLINE Overview, 2021). This database was used to search words in MEDLINE for misinformation and information. For misinformation they used search words such as false news, misinformation, and hoax. For factual disease-outbreak information they searched words such as COVID-19, Coronavirus, and viruses. All information was screened based on two factors: title-abstract review and full-text review. Information in-text was reviewed to assure it was relevant to the research questions. Twitter was found to have neutral content meaning that misinformation is equally as prevalent as factual information (Chowdhury et al. 2021). This supports Roozenbeek et al. (2020) because the predominant user on Gab identifies as far right politically and right-winged political identity shows a correlation to misinformation intake (Roozenbeek et al., 2020). In agreement to Chowdhury et al. (2021), Cinelli et al. (2020) followed a similar protocol in their analysis, but in addition assessed the magnitude the content was amplified. They also analyzed characteristics of the posts such as number of shares, number of likes, and prevalence of comments on social media platforms such as Twitter, Reddit, Instagram, Youtube, and Gab (Cinelli et al., 2020). In this study, information was found by searching Covid-related terms. Youtube and Twitter had the most user interactions

(Cinelli et al., 2020). Youtube had the highest amplification rate of misinformation (Cinelli et al., 2020). They then determined if sources were questionable or reliable. A questionable source was distinguished by showing extreme bias, promoting conspiracy theories, not citing sources of credible information, and providing information that was unverifiable and had no evidence. A reliable source was distinguished by not showing any of the above questionable characteristics.

### **Twitter**

Twitter, a microblogging social media platform, has exclusively been analyzed in regard to fueling the propagation of misinformation. By analyzing trending hashtags associated with Covid-19, such as *#Coronavirus*, *#Corona*, and *#2019cov*, the magnitude of misinformation can be examined (Kouzy et al., 2020). In an analysis of 673 total tweets containing COVID-19-related hashtags, 24.8% contained misinformation and an additional 17.4% contained unverified information regarding COVID-19 (Kouzy et al., 2020). Regarding vaccines in general, a separate analysis found that of 2,580 tweets, 13% had negative viewpoints toward the vaccine, while 54% remained neutral (Love et al., 2013). Negative tweets included content involving distrust or dangers of vaccines while positive tweets promoted vaccines and supplied verifiable information (Love, et al., 2013).

## **I. INTRODUCTION TO PAPER AND PURPOSE**

### **Limitations of Prior Studies**

While studies have been done in terms of how misinformation can influence vaccine intent, there are limitations within these studies which need to be addressed. Using hashtags to search for content for an analysis fails to include any content that did not use those specific hashtags. The understanding of the COVID-19 vaccination has changed over time; therefore, using a limited time frame won't depict the change in perception. However, studying the initial rollout of the COVID-19 vaccine can allow public health experts and scientists analyze how individuals are perceiving information. Additionally, it can be beneficial in understanding and creating effective communication strategies to maximize the amount of information understood by the public.

### **Importance and Purpose**

COVID-19 misinformation is likely preventing the mitigation of COVID-19 (Xiang and Lehmann 2021). Social media platforms are permeated with misinformation resulting in users becoming coaxed (Wu et al. 2019). This is exceedingly harmful for public health and could potentially lead to further fatalities. Reducing online misinformation is imperative to reaching herd immunity given the connection between obtaining scientifically accepted information and compliance with preventative practices such as wearing masks and vaccination (Pierri et al. 2021). Analyzing how misinformation spreads can be beneficial for implementing and managing disease outbreaks and allow public health matters to be communicated with more universal

agreement. There are few strategies implemented into social media to delineate factual information from misinformation. A nudge receiving an accuracy reminder on social media has previously enhanced users' ability to distinguish factual content from misinformation by threefold (Pennycook et al., 2020). This is an effective start to reduce the continuation of misinformation. In our research, we aim to determine the comprehensive themes of information and misinformation involving the COVID-19 vaccine, the degree at which it is propagated, and to determine if COVID-19 misinformation has a relationship to vaccine sentiment in a social media environment. We will conduct our analysis using content extracted from the social media platform Twitter.

### **Research Questions**

This study aims to analyze how the spread of misinformation can be harmful for managing disease outbreaks and gaining trust among the public. In my research, we aim to determine the comprehensive topics and sentiments associated with the COVID-19 vaccine, along with the magnitude of misinformation. This study focuses on the following research questions:

- 1.) What are the major topics associated with the COVID-19 vaccine on Twitter?
- 2.) What is the overall sentiment associated with the COVID-19 vaccine?
- 3.) What is the ratio of misinformation to information on Twitter regarding the COVID-19 vaccine?

4.) How does sentiment change based on the topic of discussion and type of information present in the discussion?

### **Experimental Design**

*1) What are the major topics associated with the COVID-19 vaccine on Twitter?*

The purpose of this experiment is to determine the topics associated with misinformation and misconceptions regarding the COVID-19 vaccine observed on Twitter. To determine this, I represented the dataset as term frequency-inverse document frequency (TF-IDF) with unigrams and bigrams. I chose TF-IDF rather than a bag-of-words model because TF-IDF reflects the importance of words within a document unlike the bag-of-words model. The bag-of-words model would not be suitable in this analysis because it will only extract featured text as a unigram or bigram and does not show the relevance of the words within the document. Additionally, the bag-of-words model would create a large and sparse vector due to the fact that new words in each tweet would be viewed. Based on the TF-IDF representation, I used Latent Dirichlet Allocation (LDA) topic modelling. LDA allows underlying topics being discussed in tweets within our dataset to be revealed (Blei et al. 2003). This then allowed us to use the coherence score to determine the number of topics found (Kapadia, 2020). I predicted that there will be misinformation found in topics regarding the efficacy and side effects of the vaccine.

*2) What is the overall sentiment associated with the COVID-19 vaccine?*



The purpose of this part of my experimental design is to describe the overall sentiment on Twitter that is associated with the COVID-19 vaccine. First, I used the NRC lexicon to score our dataset for relevant emotions and sentiments. The NRC lexicon was comprised of 14,182 words that were labeled according to positive or negative sentiment and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). The NRC lexicon scores ranged between 0 and 1. A score of 0 would indicate no association to the sentiment or emotion whereas a score of 1 would indicate a strong association to the sentiment or emotion. Our study primarily focused on overall sentiment (positive and negative) rather than the specific emotion.

*3) What is the ratio of misinformation to information on Twitter regarding the COVID-19 vaccine?*

In order to determine the types of information found in our dataset, we manually coded the tweets as either information, misinformation, opinion, or question. We coded the top 100 most positive (based on association measures with positive sentiment) and top 100 most negative (based on association measures with negative sentiment) tweets associated with each topic. Researcher 1 (E.B.) coded all of the tweets, then researcher 2 (A. A.) and researcher 3 (W.R.) coded a randomly selected 100 tweets from both the positive and negative datasets. This allowed us to determine the inter-rater reliability of our codes. After coding, we were able to quantify the amount of information,

misinformation, opinions, and questions found in each topic for both the positive and negative dataset.

4) *How does sentiment change based on the topic of discussion and type of information present in the discussion?*

To determine how sentiment varied as a result of the topic of discussion and the type of information present, I analyzed the sentiment scores calculated across the eight topics of discussions and compared them across the eight topics of both sentiments. This allowed me to visualize how the sentiment levels varied as a result of the topics of discussion changing. Next, I analyzed the sentiment associated with the different types of information found in each topic in order to determine how it affected the overall sentiment and within each topic.

## **II. METHODS**

### **Data Preparation**

#### **Data Collection**

We collected tweets from December 1, 2020 to February 28, 2021. We filtered tweets by these dates because between December 1, 2020 and February 28, 2021, the Pfizer, Moderna, and Johnson and Johnson vaccines were all granted emergency use authorization by the FDA (“A timeline of Covid-19 developments”, 2021). It is important to note that while COVID-19 vaccines were beginning to be administered during this time, the first variant of COVID-19 was also established (“A timeline of Covid-19 developments”, 2021). To gain access to tweets, we used sncrape. By using sncrape, we were able to use the keyword “COVID-19 vaccine” to collect COVID-19 vaccine specific tweets. Using the keyword, we gathered 1,386,390 tweets. Following this, the dataset was cleaned in order to build a relevancy classifier. This means that duplicated tweets, punctuation marks, hashtags, @ mentions, retweets, spaces, and stop words were removed (Van den Broeck et al. 2005). Then the dataset was converted to all lowercase letters and the lemmatizing process began. Lemmatizing allowed the same words being used differently to be recognized as the same word by being converted into their base form. For example, “carrying” and “carried” would be converted as their base word, “carry”. After cleaning and lemmatizing, we were left with a dataset of 916,686 tweets.

## **Data Annotation**

With the cleaned and lemmatized dataset, we sampled 5000 tweets randomly from the entire dataset of 916,686 tweets. All three researchers manually coded 5,000 tweets as relevant or irrelevant to the COVID-19 vaccine. Relevant tweets referred to tweets that contained a knowledge statement pertaining to the COVID-19 vaccine. Any other tweet was labeled as irrelevant. We then calculated the interrater reliability between all three researchers using Cronbach's alpha. This was calculated as 0.79. Out of the 5000 tweets that were manually coded, researcher 1 (E.B.) labeled 1,122 tweets as relevant. With the relevant tweets coded by researcher 1 (E.B.), a relevancy classifier was built in order to label the remaining tweets accordingly.

## **Relevancy Classification**

Using the 5,000 tweets coded by researcher 1 (E.B.), the tweets were converted into four representations: unigram bag of words, unigram and bigram bag of words, TF-IDF with unigrams and TF-IDF with unigrams and bigrams. Each of these representations were used to train the algorithms used in machine learning. Three machine learning algorithms were used: Support Vector Classification (Linear SVC), Logistic Regression (LR), and Naïve Bayes (NB). Each of the algorithm's performances were compared. We found that TF-IDF with unigrams and bigrams and the LR algorithm had the best ability to categorize our dataset. We additionally used synthetic minority over-sampling technique (SMOTE) to train the relevancy classifier on the 1,122 relevant tweets and the remaining 3,878 irrelevant tweets. We found that SMOTE with TF-IDF

had the best performance; therefore, we used this method to categorize our 916,686-tweet dataset as either relevant or irrelevant. After this, we were left with 210,657 tweets that were relevant to the COVID-19 vaccine. We observed that about 22.4% of the 5,000 sampled tweets were relevant and about 23% of the total corpus were relevant. This similarity in distributions is evidence that the sampled tweets that trained the relevancy classifier were representative of the corpus.

### **Topic Modelling**

We used a coherence score to find the number of topics that were hidden in our dataset (Kapadia, 2020). Eight topics were found to be optimal in describing our dataset according to the coherence model. The coherence model determines the number of optimal topics by comparing the similarity of words in each group (Kapadia, 2020). Latent Dirichlet Allocation (LDA) was run on the relevant tweets using eight topics. The LDA model was fitted using Python gensim package. Each topic was named by the researcher 1 (E.B.) after looking at the top 50 words and representative tweets for each topic.

### **Sentiment and Emotion Analysis**

The sentiments and emotions were given scores by using the National Research Council (NRC) lexicon. The NRC lexicon scored positive and negative sentiments and the following emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The scoring was determined on a scale from 0 to 1. A score of 0 indicated no association whereas a score of 1 indicated a strong association to a particular sentiment or emotion.

## **Statistical Analysis**

In order to analyze how positive or negative sentiment was distributed across the topics and if the differences were significant, Kruskal-Wallis tests were done using SPSS 27.0. We chose a non-parametric test because our distribution of sentiment was non-normal, so a statistical test that does not assume normal distribution was ideal. In the entire corpus, the significance of sentiment differences was evaluated at a 95% confidence level. Individual Wilcoxon Rank Sum tests were done at a 95% confidence level which used Bonferroni-corrected p-values. These were done to determine the specific significant differences between topics and to test the null hypothesis that the distributions of two sentiments i.e., positive and negative are the same.

## **Misinformation Analysis**

To examine the misinformation and information in our dataset, we selected the top 100 positive and negative tweets per topic for which that topic had an 80 percent contribution or higher. Each tweet was manually labeled as misinformation, information, question, or opinion by the first researcher (E.B.). Tweets that contained misconceptions or left out crucial details about COVID-19 vaccine were labeled as misinformation. The tweets that included factual information about the COVID-19 vaccine were labeled as information. Tweets that did not provide any information or misinformation were categorized into either opinion or question. If a tweet was only opinionated without any information or misinformation it was labeled as opinion and tweets that asked a question were labeled as questions. Then the other two researchers (A.A. and W.R.) coded 100

tweets each that were randomly selected. Interrater reliability was calculated using Cohen's Kappa. For the positive sentiment dataset, the kappa value was 0.849 between researcher one (E.B.) and researcher two (A.A.) with a 72% agreement. The kappa value was 0.554 between researcher one (E.B.) and three (W.R.) with 72% agreement. For the negative sentiment dataset, the kappa value was 0.735 between researcher one (E.B.) and researcher two (A.A.) with 87% agreement. The kappa value was 0.664 between researcher one (E.B.) and three (W.R.) with 78% agreement.

## **IV. RESULTS**

### **Relevancy Classification**

Upon testing different algorithms for the relevancy classifiers, the accuracy fell between 79% and 86%. F1 scores were calculated to rate their performances. We found that F1 scores, which provide a measure of classifier performance that takes into account class imbalance, fell between 0.55 and 0.87. By looking at Table 1, the results from the relevancy classification show that without using SMOTE, TF-IDF with unigrams and bigrams and logistic regression was the most accurate. TF-IDF with unigrams and bigrams and logistic regression without SMOTE has an 87% accuracy and an F1 score of 0.87. Because we were aiming to have a dataset of only relevant tweets, it is additionally important to note that this algorithm had the fewest false negatives. Therefore, TF-IDF with unigrams and bigrams and logistic regression was the most fitting model to use for this study in order to classify the entire dataset as relevant or irrelevant. Upon using this classifier model, 210,657 tweets were classified as relevant. Table 2 shows examples of representative tweets for relevant and irrelevant tweets. For future topic and sentiment analyses, the relevant tweets were used.



**Table 1: Results of Relevancy Classification**


Model	Test Set				Confusion matrix			
	Acc. (%)	F-1	P	R	TP	FP	FN	TN
	Unigrams (SVC)	79.83	0.69	0.72	0.68	167	127	198
Unigrams (LR)	82.08	0.70	0.77	0.68	149	77	216	1058
Unigrams (NB)	80.04	0.71	0.72	0.70	184	141	181	994
Unigrams and bigrams (SVC)	79.97	0.69	0.71	0.68	162	120	203	1015
Unigrams and bigrams (LR)	82.71	0.70	0.79	0.67	145	64	220	1071
Unigrams and bigrams (NB)	79.34	0.70	0.71	0.70	185	152	180	983
TF-IDF with unigrams (SVC)	82.99	0.71	0.79	0.68	148	63	217	1072
TF-IDF with unigrams (LR)	80.95	0.63	0.80	0.61	97	41	268	1094
TF-IDF with unigrams (NB)	79.62	0.55	0.86	0.56	59	22	306	1113
TF-IDF with unigrams and bigrams (SVC)	82.85	0.70	0.80	0.67	144	61	221	1074
TF-IDF with unigrams and bigrams (LR)	81.31	0.62	0.83	0.61	93	32	272	1103
TF-IDF with unigrams and bigrams (NB)	81.03	0.61	0.85	0.60	85	28	280	1107
SMOTE using TF-IDF with unigrams and bigrams (LR)	86.48	0.87	0.87	0.87	104	227	109	896

Note: TP- True positives, FP- False positives, FN- False negatives, TN- True negatives, Acc.- Accuracy, P- Precision, R-Recall

Note: SVC- Support Vector Classification, LR- Logistic Regression, NB- Naïve Bayes

**Table 2: Relevant and Irrelevant Representative Tweets**

Relevant Tweet	Irrelevant Tweet
#WATCH: Mandy Cohen says COVID-19 is NOT in the vaccine. The vaccine imitates the virus to force antibodies to fight off Covid if it enters your body	Grenada scheduled to receive 45,600 Covid-19 vaccines
How mRNA Vaccines Work - Lifehacker #health #covid-19 #coronavirus	FDA issues emergency use authorization for Pfizer/BioNTech Covid-19 vaccine - CNN

<p><b>BREAKING:</b> The Moderna #coronavirus vaccine has been authorised for use in the UK.</p> <p>The jab has been developed in the US and trials have shown its 94% effective against Covid-19.</p> <p>The UK government's ordered seven million doses but it's likely to only be available from March.</p>	<p>First trucks carrying the COVID-19 vaccine in the US have left the Pfizer facility</p>
<p>Side effects: Side effects that have been reported with the Pfizer-BioNTech COVID-19 Vaccine include:</p> <ul style="list-style-type: none"> <li>• injection site pain • tiredness</li> <li>• headache</li> <li>• muscle pain</li> <li>• chills</li> <li>• joint pain</li> <li>• fever</li> <li>• injection site swelling</li> <li>• injection site redness</li> <li>• nausea</li> <li>• feeling unwell</li> </ul>	<p>@NathCe3 @OSTERElizabeth1 @raoult_didier #Canada #Vaccin.Les vaccins immunisent? Pour limiter la propagation des nouveaux variants, les soignants doivent s'isoler près d'un mois s'ils sont en contact, à leur domicile, avec une personne atteinte de la COVID-19. Même s'ils ont déjà été vaccinés..</p>
<p>How a COVID-19 Vaccine Can Destroy Your Immune System</p>	<p>Important! Please sign this!PETITION: No to government and corporate penalties for refusing COVID-19 vaccine</p>
<p>COVID-19 syringes will have microchips on outside, not in vaccine  </p>	<p>NEW: US Sen. Mike Rounds says he'll get COVID-19 vaccine.</p>
<p>COVID-19 vaccine: Do I still get it even if I've had the virus?</p>	<p>Biden says 'almost everybody' could have COVID-19 vaccines by 'the end of summer'</p>
<p>i dont know who needs to hear this but i promise you the government is not tracking you through a microchip in the covid-19 vaccine</p> <p>they can do that with the ten digit number you were given at birth if needed</p>	<p>Only 1 year ago today, the WHO gave COVID-19 its official name.</p> <p>It's incredible to think how far science has taken us: with vaccine rollouts around the world.</p> <p>It wouldn't have happened without the contributions of #womeninscience #WomenScienceDay</p>
<p>Vaccines are the greatest advancement in medical history. They have saved more lives than antibiotics or surgery. We have created one for Covid-19. The science says its safe and works great. Yet for some reason that is be-fuckin-ond me, a lot of people don't want it. </p>	<p>5% of Montana fully immunized, 7.2% of Montana received first dosage of vaccine, 9.1% of Montana confirmed COVID-19 positive #mtnews</p>

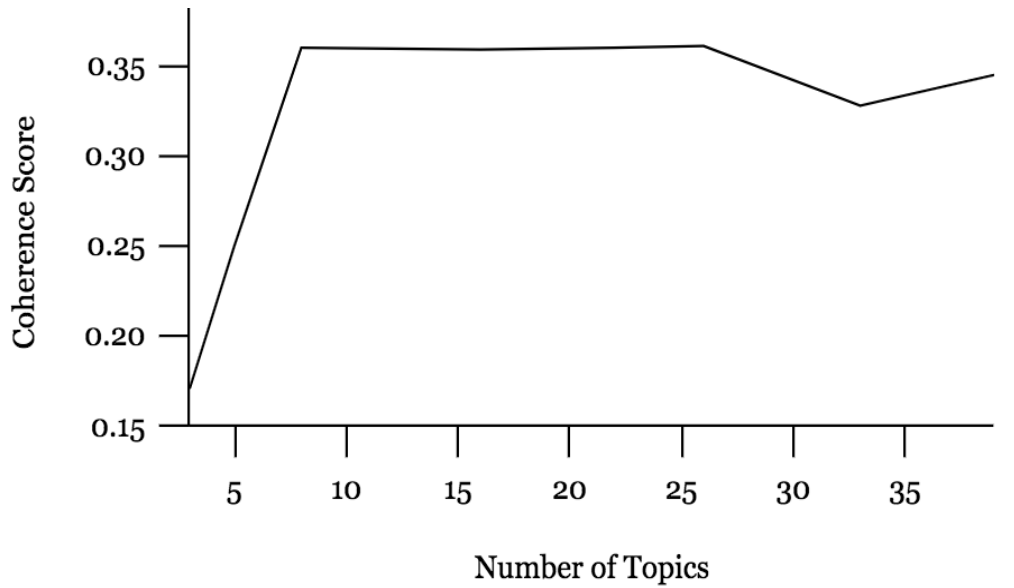
<p>“Life-threatening responses seen in at least eight people could be linked to polyethylene glycol, known to trigger reactions to some drugs” in “Suspicious grow that nanoparticles in Pfizer’s COVID-19 vaccine trigger rare allergic reactions”</p>	<p>A Disproportionate Share of COVID-19 Vaccines Are Going to White People #SmartNews</p>
---	---

### Topic Modelling

Eight topics were found within our dataset by using a coherence score. Figure 1 shows the coherence score plotted against the number of topics. Eight topics were determined because this is the point with the highest topic coherence score. When a lower or higher number of optimal topics was examined, the coherence score is less than that found in Figure 1. Because the maximum coherence score is desired, eight is the optimal number of topics.

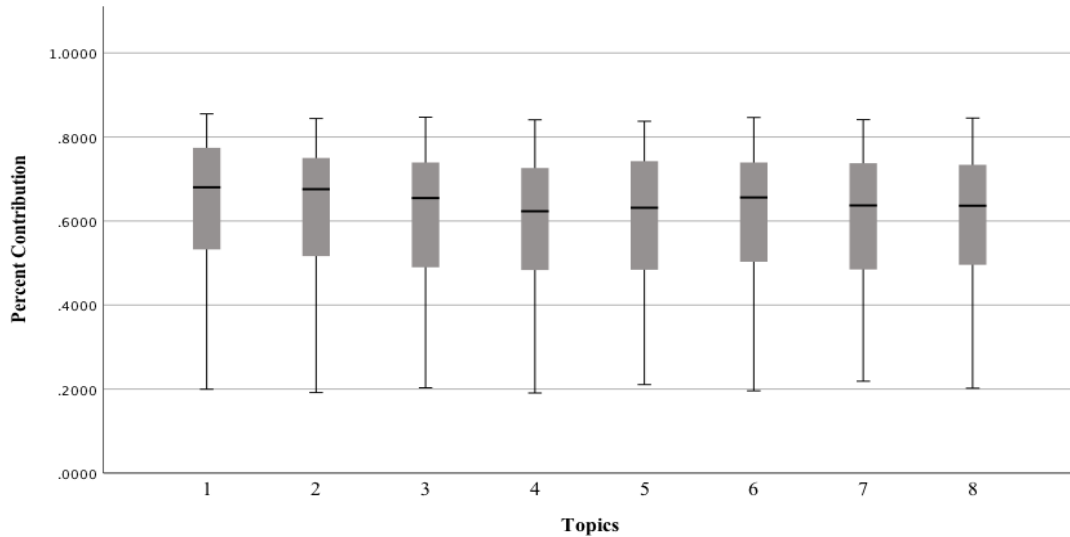
Tweets were categorized into each topic by running LDA on Python using the genism package. By analyzing representative tweets and the top words associated with each topic found, the topics were manually named. Representative tweets for each topic can be referenced in Table 3. The names for the eight topics were: *Preventative Measures and Risk Factors*, *Vaccine Trials and Efficacy*, *Vaccine Side Effects*, *Vaccine Information and Myths*, *Vaccine Safety Controversies*, *Vaccine Trial Data and Results*, *Adverse Reactions and Distrust in the Vaccine*, and *Politics and Rumors About the Vaccine*. The distribution of each topic’s percent contributions and means is shown in Figure 2.

**Figure 1. Coherence Scores Plot**



*Note: This figure shows how the coherence score varies with different numbers of optimal topics. The point of the initial spike represents the best fitting number of topics.*

**Figure 2. Topic Percent Contribution Distributions and Means of Individual Topics**



*Note: Topic 1 – Preventative Measures and Risk Factors, Topic 2 – Vaccine Trials and Efficacy, Topic 3 – Vaccine Side Effects, Topic 4 – Vaccine Information and Myths, Topic 5 – Vaccine Safety Controversies, Topic 6 – Vaccine Trial Data and Results, Topic 7 – Adverse Reactions and Distrust in the Vaccine, and Topic 8 – Politics and Rumors About Vaccine.*

**Table 3: Representative Tweets per Topic**

<b>Topic Name</b>	<b>Top 10 Words</b>	<b>Contribution Percent</b>	<b>Text Example</b>
Topic 1. Preventative Measures and Risk Factors	Get Still Mask Virus Spread Wear Prevent People Stop Protect	82.06%	I've been reading about the three Covid-19 vaccines. Because Trump made this virus get so out of control, the only responsible choice is to get whichever is available first, as soon as we can. I have some concerns but hopefully we'll be fine. More worried about the virus.

83.30%	<p>Pretty extreme ignorance and inability to look at even basic risk analysis- if a new vaccine has a .01% chance of causing side effects but getting COVID-19 has a 1% chance (lets call it 0.5% since you aren't in risk groups but 44 is not young) of killing you....  <a href="https://t.co/cBsedbGAKI">https://t.co/cBsedbGAKI</a></p>
84.14%	<p>So far it is looking very hopeful indeed that the full data for these two vaccines are highly effective at preventing from Covid-19 infection, and severe disease. Safety also looking v good w *large* trials. What we don't know still is how long immunity lasts, but trial ongoing.</p>
81.41%	<p>@Sosa04537073  @foodyumpizza  @koochykoo  @milneorchid  @piersmorgan i could say scientist all around the world would be able to differentiate the genome between the influenza virus and covid-19 virus (yes even those that don't develop vaccines :o ). But how do you know the flu is real? maybe it was bacterial pneumonia all along</p>
82.76%	<p>How can a safe vaccine be made so quickly? Scientists already began research for coronavirus vaccines during previous outbreaks caused</p>

Topic 2. Vaccine Trials and Efficacy	Variant Effective Study South New Say Find Dose Prevent Show	80.21%	<p>by related viruses such as SARS and MERS. This research provided a head start for rapid development of a COVID-19 vaccine <a href="https://t.co/I2tVMWHX3g">https://t.co/I2tVMWHX3g</a></p> <p>@SpingoC @LeightonRW07 @kateferguson4 @10DowningStreet 1. Yes, specifically flu vaccine. 2. Vaccines for communicable diseases always reduce transmission. We'll know more about reductions in morbidity once a COVID-19 vaccine is widely distributed. 3. Driving is a false equivalency as it's not communicable among the population.</p>
		81.30%	<p>Moderna vaccine applies for Emergency Use Authorization to @US_FDA - Only 11 people who received two doses of the vaccine developed COVID-19 versus 185 symptomatic cases in a placebo group (94.1% efficacy). 100% efficacy for severe disease. <a href="https://t.co/dkdcspX0us">https://t.co/dkdcspX0us</a></p>
		81.97%	<p>@JoeBiden @andrewcuomo Considering the proportion of fatalities to COVID-19 does not make a 90% or 95% success rate that great. Doubt any vulnerable Americans who get those bad 0.1% side effects from prescriptions are confident</p>

Topic 3. Vaccine Side Effects	Effect Side Dose Second First Day Get Say People Receive		in effectiveness OR safety of these vaccines for us
		83.32%	The strong early results for two leading Covid-19 vaccines hav implications that go far beyond the current pandemic:They suggest the time has come for a gene based tech that could provide new treatments for cancer,heart disease & other infectious diseases. <a href="https://t.co/D1Yj6Bkaex">https://t.co/D1Yj6Bkaex</a>
		82.32%	Miguel @migsebastiang: Today is World AIDS Day. Although there are effective treatments, there are still many victims, especially in poor countries. Scientists are still searching for a vaccine, an indication of how lucky we seem to be with COVID-19 vaccines. H/T @Marta_g_rguez <a href="https://t.co/EN0s1pG0an">https://t.co/EN0s1pG0an</a>
		81.01%	@bbc5live I can't have the Covid-19 or ANY VACCINE! I have a rare autoimmune condition, a variant of MS, which is idiopathic - cause unknown - although c.50% had their attack after a vaccination of some sort. Therefore none of us can risk it. We will need a special pass!



82.69%	<p>I wasn't allowed 2 say anything til today, but I volunteered 4 the COVID-19 vaccine trial thats been developed in Russia. My 1st dose was this morning &amp; wanted to let every1 know that its completely safe, w no side effects whatsoever and that I feel</p> <p>ÑˆĐ²Ñˆ Ñˆ% Đ½ Ñˆ,,  Đ°Đ»ĐμÑˆ%  ÑˆĐ»Ñˆ†ÑˆÑˆÑˆ% Ñˆf.</p>
82.81%	<p>Because covid-19 vaccines may cause "mild" side effects (but bad enough to make you miss a day of work), @DrPaulOffit notes that hospitals shouldn't vaccinate entire ER staff at once, or many could be out sick the next day and unable to work. #jamalive</p>
83.56%	<p>@MichaelYeadon3 Have people already died in the Covid 19 vaccine trials? (I saw somewhere that Bill Gates says he is expecting approx 700,000 deaths from the vaccine (video clip from May 2000 I think)). How dangerous is injecting mRNA compared to boosting T-cells &amp; trusting the human body?</p>
82.31%	<p>@Ravenrider2020 the PCR test in a serious study, a so-called Sanger sequencing must be used. This is the only way to make reliable statements on the effectiveness of a vaccine</p>

Topic 4. Vaccine Information and Myths

Get Fact Myth Work Misinformation Immune Explain Virus Take Protein

81.39%

against Covid-19. On the basis of the many different PCR tests of highly varying quality, neither the risk of disease/

It's been a long journey to finding the Covid-19 vaccine. Fortunately, two major companies manufacturing the vaccines say they are over 94% effective in preventing the virus. Check out the article below to see what's expected in the near future.  
<https://t.co/obHoMVayBo>

82.67%

@EU\_Commission False & criminal misinformation. The most effective defense against COVID-19 is healthy & strong immune system. Here is my little educational thread about vaccines, infectious diseases & how it all relates to our biology, health and human immune system:  
<https://t.co/QME0IE6yX5>

82.44%

@katy\_milkman on COVID, Vaccines & BI:  
Social proof is a great way to ensure that people act similar to their peers: "after getting jabbed, people should be encouraged to text their friends, post on social media & sport got the covid-19 vaccine stickers"

Topic 5. Vaccine Safety Controversies	Safe Get People Take Hall Know Community Town Say Virus		#behavioraleconomics <a href="https://t.co/04dDKApq3N">https://t.co/04dDKApq3N</a>
		82.52%	@maddow further to this, it is not explained in any media reporting. If you are infected with active COVID-19 you will be unable to get take vaccine. The third stage testing includes tests where those after boosted immunity kicks are exposed to a virus load.
		82.37%	'Covid_19 vaccine contains a spike protein called syncytin-1 (HERV-W) vital for the formation of human placenta in women. If the vaccine works so that we form an immune response AGAINST the spike protein, we are also training the female body to attack syncytin-1, which '
		80.05%	@mtmalinen Guess what else is relatively untested? Long term health effects of getting COVID-19.  But, if there are side effects to the vaccine, you wont infect other ppl with them. Not true of the virus.
		82.21%	@DoctorChristian I understand that the covid 19 research has been heavily funded. Apparently under usual circumstances, vaccine research is grant funded which is a lengthy process

Topic 6. Vaccine  
Trial Data and  
Results

Question  
Answer  
Trial  
Efficacy  
Phase

		<p>&amp; can be extremely time <a href="https://t.co/jld1bK7pa7">https://t.co/jld1bK7pa7</a> together with expertise &amp; commitment has produced rapid results.</p>
82.38%	<p>Stop The COVID-19 vaccine in the United States, too! It has not been adequately tested. Its a new vaccine type that needs many more trials. Unknown Long term side effects. So many risks.. Follow the big Pharma  <a href="https://t.co/ZsfdRf1quj">https://t.co/ZsfdRf1quj</a></p>	
80.83%	<p>No, Covid-19 won't be over in 2021.</p> <p>It will take AT LEAST a year to produce enough vaccines for the whole population, to convince enough people to TAKE the vaccine (humans are idiots and assholes), and to build up herd immunity.</p>	
80.53%	<p>Bit ironic "America First" administration hired a foreigner to save their nation, eh? &gt; Deliver a safe, effective COVID-19 vaccine in less than a year? Impossible. Meet Moncef Slaoui.  <a href="https://t.co/3V4DcAVpir">https://t.co/3V4DcAVpir</a>  #SmartNews</p>	
80.14%	<p>Oh, the attachment is New Jersey news article from this morning re: European Medicines Agency will convene a meeting on</p>	

<p>Information Effective Expert Safety Join</p>		<p>Dec. 29 to decide if there is enough data about safety &amp; efficacy of COVID-19 vaccine developed by Pfizer &amp; BioNTech for it to be approved, regulator said Tues.</p>
	<p>83.00%</p>	<p>Deliver a safe, effective COVID-19 vaccine in less than a year? Impossible. Meet Moncef Slaoui. As science adviser of Operation Warp Speed, the veteran pharmaceutical executive helped usher a coronavirus vaccine to the US market in record time. <a href="https://t.co/GwXZzpnHrt">https://t.co/GwXZzpnHrt</a></p>
	<p>82.44%</p>	<p>@sc_g_cox_70_33 @SCEMD Currently, clinical trials are evaluating investigational COVID-19 vaccines in many thousands of study participants to generate scientific data and other information for the FDA to determine their safety and effectiveness. Learn more at <a href="https://t.co/os2XhcRojM">https://t.co/os2XhcRojM</a></p>
	<p>81.99%</p>	<p>Were working closely with experts in logistics, technology, patient care, pharmaceuticals and business to make sure we can get the COVID-19 vaccine into communities across Ontario. As General Hillier said, well be ready by December 31st.</p> <p>Find out more:</p>

Topic 7. Adverse Reactions and Distrust in Vaccine	Reaction Report Woman Adverse Pregnant Health Allergic Severe Death Die	81.82%	<a href="https://t.co/EFzVCaLV6w">https://t.co/EFzVCaLV6w</a> <a href="https://t.co/Q04aQCGnhF">https://t.co/Q04aQCGnhF</a> @NatureMedicine The length of immunity following Covid-19 vaccination is also unknown. Large, carefully designed post-marketing studies will be essential to track vaccine failures (infections following vaccination) and adverse events. <a href="https://t.co/pEAMQYLNji">https://t.co/pEAMQYLNji</a>
		81.90%	Some health care workers have voiced concerns about safety and potential side effects from COVID-19 vaccines. They want to see clear data on safety and efficacy before they sign on to get a new vaccine, Mahon says. <a href="https://t.co/YCM0LtOxQE">https://t.co/YCM0LtOxQE</a>
		81.72%	Due to distrust of, and systemic racism in, the healthcare system, "only 32% of Black adults say they would definitely or probably take a COVID-19 vaccine" according to Pew Research Center. In CA the percentage is 30%. <a href="https://t.co/whNaOeQGm3">https://t.co/whNaOeQGm3</a>
		83.19%	@TurnerKebab The UK government KNOW this is going to harm people. The MHRA have put out a £1.5m tender: "The MHRA urgently seeks an Artificial Intelligence (AI)

Topic 8. Politics and Rumors About Vaccine	Know Take Die Get People Test Here News Positive Need		software tool to process the expected high volume of Covid-19 vaccine Adverse Drug Reaction (ADRs)..."
			<a href="https://t.co/KFj9CueLDF">https://t.co/KFj9CueLDF</a>
		82.86%	@pjakma @florian_krammer I disagree. You need to look at this from a risk perspective. And the overall risk profile of COVID-19 is much worse than the vaccine even for younger and healthier adults than the ones in the risk group. After all the disease can result in death and long-term effects.
		82.04%	@JosefKalfsGran The Gov are expecting a high number of adverse reactions as we can see in the link below - ' The MHRA urgently seeks an Artificial Intelligence (AI) software tool to process the expected high volume of Covid-19 vaccine Adverse Drug Reaction (ADRs) ' <a href="https://t.co/cPxHtHPrbi">https://t.co/cPxHtHPrbi</a>
		80.88%	@simontoleman @markste76948257 Testing people for antibodies wouldnt potentially stop people who have covid 19 from dying? Also then you really are talking about forcing people into testing? As far as I can gather no one is forced into taking the vaccine?

		Im sure people do care about flu..
	81.45%	@KathyConWom Hancock said govt "looking at the technology" so people can show they had a Covid-19 vaccine: "You'll probably find that restaurants bars cinemas & other venues will also use that system - as they have done with the test and trace app." <a href="https://t.co/m7b7rJyUmm">https://t.co/m7b7rJyUmm</a>
	83.16%	@BurroughsWarren @VBermoothes 75,000 people died from COVID 19 at the moment, government knocked off the death toll figures to save face, and the vaccine the only hope we got getting out of lock down for good, and masks do stop the spread of transmission to people near you, you covid deniers going to kill a
	82.79%	Thinking about the COVID vaccine, I still have my qualms. However, people forget that we still do not know the long term affects COVID-19 can have, especially neurologically as it crosses the blood brain barrier. Things to keep in mind when questioning the s/e of the vaccine.
	81.24%	@IAmTomHall Maybe because the flu can be several different viruses and the general vaccine only



		works against the most dangerous varieties, and in the USA only 30-40% of people get this vaccine. COVID-19 is one single virus, and has a lot more attention currently.
--	--	--

*Note: All representative tweets shown have at least an 80 percent contribution.*

### Sentiment Analysis

The NRC lexicon scored tweets in each topic which identified their sentiment as either positive or negative. Representative tweets of positive and negative sentiment for each topic are shown in Table 4 along with the emotion given by NRC lexicon.

**Table 4. Representative Tweets of Positive and Negative Sentiment for the Eight Topics Selected to Categorize the Subject of Tweets**

Topics	Top 5 Positive Tweets	Emotion	Top 5 Negative Tweets	Emotion
<b>Preventative Measures and Risk Factors</b>	<p>Childrens survival depends on dedicated health workers being able to work safely in the community day after day. Thats why initial doses of COVID-19 vaccines must be prioritized for nurses, doctors and other essential workers. Our health depends on their health.</p> <p>#VaccinesWork</p>	Trust	<p>@eatsruns The vaccine for corona viruses has been in the works for years. There was SARS and MERS before COVID 19. This virus affected so many more ppl than those viruses so thats why the vaccine came to light.</p> <p>Its been in the making for longer than most folks know.</p>	Trust

<p>QUESTION: Health experts report some patients have had similar side effects that compare to COVID-19 symptoms after taking the vaccine, however can be managed. Would you still take the vaccine when widely available? Let me know your thoughts!</p>	<p>Anticipation</p>	<p>You say "There is NO science", How about you tell the 1.67Million people that are dead and their loved ones that there's no science to back up their deaths from #Covid_19. Last I checked you need science to find a virus and to make a vaccine</p>	<p>Anger</p>
<p>Ppl keep talking about how these vaccines have been rushed as if Covid-19 hasn't been around for a year and there aren't similar diseases for scientists to learn from. Countries outside the West have already started vaccinating their popuations. Ppl have been doing the work.</p>	<p>Trust</p>	<p>Today I chose to receive the COVID-19 vaccine. I will still be social distancing, masking, and taking all measures to slow the spread of the virus. Second vaccine dose in 3 weeks!</p>	<p>Trust</p>
<p>The search for a COVID-19 vaccine stands out from other developments because it is the first one the ENTIRE science community on Earth has come together to develop...and this is why the vaccine is being developed at a faster rate than other vaccines.</p>	<p>Trust</p>	<p>@DrOz Yeah but we dont know what the side effects are over the years. It might be the vaccine who kills more people then Covid-19 as far as we know. Plus its not a vaccine if we still carry the virus without symptoms, others around us would still get infected.</p>	<p>Fear</p>
<p>Dr. Moncef Slaoui, the head of the U.S. vaccine development effort, said Sunday he believes the COVID-</p>	<p>Trust</p>	<p>Experts remain unclear about whether someone who gets vaccine is still able to contract &amp; pass</p>	<p>Anger</p>

	19 vaccine effectiveness could last for many, many years, with people who are more vulnerable requiring a booster every three to five years.		on CCP Virus? No way! #NotYourGuineaPig	
<b>Vaccine Trials and Efficacy</b>	@touchmywoowoo @DJSiri "Moderna's latest findings, according to a company press release, showed that of 196 people in the clinical trial who caught COVID-19, 185 of them had received the placebo, while only 11 received the active vaccine. That works out to an effectiveness rate above 94%."	Trust	People With Certain Allergies Shouldnt Get COVID-19 Vaccine @inspectorgajicc @ktrrrrrrose @Double17Darts @JugglingFitness @TorontoStar The studies and basis of this vaccine was based off of the 2005 coronavirus as it is said to be the most similar mutation of the virus to covid-19. Its not as much that they have only been studying the vaccine for 9 months, the data they began with is from the virus 15 years ago	Trust
	Moderna announced the final results of its 30,000-person efficacy trial for its vaccine: Only 11 people who received two doses of the vaccine developed COVID-19 symptoms after being infected with the SARS-Cov-2 vs.185 symptomatic cases in a placebo. That is an efficacy of 94.1%.	Trust	The new strain of #Covid_19 virus has raised questions about the effectiveness of all vaccines that are being applied on human for mass vaccination. Seems the vaccine don't go in vein. #secondwave #mutation #coronavirus	Anticipation
	7/12 If a PEG-containing mRNA vaccine for Covid-19	Anger	Lowcountry doctor says the new mutation to COVID-19 is a	Trust

gains FDA approval, the uptick in exposure to PEG will be unprecedented and potentially disastrous.

matter of one or two of the proteins within the virus, and in order for the vaccine to be ineffective, the virus would likely need to mutate significantly more

Moderna documents and publications indicate that the company is well aware of safety risks associated with...

<p>UPDATE: Pfizer, BioNTech Report 'Data from 43,448 participants, half of whom received BNT162b2 and half of whom received placebo, showed that the vaccine candidate was well tolerated and demonstrated 95% efficacy in preventing COVID-19...' \$BNTX \$PFE</p>	<p>Trust</p>	<p>@realDonaldTrump As COVID-19 cases rise, baseless claims about vaccines and the virus have spread online, including a doctor's photo that was misrepresented on social media to falsely claim the pandemic is fake.</p>	<p>Anger</p>	
<p>@steve67spd @AlexBerenson @CDCgov Since you like numbers so well, read these results from the Pfizer trial of over 30k people that shows getting the vaccine reduces severe cases of covid-19 by less than 1% compared to literally doing nothing.</p>	<p>Anger</p>	<p>Coronavirus pandemic has infected more than 68.9 million people. Here's the latest on global Covid-19 crisis:  Canada approves Pfizer vaccine Turkey reports over 31,000 new infections Italy reports fewer virus deaths, infections</p>	<p>Anticipation</p>	
<p>'Doing nothing', you'll note, has zero side effects.</p>				
<p><b>Vaccine Side Effects</b></p>	<p>In Year 1 we have been very busy</p>	<p>Joy</p>	<p>Sinovac and Sinopharm, Chinas</p>	<p>Trust</p>

<p>scientists. We haven't developed a vaccine for COVID 19, but we do know how to look after ourselves to stay healthy. The children sorted pictures into different categories of staying healthy and wrote about their findings.</p>		<p>vaccine makers, are testing three vaccine candidates in phase III trials that could be distributed without cold storage. They uses inactivated forms of COVID-19 to induce immune responses.</p>	
<p>COVID-19 VACCINE: With potentially just WEEKS until a COVID vaccine hits hospitals, doctors are getting ready.</p> <p>Once it's available to the community, doctors say things could look similar what we've seen with mass COVID testing sites.</p>	<p>Anticipation</p>	<p>Covid-19 vaccine.... Developed in less than a year!!!! There is something wrong here..think about it... don't forget that for the first time the drug companies have been given immunity to prosecution for whatever the side effects will be no matter how serious!!</p>	<p>Anger</p>
<p>Covaxin clinical trials are based on a 2-dose schedule, given 28 days apart. Vaccine efficacy will b determined 14 days post 2nd dose. It has been designed to b efficacious when subjects receive both doses-Bharat Biotech on Haryana Anil Viz getting COVID-19 positive</p>	<p>Trust</p>	<p>The side effects of the COVID-19 vaccine wont manifest immediately.</p> <p>Dont you know what happens when the human body is exposed to radiation?! It modifies the DNA. Therefore it takes years to see/feel the side effects.</p> <p>Where do you think cancers come from?</p> <p>All about the DNA.</p>	<p>Disgust</p>

	<p>Test humans wanted after vaccine</p> <p>Study to Describe the Safety, Tolerability, Immunogenicity, and Efficacy of RNA Vaccine Candidates Against COVID-19 in Healthy Individuals - No Study Results Posted -  <a href="https://t.co/Z5Lge2dXhp">https://t.co/Z5Lge2dXhp</a>  <a href="https://t.co/c137g2uF76">https://t.co/c137g2uF76</a></p>	Anticipation	<p>"People who get a COVID-19 vaccine should be monitored for at least 15 minutes after getting vaccinated, according to the CDC."</p> <p>"side effects include tiredness, headache, muscle pain, and chills. The agency said they go away after several days"</p>	Trust
	<p>I think [it] is going to be about a little less than 50 percent of people are willing to get a vaccine at this point."</p> <p>Theyre predicting well need around 70 percent of the population to get vaccinated in order to completely conquer COVID-19</p>	Anticipation	<p>Sounds like the China Virus vaccine gives you the China Virus Each vaccine is slightly different but, in general, experts say you may experience the following side effects with either COVID-19 vaccine:</p> <p>Fever  Fatigue  Headache  Joint pain  Muscle aches</p>	Anticipation
<b>Vaccine Information and Myths</b>	<p>@asplint  @Melissa2664  @uche_blackstock  @lizditz I understand that the vaccine will be safer than COVID-19 and I think a lot of people that still have some concerns and are asking questions do. We still want to better understand what risks might</p>	Anticipation	<p>I don't know who needs to hear this but here you go.</p> <p>mRNA vaccines do not use the live virus that causes COVID-19.</p> <p>mRNA never enters the nucleus of the cell, which is where our</p>	Trust

come with the vaccine, especially since it utilizes newer technology.		DNA (genetic material) is kept.	
<p>@JettGoldsmith @FrogNews @ABC</p> <p>Dr. Anthony Fauci explained in an interview that the actual goal of the early COVID-19 vaccines isnt even to block infections, a detail that people should be aware of even before considering whether to get a vaccine or which vaccine to go for. "</p>	Anticipation	the following two are the newest types: 5. mRNA vaccines (e.g., zika virus under development, Moderna & Pfizer's COVID-19 vaccine) directly code for virus antigenic proteins and stimulates B-cell and T-cell response (which, that double response is what we want!)	Trust
<p>@Purplemoonz @Holbornlolz</p> <p>Possible reasons for this are explained in this article. In short, it relates to a developed immunity to the vector in which the RNA is delivered to the body. An interesting read.</p> <p>The Economist   Another covid-19 vaccine joins the party</p>	Trust	The quick development of vaccines for Covid-19 was made possible by decades of research on the virus that causes AIDS. Everything we do with every other pathogen spins off of things weve learned with HIV, says Dr. Anthony Fauci.	Trust
<p>@SteveFDA Dr. Anthony Fauci explained in an interview that the actual goal of the early COVID-19 vaccines isnt even to block infections, a detail that people</p>	Anticipation	@maddow further to this, it is not explained in any media reporting. If you are infected with active COVID-19 you will be unable to get take vaccine. The third stage testing includes	Trust

	should be aware of even before considering...		tests where those after boosted immunity kicks are exposed to a virus load.	
	If you are going to refer to the Tuskegee Study as the reason you will not take the Covid-19 vaccine, do your research. The subjects of the study NEVER received the vaccine. Stop relying on social media to educate you.	Anger	They make it very clear that the Covid-19 vaccines (at least, the ones due to be licensed in the UK in the near future) are NOT attenuated live virus vaccines. They cannot reproduce or replicate in the body, so they cannot cause an infection. 9/11	Trust
<b>Vaccine Safety Controversies</b>	I have a question for all the so called Christians out there. There is evidence of human aborted tissue within the development of the vaccine that will be injected into your body for the treatment of Covid 19, my question do you still feel strongly about abortions?	Anger	@realDonaldTrump I refuse to put aborted fetus cells in my body or my kids.  Thoughts? Woman Shows COVID-19 Vaccine Has Recombinant DNA & Lung Tissues From An Aborted Fetus! "It's In Front Of Your Face, Do Your Research" - <a href="https://t.co/1zkqYVr81e">https://t.co/1zkqYVr81e</a> via @Worldstar #WSHH #WORLDSTAR	Disgust
	All the information and misinformation around the upcoming covid-19 vaccine has me thinking about one day in the last class I ever took in college, and I think it's one of those experiences others could benefit	Joy	@UWMolES @PrecisionNano Why RNA vaccines for Covid-19 raced to the front of the pack  Developing and testing a new vaccine typically takes at least 12 to 20 months.	Anticipation



from, so here we go to thread town!

However, just over 10 months after the genetic sequence of the SARS-CoV-2 virus virus was published, two pharma @Patrick39949153 @SJAMcBride Why RNA vaccines for Covid-19 raced to the front of the pack

<p>Betsey Tilson says she recognizes some people think development of COVID-19 vaccine has been rushed. But Tilson says there were no safety shortcuts even as it was quickly developed. She says developed not due to compromises in safety &amp; efficacy. #nced #coronavirus</p>	<p>Trust</p>	<p>Developing and testing a new vaccine typically takes at least 12 to 17 months. However, just over 10 months after the genetic sequence of the SARS-CoV-2 virus virus was published, two pharma Co</p>	<p>Anticipation</p>
<p>Nor can we assume that just because the vaccine is available that vulnerable groups will come forward. There is already some research which suggests that there is Covid-19 vaccine hesitancy (different from anti-vaxxers) among some BME communities. This needs to be addressed.</p>	<p>Anticipation</p>	<p>@suhaib1983BB @DrSusanNasif Explained: Why RNA vaccines for Covid-19 raced to the front of the pack  Developing and testing a new vaccine typically takes at least 12 to 20 months. However, just over 10 months after the genetic sequence of the SARS-CoV-2 virus</p>	<p>Anticipation</p>
<p>JOKING COVID: Is getting "natural immunity" w/ help from your own blood's DNA better than a vaccine? My</p>	<p>Anticipation</p>	<p>Why RNA Vaccines for COVID-19 Raced to the Front of the Pack - Developing and testing a new vaccine typically takes at least</p>	<p>Anticipation</p>

	answer is a big YES. Why? Covid-19 has been mutating/transforming itself from Covid-19 into Covid-20 during Trump's last 9 mons. You've to take new stronger vaccine next year		12 to 18 months. However, just over 10 months after the genetic sequence of the SARS-CoV-2 virus was published, two...	
<b>Vaccine Trial Data and Results</b>	The European Medicines Agency will convene a meeting on December 29 to decide if there is enough data about the safety and efficacy of the Covid-19 vaccine developed by Pfizer and BioNTech for it to be approved, the regulator said Tuesday.	Anticipation	@Ravencroft82 @Christo49707680 @minimoss4 @ginacarano No, I said they didn't do things with every data they had about other corona researches because other corona viruses weren't as dangerous as covid 19. But they used them now so that is why they created the vaccine faster according to the presentation I saw about the topic.	Trust
	Vaccine developers who have already reported promising phase III trial results against COVID-19 estimate that, between them, they can make sufficient doses for more than one-third of the worlds population by the end of 2021.	Anticipation	The FDA has approved Modernas COVID-19 vaccine. Thats the second one approved by the FDA to fight the virus. Many of you have questions as these vaccines are rolling out, so were going to the experts to find the answers. Below is a thread of questions weve answered so far	Trust
	The European Medicines Agency will convene a	Anticipation	Nanotechnology: Part of COVID-19 vaccines but potential	Anticipation

meeting on Dec. 29 to decide if there is enough data about the safety and efficacy of the COVID-19 vaccine developed by Pfizer and BioNTech for it to be approved, the regulator said Tuesday.

still hindered in #Europe  
Nanotechnology has contributed to the production of vaccines against the COVID-19 virus. However, in Europe, a de-centralised and vague regulatory framework prevents ...  
<https://t.co/HoIs16kg95> #EU

	<p>Dr Amoth: Kenya is a member of the global vaccine initiative! AstraZeneca-Oxford COVID-19 vaccine may probably work best for us! We are in discussions with other countries! Any other potential vaccine will be analysed based on safety, immunogenicity, efficacy and economics</p>	<p>Anticipation</p>	<p>@MrRaid @NedaaAlkhamis @Engr_Salwa Deisher questioned the lack of transparency shared by COVID-19 vaccine manufacturers at the various stages of production, raising concerns over the data and the reports of negative side effects in test groups. Vaccine companies not releasing the data</p>	<p>Fear</p>
	<p>This is the second COVID-19 vaccine to complete a Phase 3 clinical trial, which includes at least 30,000 people. It will join the Pfizer vaccine in its review by FDA scientists, which may take a few weeks.</p> <p>The Moderna application was based on the results of its clinical trial:</p>	<p>Anticipation</p>	<p>I'm skeptical about any and all COVID-19 vaccines because there hasn't been nearly enough time and test trials to know of any possible detrimental and/or long-term side effects. But I'm an essential employee as a 911 calltaker/dispatcher. If given the option, I'm not taking it.</p>	<p>Anticipation</p>
<p><b>Adverse Reactions</b></p>	<p>Some health care workers have voiced</p>	<p>Anticipation</p>	<p>The US government is hoping to give a</p>	<p>Anticipation</p>

<b>and Distrust in Vaccine</b>	<p>concerns about safety and potential side effects from COVID-19 vaccines. They want to see clear data on safety and efficacy before they sign on to get a new vaccine, Mahon says.</p>		<p>#COVID-19 #vaccine to 100 million people by the end of February. The number represents all the nations #frontline #health workers, the #elderly, and people with underlying conditions. Vaccines will be from Moderna and Pfizer/BioNTech.</p>	
	<p>Well, Kylee:</p> <p>-2,380 Americans under the age of 35 have died from COVID-19.</p> <p>-Post-COVID Syndrome can cause strokes, heart problems, and other serious injuries in all ages</p> <p>-We *can* know the long-term effects of a vaccine. We've used this vaccine technology for years/decades.</p>	<p>Trust</p>	<p>@richardzussman [NEW POST] COVID-19 Vaccine Warnings Women Should Not Get Pregnant for at least 2 months after Vaccination - The government has said that women should avoid becoming pregnant until at least two months after being vaccinated against Covid-19.</p>	<p>Fear</p>
	<p>U.S. health officials closely tracking possible side effects of the first authorized COVID-19 vaccine say they have seen six cases of severe allergic reaction.</p>	<p>Anticipation</p>	<p>The first American to get an severe allergic reaction, at home recovering fine now</p> <p>The positives outweigh the negative so far..</p> <p>Alaska health care worker had severe allergic reaction to</p>	<p>Trust</p>

			Pfizer Covid-19 vaccine..	
	U.S. health officials closely tracking possible side effects of the first authorized COVID-19 vaccine say they have seen six cases of severe allergic reaction out of more than a quarter-million shots given.	Anticipation	We are hearing about people experiencing allergic reactions to the covid-19 vaccine. What do you think is happening to babies whose parents report huge changes after getting their vaccines (which is denied by doctors)?	Fear
	"Do not administer the Moderna COVID-19 Vaccine to individuals with a known history of severe allergic reaction (e.g., anaphylaxis) to any component of the Moderna COVID-19 Vaccine."  Moderna press release makes very interesting reading.  More to follow.	Anticipation	According to a Pew research study done in mid-November, among Black Americans, 71% know someone who has been hospitalized or died because of COVID-19, but fewer than half of Black adults polled said they would take a vaccine if it were available.	Sadness
<b>Politics and Rumors About Vaccine</b>	The world's been waiting for a vaccine " almost since we first heard of COVID-19. When and where can you get yours? While much is still being worked out, we're dedicated to sharing as much information as we can as quickly as possible.	Anticipation	THE COVID 19 IS A FLU VIRUS WITH A 99.9996 SURVIVAL RATE. HOW CAN YOU HAVE A VACCINE FOR A FAKE VIRUS UNLESS SOMETHING ELSE IS ACTUALLY HAPPENING. MAYBE IT IS THE HIV VIRUS INSIDE EACH VACCINE THAT WILL CAUSE	Anger

AIDS AND DEPOPULATION BY 95% OF THIS WORLD?

<p>As you know dear people I am quite a calm person but anti vaxers really can fuck off! "If 99% of people recover from COVID 19 why do we need a vaccine?" one percent of the UK population is approx 678,860 people - exactly how many of those people are they saying don't matter?</p>	<p>Anger</p>	<p>@JackPosobiec Let's skip over the dangers of that type of vaccine and move to why it wouldn't work for COVID-19: The antibodies don't last. I know two people who have had it twice. Tested positive, recovered, tested negative, and a few months later tested positive again.</p>	<p>Anger</p>
<p>14. Anyone who after reading all this still wants to get injected with the mRNA vaccine, should at the very least have their blood checked for COVID-19 antibodies. There is no need for a vaccine in persons already naturally immunized.</p>	<p>Anticipation</p>	<p>Peolle think it's weird that we have a vaccine for COVID-19 already saying "how is that possible when they say they've never seen this virus before?" How many fucking doctors and researchers are in the world? You know the internet is a thing..right?</p>	<p>Anger</p>
<p>Whats the point in giving 1m people 1 dose each (when they require 2 to protect them?) surely better to give 500k 2 doses each and protect 500k of the most vulnerable! COVID-19: Tony Blair suggests using up UK vaccine stock on first doses</p>	<p>Anticipation</p>	<p>But how did they come up with a vaccine so fast if it takes years for others! COVID-19 is a new strain of a already known virus, and thats what many people forget. These people have gathered data on this shit already for years.</p>	<p>Anger</p>

Mencap have a variety of easy read guides relating to COVID-19 for people with disabilities. Topics include vaccines, social distancing and self isolating. See their website here. #safeguarding #learningdisabilities @hantsconnect	Trust	@Diana_Damifino @miltsdad No Covid-19 is not the even close to the pandemic of 1918 if you ppl knew how to do Math you'd know that! vaccines for a virus that kills less then 1%. and isn't tested 4-7 years like its suppose to, what more could go wrong..... side effects like the ones that harmed people	Anger
---	-------	---	-------

*Note: The top 5 tweets for each sentiment were chosen because they had the highest percent contribution for the given topic.*

### Statistical Analyses

Using the results from the Kruskal-Wallis tests (Table 5), we observed that there are statistically significant differences in positive ( $H(7)=10707.3$ ,  $p < 0.001$ ) and negative ( $H(7)=11735.8$ ,  $p < 0.001$ ) sentiments across the 8 topics.

**Table 5. Kruskal-Wallis Test Results in Positive and Negative Sentiment**

	df	H test statistic	P-value	Median
Positive	7	10707.3	< 0.001	0.182
Negative	7	11735.8	< 0.001	0.083

*Note: 0.05 significance level and a 95% confidence interval*

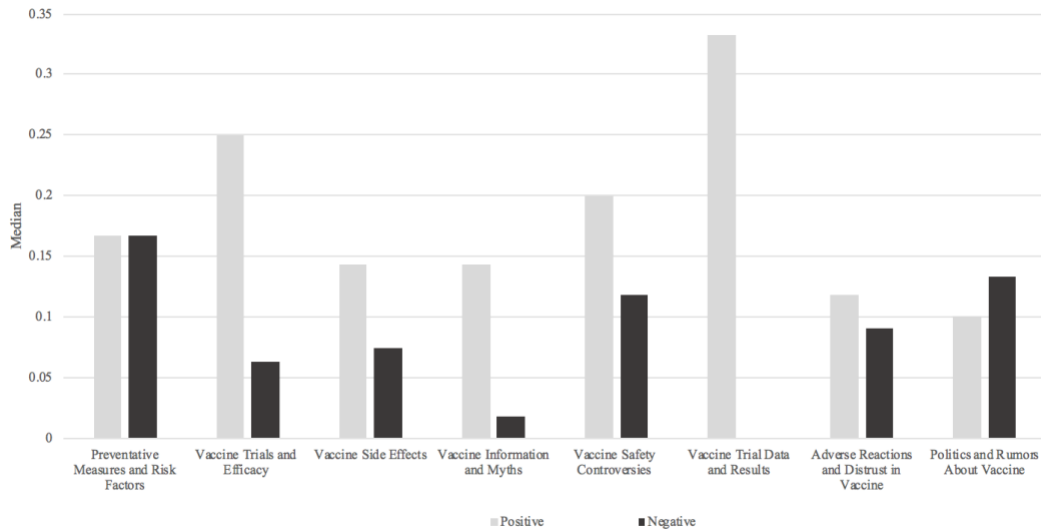
Post-hoc Mann-Whitney tests using Bonferroni-adjusted p-values showed that the topics had statistically significant differences in the amount of positive sentiment. The

only exception to this was observed in Topic 3 (Vaccine Side Effects) and Topic 4 (Vaccine Information and Myths). These two topics were found to be statistically similar in terms of their sentiment association with an adjusted p-value of 0.896. For negative sentiment, Topic 2 (Vaccine Trials and Efficacy), Topic 3 (Vaccine Side Effects), and Topic 7 (Adverse Reactions and Distrust in Vaccine) had statistically similar sentiment with adjusted p-values ranging from 0.45-1.00. Topic 7 (Adverse Reactions and Distrust in Vaccine) was the most statistically similar to Topic 4 (Vaccine Information and Myths) with an adjusted p-value of 0.072. All other differences were statistically significant.

Tweets with positive sentiment (Median=0.182) had a higher average association than those with negative sentiment (Median=0.083). Figure 3 shows the median values across the topics in both sentiments. *Vaccine Trial Data and Results* had the highest association with positive sentiment, with a median value of 0.33 while also having a median value of 0.00 for negative sentiment. *Politics and Rumors about the Vaccine* was the only topic where the negative value was larger than that in the positive sentiment. In both positive and negative, the median association score was equal (0.167) in the *Preventative Measures and Risk Factors* topic, suggesting equal prevalence. In all other topics, the positive value was greater than the negative value.



**Figure 3. Median Scores of Positive and Negative Sentiments Compared Among 8 Topics**



The top 25 words associated with each sentiment for each topic is shown in the word clouds in Figure 4. Consistency among topics was observed in both positive and negative sentiments. For example, words such as dose, efficacy, and study were positive words among all topics while words such as government, death, and virus were negative among all topics. By examining the word cloud associated with *Preventative Measures and Risk Factors*, “effective”, “protect”, and “efficacy” are all words highly associated with positive sentiment inferring that the discussion is primarily focused on how effective the COVID-19 vaccine is. Words such as “mask”, “wear”, “avoid” were commonly associated with negative sentiment inferring that discussion is regarding preventative measures that public health professionals have advised individuals to take. By examining

the word cloud associated with *Vaccine Trials and Efficacy*, “safe”, “effective”, and “candidate” were all words associated with positive sentiment. This suggests that discussion is based around the vaccine trials. Words associated with negative sentiment in this topic included “infection”, “death”, and “infectious” suggesting that discussion is focused on the virus and pandemic. By examining the word cloud associated with *Vaccine Side Effects*, “safe”, “effective”, and “dose” were all words associated with positive sentiment, suggesting that discussion is based around the safety and effectiveness of the vaccines. Words associated with negative sentiment in this topic included “pain”, “symptom”, and “sore”. This suggests that discussion is focused on the side effects from the vaccine. By examining the word cloud associated with *Vaccine Information and Myths*, “explain”, “expert”, and “fact” were all words associated with positive sentiment suggesting that discussion is based around debunking myths and educating individuals on the vaccine. Words associated with negative sentiment in this topic included “false”, “severe”, and “worsen”. This suggests that discussion is focused on the continual misinformation observed on social media platforms. By examining the word cloud associated with *Vaccine Safety Controversies*, “question”, “scientist”, and “doctor” were all words associated with positive sentiment. This shows that Twitter has health experts available to spread valuable information about the COVID-19 vaccine. Words associated with negative sentiment in this topic included “government”, “dangerous”, and “refuse”. This suggests that discussion is focused on questioning how trustworthy the government is in the process of the COVID-19 vaccination process. By examining the word cloud

associated with *Vaccine Trial Data and Results*, “efficacy”, “phase”, and “trial” were all words associated with positive sentiment. This suggests that discussion is based around the COVID-19 vaccine trials. Words associated with negative sentiment in this topic included “risk”, “government”, and “warn”. This suggests that discussion is focused on the validity of the COVID-19 vaccine. By examining the word cloud associated with *Adverse Reactions and Distrust in the Vaccine*, “ingredient”, “allergic”, and “effective” were all words associated with positive sentiment. This shows that positive sentiment discussions are involving all the effects of the vaccine. Words associated with negative sentiment in this topic included “adverse”, “reaction”, and “refuse”. This suggests that discussion is focused on the potential negative effects associated with the vaccine. By examining the word cloud associated with *Politics and Rumors About the Vaccine*, “good”, “truth”, and “news” were all words associated with positive sentiment. This suggests that discussion shows trust in the vaccine. Words associated with negative sentiment in this topic included “hoax”, “worry”, and “fake”. This suggests that discussion is focused on rumors and conspiracies.





anticipation are discussed at a higher rate on Twitter during this time period. Because prior research indicated that most of the vaccine hesitancy observed with the COVID-19 vaccine stems from fear of potential side effects, it is not surprising that fear has an elevated mean. In the topics, *Preventative Measures and Risk Factors*, *Vaccine Trials and Efficacy*, *Vaccine Side Effects*, *Vaccine Information and Myth*, *Adverse Reactions and Distrust in the Vaccine*, and *Politics and Rumors About the Vaccine*, trust had the highest mean, followed by fear. This suggests that while users are trusting science, there is a large number of users who are fearful of the vaccine. It is not surprising that trust and fear are prominent due to social media’s ability to polarize groups of opposing viewpoints. The topic *Vaccine Safety Controversies* shows the highest mean in trust, followed by equal scores in anticipation in fear. This suggests that while the majority of users show trust, others are likely grouped into either anticipating the vaccine or fearing its safety. In the topic, *Vaccine Trial Data and Results* the highest means are seen in trust and anticipation. This suggests that users are less fearful when discussing the trials of the vaccines. This is likely due to the high success and efficacy the trials provided.

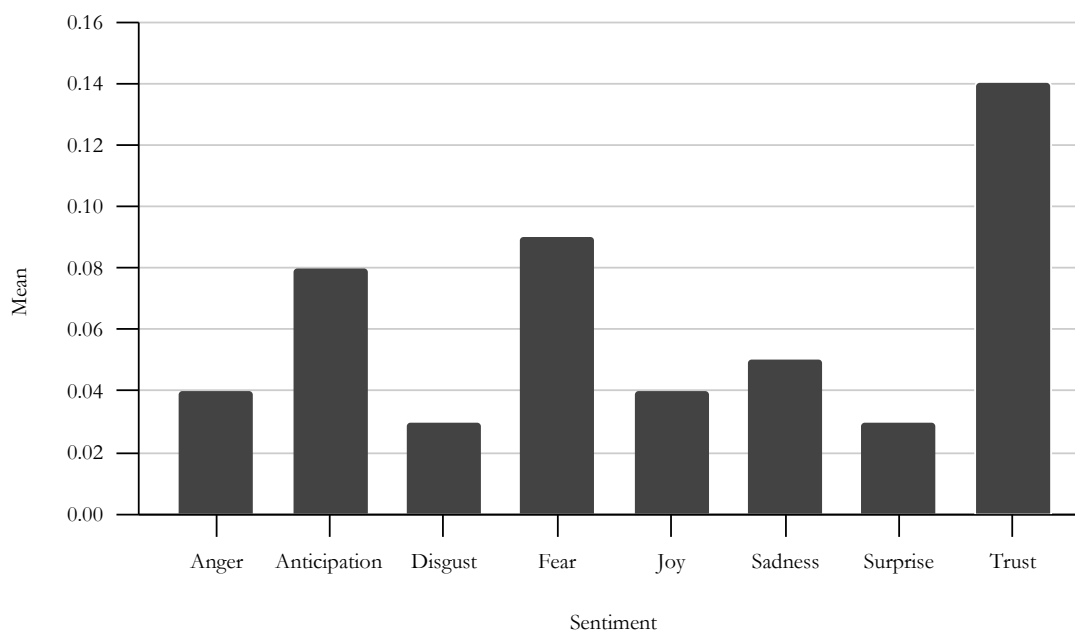
**Table 6: Mean association scores from the emotion analysis**

Topics	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
1	0.04	0.09	0.03	0.11	0.04	0.06	0.03	0.13
2	0.03	0.06	0.02	0.08	0.03	0.04	0.02	0.16
3	0.04	0.08	0.03	0.09	0.04	0.06	0.03	0.12
4	0.04	0.07	0.02	0.09	0.03	0.05	0.02	0.14
5	0.04	0.08	0.03	0.08	0.05	0.05	0.03	0.14

6	0.02	0.09	0.01	0.05	0.04	0.03	0.02	0.17
7	0.04	0.08	0.04	0.10	0.04	0.06	0.03	0.13
8	0.05	0.08	0.03	0.11	0.04	0.06	0.04	0.11
Mean								
total	0.04	0.08	0.03	0.09	0.04	0.05	0.03	0.14

*Note: Topic 1 – Preventative Measures and Risk Factors, Topic 2 – Vaccine Trials and Efficacy, Topic 3 – Vaccine Side Effects, Topic 4 – Vaccine Information and Myths, Topic 5 – Vaccine Safety Controversies, Topic 6 – Vaccine Trial Data and Results, Topic 7 – Adverse Reactions and Distrust in the Vaccine, and Topic 8 – Politics and Rumors About Vaccine.*

**Figure 5: Comprehensive Mean Association Scores Compared**



### Misinformation Analysis

To study the quality of information being shared on Twitter, only positive and negative sentiments were used rather than all 8 of the emotions. Of the sampled 1600 tweets, four types of information were determined: information, misinformation,

opinions, and questions. Information was categorized by any statement that contained verifiable information while misinformation was categorized by misconceptions or unverifiable information being portrayed as factual information. Table 7 shows text examples of each of the four types of information and how positive and negative sentiment can affect the text.

**Table 7: Representative Tweets for Information, Misinformation, Opinion, and Question**

	<b>Positive</b>	<b>Negative</b>
<b>Information</b>	<p>The vaccine made by Pfizer &amp; BioNTech provides strong protection against Covid-19 about 10 days of the first dose. their two-dose vaccine had an efficacy rate of 95% after 2 doses administered 3 wks apart. Many participants did experience aches, fevers and other side effects</p> <p>UPDATE: Pfizer, BioNTech Report 'Data from 43,448 participants, half of whom received BNT162b2 and half of whom received placebo, showed that the vaccine candidate was well tolerated and demonstrated 95% efficacy in preventing COVID-19..</p>	<p>@HalHewgley @NC_Governor 1. One thing to bring up: Covid-19 is part of Coronavirus family that in humans causes the common cold, etc. If enough people take the vaccine the virus should spread significantly slower and avoid further rapid mutations that would make it more dangerous.</p> <p>Social care workers can now get the free, MHRA approved vaccine. Don't forget, you need two doses to be protected against Covid-19, so you can keep caring for others. Find out about the roll-out at <a href="https://t.co/wM9WSJq3My">https://t.co/wM9WSJq3My</a> <a href="https://t.co/704aPha2De">https://t.co/704aPha2De</a></p>
<b>Misinformation</b>	<p>Anyone who after reading all this still wants to get injected with the mRNA vaccine, should at the very least have their blood checked for COVID-19 antibodies. There is no need for a</p>	<p>@CodeMonkeyZ Hear me out: We know COVID-19 came from China, likely developed as a weapon and released on purpose to effect the election. Democrats in</p>



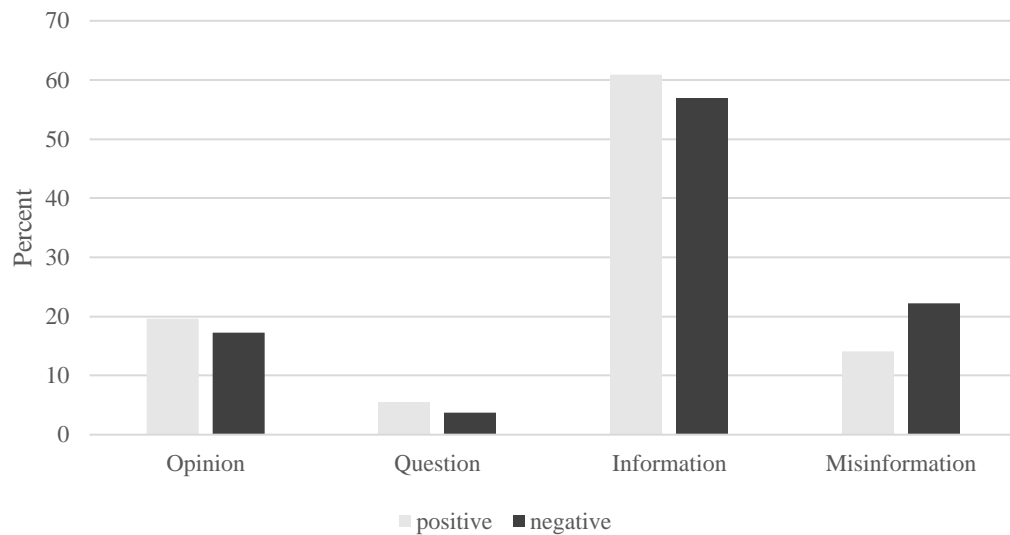
	<p>vaccine in persons already naturally immunized.</p> <p>Did u know that Pfizer COVID-19 vaccine contains protein capsid that is similar to syncytin protein which is essential for placenta development? Meaning? It can create autoimmune response in the females that can lead to miscarriages.</p> <p>Did someone mention population control?</p>	<p>power don't seem to be getting the virus. Could it be China gave them a pre-developed vaccine before releasing the virus?</p> <p>Theres a COVID-19 variant virus going around in multiple countries as we speak. Meaning that they are unsure if the vaccine currently available will work for this new variant. Meaning that they may have to develop a different vaccine for this new variant virus</p>
<p><b>Opinion</b></p>	<p>Frontline workers and the working class should be the first set of people getting the vaccine. Those who are required to work in person among others. Especially for nurses and doctors who consistently work with COVID-19 patients everyday.</p> <p>So grateful to the brave men &amp; women who participated in clinical trials testing covid-19 vaccines including Michael Pearson. He signed up for the #Moderna trial b/c he wanted to make sure African Americans were represented in the studies. @6abc  <a href="https://t.co/3PAW77nFRU">https://t.co/3PAW77nFRU</a></p>	<p>I feel really sorry for those people who are going to be on their own this Christmas. This Covid virus is out of control. We all need the vaccine as soon as possible and get back to living a normal life  #Covid19UK  #CovidChristmas  #Covid_19</p> <p>Today I chose to receive the COVID-19 vaccine. I will still be social distancing, masking, and taking all measures to slow the spread of the virus. Second vaccine dose in 3 weeks!</p>
<p><b>Question</b></p>	<p>How quickly can a COVID-19 vaccine slow the spread of the virus? Once someone is vaccinated, can that person or their family ease up on wearing masks and social distancing? As vaccinations begin across the</p>	<p>@carlheneghan We all know that COVID viruses mutate which is why we need a new flu vaccine every year, and when a couple of years ago, they got it wrong, we had many many more deaths. So,</p>

U.S., @DrLaPook answers questions about the road ahead. <a href="https://t.co/LCyAR1eo0Z">https://t.co/LCyAR1eo0Z</a>	is the COVID-19 vaccine still going to work on the new variant, which has miraculously appeared.
@salilstatistics Hi Salil... just wondering if you looked at the Pfizer vaccine study? they were saying that about 162 out of 17511 in the placebo arm got Covid-19. That leaves a huge number 17349 that didn't get Covid .do you have any thoughts? Perhaps these people were wearing masks?	No. We want to see the data of clinical trials and his reviews & Peer reviewed articles on COVID 19 vaccine ,Has he taken one ? How do they know that the new variant will not impact the vaccine? Have they studied the new variant 2 find out the impacts or it will 2 harm them

From this analysis, it was determined that factual information was found at higher quantities than misinformation, opinions, or questions. It was determined that in both sentiments, information was present at a higher ratio than misinformation. However, Figure 6 shows that while tweets with positive sentiment contained more factual information, misinformation was found at a higher ratio when examining negative sentiment. *Politics and Rumors about the Vaccine* and *Vaccine Information and Myths* did not follow the general trends observed and were exceptions. *Politics and Rumors about the Vaccine* showed more information and misinformation in the negative dataset in comparison to the positive dataset. *Vaccine Information and Myths* showed more informational tweets in the negative dataset in comparison to the positive dataset. Table 8 provides the actual number of findings in each type of information. In the positive dataset, there was more opinions by 1% and more questions by 4% in comparison to what was observed in the negative sentiment dataset. Overall, I found that the data was 61%

information in the positive dataset and 57% information in the negative dataset. The positive dataset had a 14% misinformation and the negative dataset had 22% misinformation.

**Figure 6: The Distribution of Information in Positive and Negative Sentiment**



**Table 8: Categorization of Representative Tweets**

Topic	Categorization	Positive Percent	Negative Percent
Preventative Measures and Risk Factors	Opinion	18	17
	Question	5	1
	Information	67	52
	Misinformation	10	29
Vaccine Trials and Efficacy	Opinion	7	9
	Question	7	6
	Information	76	72

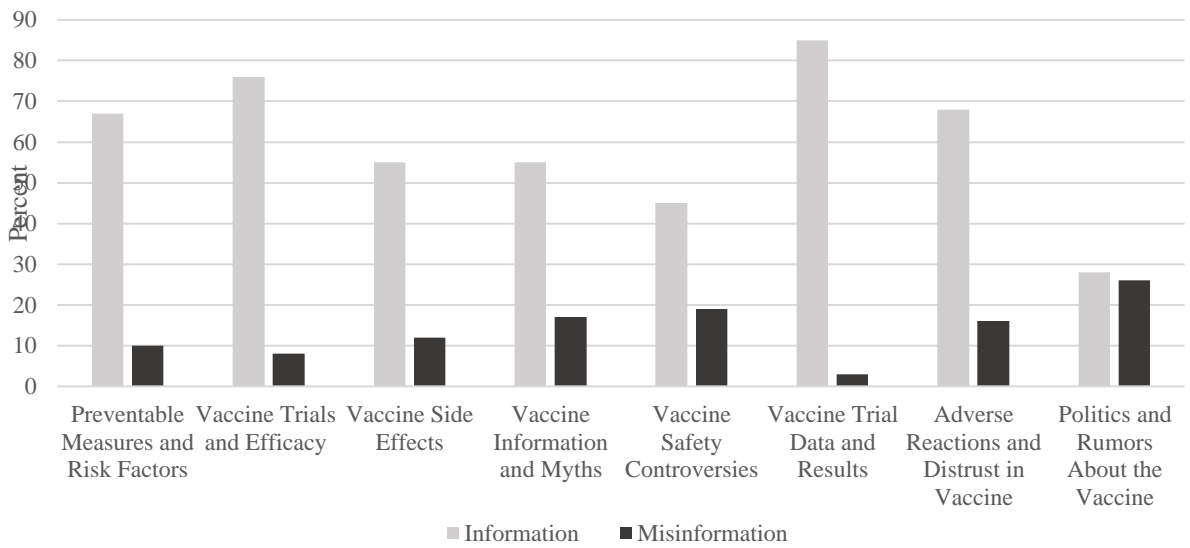
	Misinformation	8	11
Vaccine Side Effects	Opinion	28	21
	Question	3	3
	Information	55	49
Vaccine Information and Myths	Misinformation	12	25
	Opinion	21	12
	Question	5	1
Vaccine Safety Controversies	Information	55	69
	Misinformation	17	16
	Opinion	29	31
Vaccine Trial Data and Results	Question	5	5
	Information	45	33
	Misinformation	19	29
Adverse Reactions and Distrust in Vaccine	Opinion	5	6
	Question	5	2
	Information	85	85
Politics and Rumors About Vaccine	Misinformation	3	5
	Opinion	10	8
	Question	4	7
Adverse Reactions and Distrust in Vaccine	Information	68	52
	Misinformation	16	31
	Opinion	36	31
Politics and Rumors About Vaccine	Question	9	4
	Information	28	36
	Misinformation	26	29

To further examine the quality of information, each topic was analyzed. Because the objective of this analysis was to determine the ratio of misinformation to information, only information and misinformation were examined. Figure 7 shows how information and misinformation fluctuate between topics and sentiment. In both positive and negative sentiment, *Vaccine Trial Data and Results* shows the topic with the most information

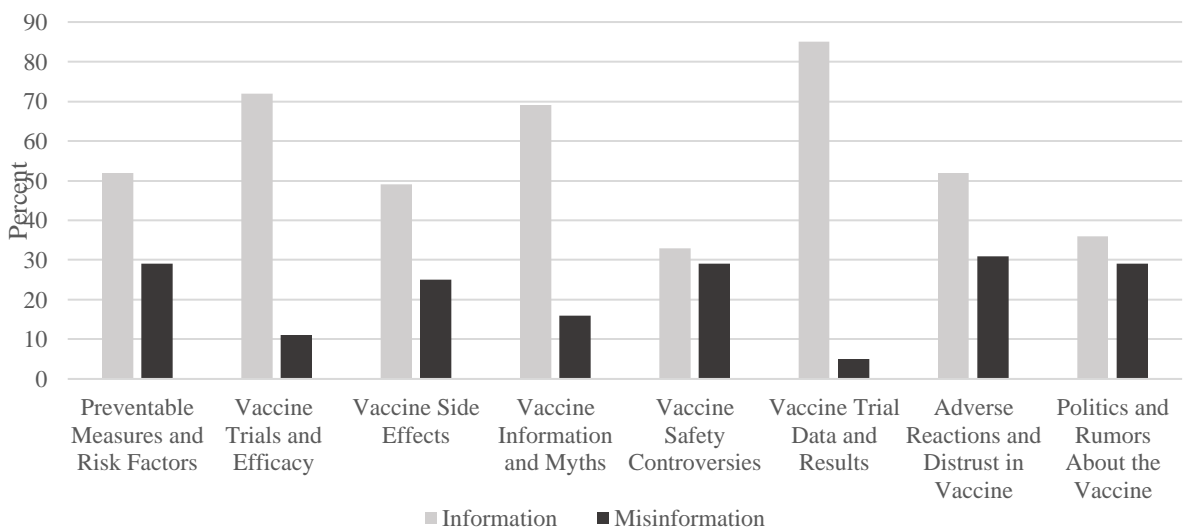
while also containing the least misinformation. Because this topic has a low amount of misinformation in both sentiments, it suggests that sentiment is not impacting the type of information present in these topics. By looking at Figure 7a and 7b, the increase in misinformation in negative sentiment is evident. *Preventative Measures and Risk Factors* shows the largest increase in misinformation (19% increase) followed by *Adverse Reactions and Distrust in Vaccine* which had a 15% increase. Interestingly, *Vaccine Trial Data and Results* had no difference in the percent of misinformation.

**Figure 7: The Distribution of Information Between Topics in Positive (7a) and Negative (7b) Sentiment**

**(a)**



**(b)**



*Note: Figure 7 does not show opinions or questions in positive or negative sentiment*

## V. DISCUSSION

This study provides data that can be beneficial to understanding how social media can play a role in promoting specific sentiments regarding the COVID-19 vaccines. Through sentiment analyses, user's emotions could be studied. By using a topic modelling analysis, key topics of discussion regarding the COVID-19 vaccine could be monitored. With this, the rate of certain themes in discussion could be monitored and compared to provide information on what aspects of the COVID-19 vaccine are discussed. Following the topic modelling analysis, the themes included *Preventative Measures and Risk Factors*, *Vaccine Trials and Efficacy*, *Vaccine Side Effects*, *Vaccine Information and Myths*, *Vaccine Safety Controversies*, *Vaccine Trial Data and Results*, *Adverse Reactions and Distrust in the Vaccine*, and *Politics and Rumors About the Vaccine*. The high positive mean in the topic *Vaccine Trial Data and Results* suggests that the COVID-19 vaccine trials have a positive impact on Twitter users. Negative sentiment was shown to be associated with questioning the intentions of politicians and the government. This suggests that users lack trust in the government's role in the COVID-19 vaccines. Because "government" was associated with negative sentiment among all topics and that fear and anticipation had higher means than almost all of the other emotions examined shows that the government needs to address these feelings that are substantially present in the data. To reduce the lack of trust, the government should prioritize validating and easing fearful emotions that are being amplified through social media platforms by being transparent about COVID-19 findings (Rossi et al., 2020). The

government should address any adverse events to show their intentions. Public figures should work towards spreading awareness and knowledge about the vaccines.

These data additionally show how social media platforms can have an impact on amplifying information and misinformation which can further be used to understand how to communicate effectively with the public and build their trust. For example, our analyses allow us to understand which topics contain the most information. When analyzing the type of information, there was a higher proportion of information, questions, and opinions found with positive sentiment in comparison to that found with negative sentiment. When examining the positive sentiment tweets further, the dataset shows that the information, opinions, and questions tweets generally are either regarding expert opinions or containing scientific information on the COVID-19 vaccine. In the positive dataset, information was mostly found in discussions based on the vaccine's trial data, results, and efficacy. Misinformation in this data set was mostly found involving myths, adverse reactions, politics and rumors, and vaccine information. In comparison, the analysis shows that negative sentiment had a higher ratio of misinformation to information. When the negative sentiment tweets were further examined, it was found that these tweets generally contained myths, rumors, politics, and non-scientific information about the COVID-19 vaccines. In the negative sentiment tweets, information is mostly found in discussion based on the vaccine's trial data, results, and efficacy. Misinformation in the negative dataset is mostly found in discussions regarding preventative measures, myths, safety, adverse reactions, and politics involving the



COVID-19 vaccines. Fact checking information found online should be emphasized to attempt to lower the rate misinformation spreads.

With these data, we can model communication styles in tweets in well-understood topics to tweets that exhibit more misinformation in order to attempt to dismantle the spread of further misinformation. These data can be used to predict what the public will perceive correctly and incorrectly in future similar situations. To reach more users on Twitter, information should be tailored to meet the needs of different user demographics in order to spread knowledge. Experts should communicate at the level of the audience and communicate with empathy to avoid defensive reactions. Because vaccine hesitancy can stem from not fully understanding how vaccines work within the body, providing resources explaining the science behind vaccines could potentially decrease the distrust and hesitancy observed. Because 34% of the total factual information found in tweets with positive sentiment was regarding vaccine trial data, results, and efficacy, this can be interpreted that scientific data is perceived well among the public and should be used when presenting information on social media platforms. On the other hand, tweets with negative sentiment regarding adverse reactions, preventative measures, safety, and politics and rumors about the vaccine accounted for 67% of the total amount of negative sentiment misinformation. This provides information on what the public does not perceive correctly so that experts can carefully communicate any adverse effects and debunking rumors.

Because the vaccine trial data is perceived well, the adverse reactions and side effects caused by the vaccine in the vaccine trials should be communicated to emphasize the low risk of a serious reaction occurring. This could reduce the amount of hesitancy stemming from potential side effects. Because prior studies suggest that politically conservative individuals are more susceptible to misinformation, the government should communicate public health matters in a uniform manner to avoid skepticism over critical information among political parties and maximize the amount of factual information (Roozenbeck et al.,2020).

Based on the distribution of misinformation found across the topics and sentiments in this study, experts can understand where the most misinformation is stemming from and also where the most information is being understood. This is crucial because it allows experts to determine the most efficient ways to maximize the amount of information understood and trust built on social media. For example, this study emphasizes that tweets discussing vaccine trial data and results have a considerably lower amount of misinformation. Therefore, experts should prioritize communicating the data found in vaccine trials because it is more likely to be identified correctly as factual information. By communicating these results, they can increase trust by showing transparent, raw results. Experts who share the results should provide a simpler summary so that the general public can interpret the meaning of the results with a better understanding.

In order to decrease vaccine skepticism, misinformation must be reduced to help the public understand the risks and benefits of the vaccine compared to the COVID-19 infection. By providing users with the maximum amount of information and eliminating misinformation, users can make an informed opinion regarding the COVID-19 vaccine. Topic modelling and sentiment analysis can help provide users with factual information because it can be used to determine the distribution of misinformation. Because we observed that trial results and efficacy have the highest amount of information and low amount of misinformation observed in both the positive and negative sentiments, information containing scientific data should be promoted on Twitter.

By communicating more effectively and minimizing mistrust, the amount of misinformation being amplified should be reduced, and therefore help increase vaccine intent (Saleh et al., 2021). Prior research suggests that positive sentiment increases an individual's reasoning abilities when prompted with misconceptions, whereas negative sentiment was found to be associated with unjustified skepticism (Shereen et al., 2020). Social media platforms should be held accountable in order to stop the amplification of misinformation. Social media platforms should incorporate verified fact checks to label posts as verified or unverified. With this, users would have easy access to knowing what is factual and what is not, therefore potentially reducing the amount of misinformation being misidentified as information. Thus, giving users the resources to make a confident opinion based on information. To increase vaccine acceptance in the future, children and

young adults should be informed of the risks and benefits associated with vaccines.

### **Limitations and Future Research**

This study has several limitations. This study only analyzed data from Twitter. Therefore, these results cannot be generalized to other social media platforms. Political view, education levels, minority status, and personal experiences have been shown to have an effect on vaccine intent (Roozenbeek et al., 2020). Because Twitter which has a younger (55.6% of users are younger than 34 according to Twitter in 2020), more educated user-base than other social media platforms this could have impacted the topic and sentiment distributions in this study. Comparing our study to other social media platforms with users of different demographics is important to observe the extent of this bias. We only collected data from December 1, 2020 until February 28, 2021 during the beginning of the initial rollout of the COVID-19 vaccine. Therefore, our results should not be generalized for the entirety of the pandemic because during the initial rollout of the COVID-19 vaccine studies were still being conducted and less information was known. We only analyzed English speaking tweets; therefore, our data is not representative of all demographics and is biased around English speaking countries; thus is not representative of the entire population. An additional limitation is how the definition of information and misinformation has changed over the course of the pandemic. Due to the lack of knowledge early in the pandemic, information that is now considered as misinformation was believed to be factual (vice versa). This could impact the validity of the ratios observed among information and misinformation, which again

emphasizes the importance to conduct another study towards the end of the pandemic. This study has slight bias due to my professional background in biology.

Following this study, different social media platforms should be compared in order to analyze how different demographics affect the results. Because our study found a large amount of misinformation in the topic regarding politics and rumors about the vaccine, I would examine if political views impact data differently. I would compare our results to Gab, which has a politically conservative user base. Additional studies should be done at different timestamps in the pandemic to observe if there are any temporal trends. For example, this study was done during the initial rollout of the COVID-19 vaccine, but similar studies should be done midway through and at the end of the pandemic to understand if there are comprehensive themes and how new COVID-19 variants such as the delta and omicron affect the results. Ultimately, moving forward, examining how the types of information vary from the beginning to end of the pandemic and among different platforms is necessary in order to determine comprehensive trends.

### **Implications for Improving Public Health**

This study was done to understand sentiment and the types of information observed on Twitter regarding the COVID-19 vaccine. This was crucial to public health due to the lack of agreement and distrust observed among the public, which potentially elongated the effects of the pandemic. This study has the potential to make considerable impacts on research in public health because the data can be used to understand how sentiment and misinformation can contribute towards the prevalence of COVID-19. This

study determined that negative sentiment stems mostly from discussion based around preventative measures and that misinformation mostly stems from discussion regarding adverse reactions and politics regarding the vaccine. This is beneficial because experts can use it to understand vaccine hesitancy and the spread of misinformation.

Additionally, this study has the potential to create considerable advancements in public health. This study's results show that positive sentiment and information is plentiful surrounding discussion regarding vaccine trial data. Overall trends in our data show that positive sentiment is generally associated with more information than negative sentiment and that tweets with negative sentiment have a larger ratio of misinformation to information. Overall, this provides researchers in public health insight on trends that can be used to achieve major advancements such as strengthening vaccine acceptance and trust.

Researchers can extend this study globally. Because our study only analyzed English speaking tweets, continuing this study in terms of all major languages could provide knowledge that is beneficial to public health in non-English speaking countries. This study's design can be applied beyond the COVID-19 vaccine. Because vaccination is a controversial topic, understanding sentiment and types of information regarding major vaccines is important in understand why vaccines are controversial. By analyzing tweets within our data, I observed that tweets from topics regarding the vaccine's efficacy and trial results showed more agreeance in the vaccine while tweets from topics discussing the potential side effects and rumors about the vaccine showed more

skepticism. Comparing our study's results to other vaccines would allow researchers to understand extensive themes of vaccination skepticism and agreeance.

## VI. REFERENCES

- Bewick, Viv, Liz Cheek, and Jonathan Ball. 2004. "Statistics Review 10: Further Nonparametric Methods." *Critical Care / the Society of Critical Care Medicine* 8 (3): 196–99.
- Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research*.  
[https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB\\_iframe=true&width=370.8&height=658.8](https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB_iframe=true&width=370.8&height=658.8).
- Bonnevie, Erika, Allison Gallegos-Jeffrey, Jaclyn Goldberg, Brian Byrd, and Joseph Smyser. 2021. "Quantifying the Rise of Vaccine Opposition on Twitter during the COVID-19 Pandemic." *Journal of Communication in Healthcare* 14 (1): 12–19.
- Branswell, Helen, Andrew Joseph, Maddie Bender, Angus Chen, and Rachel Cohrs. 2021. "Comparing Three Covid-19 Vaccines: Pfizer, Moderna, J&J." February 2, 2021.  
<https://www.statnews.com/2021/02/02/comparing-the-covid-19-vaccines-developed-by-pfizer-moderna-and-johnson-johnson/>.
- Centers for Disease Control and Prevention. (2022). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved March 21, 2022, from  
<https://covid.cdc.gov/covid-data-tracker/#datatracker-home>
- Centers for Disease Control and Prevention. (2022). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved March 21, 2022, from  
<https://covid.cdc.gov/covid-data-tracker/#vaccination-trends>
- Chowdhury, Nashit, Ayisha Khalid, and Tanvir C. Turin. 2021. "Understanding Misinformation Infodemic during Public Health Emergencies due to Large-Scale Disease Outbreaks: A Rapid Review." *Zeitschrift Fur Gesundheitswissenschaften = Journal of Public Health*, May, 1–21.
- Cinelli, Matteo, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. "The COVID-19 Social Media Infodemic." *Scientific Reports* 10 (1): 16598.
- "Coronavirus Disease (COVID-19) – World Health Organization." n.d. Accessed November 28, 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- Devika, C. Sunitha, and Amal Ganesh. 2016. "Sentiment Analysis: A Comparative Study on Different Approaches." *Procedia Computer Science* 87 (January): 44–49.



- Gabutti, Giovanni, Cecilia Trucchi, Michele Conversano, Giambattista Zivelonghi, and Giorgio Zoppi. 2014. "Booster Vaccination: The Role of Reduced Antigen Content Vaccines as a Preschool Booster." *BioMed Research International* 2014 (February): 541319.
- George, Darren, and Paul Mallery. 2016. "An Overview of IBM SPSS Statistics." *IBM SPSS Statistics 23 Step by Step*, 15–21.
- Haynes, Barton F. 2021. "A New Vaccine to Battle Covid-19." *The New England Journal of Medicine*.
- Hendaus, Mohamed A., and Fatima A. Jomha. 2021. "mRNA Vaccines for COVID-19: A Simple Explanation." *Qatar Medical Journal* 2021 (1): 07.
- Hiscott, John, Magdalini Alexandridi, Michela Muscolini, Evelyne Tassone, Enrico Palermo, Maria Soultioti, and Alessandra Zevini. 2020. "The Global Impact of the Coronavirus Pandemic." *Cytokine & Growth Factor Reviews* 53 (June): 1–9.
- Holshue, Michelle L., Chas DeBolt, Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce, Christopher Spitters, et al. 2020. "First Case of 2019 Novel Coronavirus in the United States." *The New England Journal of Medicine* 382 (10): 929–36.
- Huang, Yuan, Chan Yang, Xin-Feng Xu, Wei Xu, and Shu-Wen Liu. 2020. "Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19." *Acta Pharmacologica Sinica* 41 (9): 1141–49.
- Jayaweera, Mahesh, Hasini Perera, Buddhika Gunawardana, and Jagath Manatunge. 2020. "Transmission of COVID-19 Virus by Droplets and Aerosols: A Critical Review on the Unresolved Dichotomy." *Environmental Research* 188 (September): 109819.
- John, T. J., and R. Samuel. 2000. "Herd Immunity and Herd Effect: New Insights and Definitions." *European Journal of Epidemiology* 16 (7): 601–6.
- Kapadia, S. (2020, December 29). *Evaluate topic models: Latent Dirichlet allocation (LDA)*. Medium. Retrieved March 21, 2022, from <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>
- Kaur, Simran Preet, and Vandana Gupta. 2020. "COVID-19 Vaccine: A Comprehensive Status Report." *Virus Research* 288 (October): 198114.
- Konda, M., Dodda, B., Konala, V. M., Naramala, S., & Adapa, S. (2020, June 30). *Potential zoonotic origins of SARS-COV-2 and insights for preventing future pandemics through*

- one health approach*. Cureus. Retrieved April 7, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7392364/>
- Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraittem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour. 2020. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter." *Cureus* 12 (3): e7255.
- Loomba, Sahil, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson. 2021. "Measuring the Impact of COVID-19 Vaccine Misinformation on Vaccination Intent in the UK and USA." *Nature Human Behaviour* 5 (3): 337–48.
- Love, Brad, Itai Himelboim, Avery Holton, and Kristin Stewart. 2013. "Twitter as a Source of Vaccination Information: Content Drivers and What They Are Saying." *American Journal of Infection Control* 41 (6): 568–70.
- "MEDLINE Overview." 2021, February.  
[https://www.nlm.nih.gov/medline/medline\\_overview.html](https://www.nlm.nih.gov/medline/medline_overview.html).
- Memish, Z. A., Perlman, S., Van Kerkhove, M. D., & Zumla, A. (2020, March 28). *Middle East respiratory syndrome*. *Lancet* (London, England). Retrieved April 7, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7155742/>
- Nan, X., Wang, Y., & Thier, K. (2021). Health misinformation. *The Routledge Handbook of Health Communication*, 318–332. <https://doi.org/10.4324/9781003043379-27>
- Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. 2020. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention." *Psychological Science* 31 (7): 770–80.
- Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., Storgaard, M., Al Khalili, S., & Simonsen, L. (2020, September). *Comparing sars-COV-2 with SARS-COV and influenza pandemics*. *The Lancet. Infectious diseases*. Retrieved April 7, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7333991/>
- Pierri, Francesco, Brea Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. 2021. "The Impact of Online Misinformation on U.S. COVID-19 Vaccinations." *arXiv [cs.SI]*. arXiv. <http://arxiv.org/abs/2104.10635>.
- Pormohammad, Ali, Mohammad Zarei, Saied Ghorbani, Mehdi Mohammadi, Mohammad Hossein Razizadeh, Diana L. Turner, and Raymond J. Turner. 2021. "Efficacy and Safety

- of COVID-19 Vaccines: A Systematic Review and Meta-Analysis of Randomized Clinical Trials.” *Vaccines* 9 (5). <https://doi.org/10.3390/vaccines9050467>.
- Quesada, J. A., A. López-Pineda, V. F. Gil-Guillén, J. M. Arriero-Marín, F. Gutiérrez, and C. Carratala-Munuera. 2021. “Incubation Period of COVID-19: A Systematic Review and Meta-Analysis.” *Revista Clínica Espanola* 221 (2): 109–17.
- Raghupathi, Viju, Jie Ren, and Wullianallur Raghupathi. 2020. “Studying Public Perception about Vaccination: A Sentiment Analysis of Tweets.” *International Journal of Environmental Research and Public Health* 17 (10). <https://doi.org/10.3390/ijerph17103464>.
- Ritchie, Hannah, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. 2020. “Coronavirus Pandemic (COVID-19).” *Our World in Data*. <https://ourworldindata.org/coronavirus>.
- Rossi, Giovanni A., Oliviero Sacco, Enrica Mancino, Luca Cristiani, and Fabio Midulla. 2020. “Differences and Similarities between SARS-CoV and SARS-CoV-2: Spike Receptor-Binding Domain Recognition and Host Cell Infection with Support of Cellular Serine Proteases.” *Infection* 48 (5): 665–69.
- Saleh, Sameh N., Samuel A. McDonald, Mujeeb A. Basit, Sanat Kumar, Reuben J. Arasaratnam, Trish M. Perl, Christoph U. Lehmann, and Richard J. Medford. 2021. “Public Perception of COVID-19 Vaccines through Analysis of Twitter Content and Users.” <https://doi.org/10.1101/2021.04.19.21255701>.
- Shereen, Muhammad Adnan, Suliman Khan, Abeer Kazmi, Nadia Bashir, and Rabeea Siddique. 2020. “COVID-19 Infection: Origin, Transmission, and Characteristics of Human Coronaviruses.” *Journal of Advertising Research* 24 (July): 91–98.
- Shulla, Ana, and Tom Gallagher. 2009. “Role of Spike Protein Endodomains in Regulating Coronavirus Entry.” *The Journal of Biological Chemistry* 284 (47): 32725–34.
- Smith, Kendal N., Kristen N. Lamb, and Robin K. Henson. 2020. “Making Meaning out of MANOVA: The Need for Multivariate Post Hoc Testing in Gifted Education Research.” *The Gifted Child Quarterly* 64 (1): 41–55.
- Van den Broeck, Jan, Solveig Argeseanu Cunningham, Roger Eeckels, and Kobus Herbst. 2005. “Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities.” *PLoS Medicine* 2 (10): e267.

Wu, Liang, Fred Morstatter, Kathleen M. Carley, and Huan Liu. 2019. "Misinformation in Social Media: Definition, Manipulation, and Detection." *SIGKDD Explor. Newsl.* 21 (2): 80–90.

Xiang, David, and Lisa Soleymani Lehmann. 2021. "Confronting the Misinformation Pandemic." *Health Policy and Technology* 10 (3): 100520.

