EARLY RESPONSE ON TWITTER TO THE APPROVAL OF THE FIRST COVID-19 VACCINE: AN ANTHROPOLOGICAL PERSPECTIVE

By

EMMA R. WILLHARDT

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Arts

Graduate Program in Anthropology

Written under the direction of

Susan Cachel

And approved by

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey

January 2022
ABSTRACT OF THE THESIS

Early Response on Twitter to the Approval of the First COVID-19 Vaccine: An Anthropological Perspective

By Emma R. Willhardt

Thesis Director: Susan Cachel

A new strain of coronavirus, SARS-CoV-2 or COVID-19, arose in Wuhan, China in December 2019. COVID-19 quickly spread globally and has persisted for nearly 2 years. There have been over 261,000,000 reported cases and 5,200,000 reported deaths as of November 29, 2021. A variety of public health initiatives, such as mask-wearing, have been implemented to manage the transmission of COVID-19 but the most effective step toward controlling the pandemic is wide-spread vaccination. The first vaccine for COVID-19 was approved for emergency usage in the United States on December 11, 2020. However, the growing anti-vaccination movement poses a barrier to vaccinating enough individuals to slow COVID-19 transmission. The anti-vaccination movement argues that vaccinations pose serious health risks and there are natural alternatives to vaccination that work just as well. This movement has a powerful presence on social media and has been getting more mainstream exposure through the Republican party, so is likely to continue growing. Public health campaigns need to consider the current sentiment toward and narratives about the COVID-19 vaccine to successfully convince vaccine hesitant individuals to get vaccinated. This research assesses Twitter users’ sentiment toward the approval of the first COVID-19 vaccine in the 48 hours after the approval through three hypotheses: i. Average user sentiment, indicating sentiment in favor of or against the vaccine, is consistent across geographic regions, ii. Twitter users belong to online communities within the site that are created by direct interactions between users and have different average user sentiment, and iii. These communities can
be defined by common narratives. Sentiment score was calculated through trained language processing using the Natural Language Toolkit in Python. Each user’s sentiment score was then calculated, and average scores were compared across geographic regions and across online communities which were detected through a network analysis of the sampled users. Finally, some important narratives were determined through examining the most interacted with tweets and the users with the largest following within each community. This research concludes that average sentiment does not vary geographically but there is some variation between online communities. Sentiment was largely neutral or slightly positive toward the vaccine. Sentiment and narratives are likely to have changed over the year since the first vaccine was FDA approved, but this research demonstrates that in the first days after approval, Twitter users were not polarized about the vaccine.
ACKNOWLEDGEMENTS

I want to thank my advisor, Susan Cachel, for her ongoing insight and guidance as I have worked on this project the last two years. This work could not have been started, much less completed without her advice. I also want to express my appreciation for Rob Scott for helping me re-envision this research, making it the project it is today, and Carmel Schrire, who generously joined my committee at the very last minute.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES AND FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td><strong>CHAPTERS</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter 1: An introduction to coronaviruses and the COVID-19 pandemic</td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Coronaviruses</td>
<td>4</td>
</tr>
<tr>
<td>1.3 The COVID-19 outbreak</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Long-term effects of COVID-19</td>
<td>9</td>
</tr>
<tr>
<td>1.5 ‘Long COVID’: Twitter in action</td>
<td>13</td>
</tr>
<tr>
<td>Chapter 2: Vaccination Narratives and Medical Anthropology</td>
<td></td>
</tr>
<tr>
<td>2.1 Vaccine narratives</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Patient choice and the modern anti-vaccination movement</td>
<td>18</td>
</tr>
<tr>
<td>Chapter 3: Social-Media Specific Vaccine Hesitancy</td>
<td></td>
</tr>
<tr>
<td>3.1 A general synthesis of vaccine hesitancy on Twitter</td>
<td>26</td>
</tr>
<tr>
<td>3.2 Relevant social-media-based anthropology research</td>
<td>28</td>
</tr>
<tr>
<td>3.3 Social media terminology</td>
<td>33</td>
</tr>
<tr>
<td>Chapter 4: Methods</td>
<td></td>
</tr>
<tr>
<td>4.1 Hypotheses</td>
<td>36</td>
</tr>
<tr>
<td>4.2 Data aggregation and cleaning</td>
<td>36</td>
</tr>
<tr>
<td>4.3 Data analysis</td>
<td>41</td>
</tr>
<tr>
<td>Chapter 5: Results</td>
<td></td>
</tr>
<tr>
<td>5.1 Sentiment score, regional division, and community belonging</td>
<td>42</td>
</tr>
<tr>
<td>5.2 Narrative themes within communities</td>
<td>48</td>
</tr>
<tr>
<td>Chapter 6: Conclusions</td>
<td></td>
</tr>
<tr>
<td>6.1 Discussion</td>
<td>50</td>
</tr>
<tr>
<td>6.2 Limitations</td>
<td>53</td>
</tr>
<tr>
<td>6.3 Future directions</td>
<td>55</td>
</tr>
<tr>
<td>6.4 In summary</td>
<td>56</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>58</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>Appendix A</td>
<td>68</td>
</tr>
<tr>
<td>Appendix B</td>
<td>69</td>
</tr>
</tbody>
</table>
LIST OF TABLES AND FIGURES

Figure 1 .......................................................................................................... 43

Figure 2A ........................................................................................................ 45

Figure 2B ........................................................................................................ 46

Figure 3 .......................................................................................................... 47

Table 1 ......................................................................................................... 43
Chapter 1: An introduction to coronaviruses and the COVID-19 pandemic

1.1 Introduction

Social media provides opportunities for making novel connections and spreading information unlike any form of communication before. A great deal of previous work has focused on the implications of social media, but there have been limited opportunities to observe its functioning in times of public health crisis (e.g. Oh, Kwon, and Rao 2010; Harlow 2013; Prier 2017). The COVID-19 pandemic offers a chance to do such research during various stages of the public health response. So, it is important to understand reactions, effects, and ways to improve such public health measures in the future. Social media, particularly Twitter, has been underutilized in previous public health research (Sinnenberg et al. 2017). The Twitter API is free-access and allows for incredibly large data sets, so can be very effective for narrative tracking and sentiment analysis. This work is informed by this dearth and hopes to demonstrate that anthropology and public health can utilize similar data and learn from each other.

It has nearly been a full year since the first COVID-19 vaccination was approved for administration on December 11, 2020, but too few individuals have been vaccinated to successfully inhibit new strains of COVID from arising and proliferating (CDC 2021). If enough individuals have access to and consent to receiving the vaccination, this could turn the tide toward successfully managing the COVID-19 pandemic (CDC 2021). This research explores the early response to the COVID-19 vaccine through Tweets produced in the 48 hours after the first vaccine was approved. Tweets which contained ‘#covidvaccine’ were collected and the users who produced these tweets were given a sentiment score, defining their opinion toward the COVID-19 vaccine. These sentiment scores were then compared across geographic regions and online communities.
Predominant narratives from each community were also discerned through examining the users and tweets that would have been seen by the most users. The goal of this research is to discern common narratives and shared sentiment about the COVID-19 vaccines within communities of users on Twitter.

The COVID-19 pandemic began in Wuhan, China, in December 2019. Since then, it has rapidly spread across the globe and persisted for two full years, making it a major public health crisis in many countries. As of November 27, 2021, over 261 million individuals have been infected and around 5.2 million individuals have died (CDC 2021). COVID-19 is caused by the virus SARS-CoV-2, which belongs to the coronavirus family. Other members of the coronavirus family have been responsible for the SARS outbreak of 2002-2004 and the MERS outbreak of 2012, so they are well known pathogens. There are few effective treatments for COVID-19 and even mild cases may result in severe long-term effects, so vaccination is the best prospect for controlling the spread of the virus and alleviating the risk of long-term symptoms and death (CDC 2021). There is growing interest in discourse around vaccinations as the anti-vaccination movement grows in pervasiveness across social media, spreading misinformation about the COVID-19 vaccine (e.g. Hornsey et al. 2020; Hussain et al. 2021; Milani et al. 2020). This movement produces narratives of distrust in the safety and efficacy of vaccines through personal anecdotes and questionable research. Especially as new, more transmissible variants of COVID-19 arise, it is more important than ever for enough individuals to get vaccinated to significantly control the spread and the anti-vaccination movement poses a serious barrier to this goal.
The COVID-19 pandemic has provided a plethora of opportunities for novel research. There was an uptick in research investigating the spread of health-related misinformation which coincides with the growing anti-vaccination community. Many of these studies focused on social media, as this research does. Anthropology has been underrepresented in this conversation, likely preferring to do more ethnographic investigations (Wang et al 2019). However, there is a growing body of anthropological research which utilizes social media data to undertake narrative analysis and even ethnography (Breslin et al. 2020; Góralska 2020). This research attempts to demonstrate that anthropology can provide a unique lens for studies concerning vaccination sentiment. While the anti-vaccination movement, which will be discussed in Chapter 2, poses public health issues, and has heightened general skepticism toward certain types of scientific research, most vaccine hesitant individuals are well-intentioned, just concerned about the safety of vaccines (Larson et al. 2016; Stasiuk et al. 2021). Research which tries to understand and even sympathize with these concerns will be more effective at attempting to rectify this skepticism and demonstrate that most medical interventions are thoroughly tested and safe. This work aims to do so through an orientation in patient choice and illness narratives, which will also be explored in Chapter 2.

This chapter will provide an overview of the coronavirus family, the beginning of the COVID-19 outbreak, and the long-term effects of COVID-19. Chapter 2 will explore the history of vaccination narratives and medical anthropology research on vaccination. This will highlight the racialized history of medical testing and the modern antivaccination movement. Chapter 3 discusses the utility of social media for anthropological research and provides some background on sentiment tracking, which is functionally used in this research to determine public opinion on the COVID-19
vaccination. Chapter 4 outlines the methods used for Tweet aggregation and analysis. Chapter 5 provides the results of this research. Chapter 6 includes a discussion of the implications of these results, acknowledges the limitations of this study, and suggests some channels for future work in this vein.

1.2 Coronaviruses

2019-nCoV or SARS-CoV-2, more commonly known as COVID-19, is a novel coronavirus. The first coronavirus was recognized in the 1960s by Tyrrell and Bynoe (1966) when they isolated a virus from the respiratory tract of an individual complaining of a common cold. Coronaviruses are encapsulated, single-strand RNA viruses which usually cause mild infections (Holmes 2003). In subsequent years, researchers identified other strains of coronaviruses which could infect animals, as well (McIntosh, Becker, and Chanok, 1967). There is much greater variety in the animal coronaviruses than in those that habitually infect humans, providing greater opportunities for mutations and strains which cause severe infections. Some of these strains are even zoonotic, meaning that they are transmissible from animals to humans.

The first such coronavirus, severe acute respiratory syndrome, or SARS, emerged in 2002 in southern China and quickly spread worldwide (Drosten et al. 2003). Eight thousand and ninety-eight individuals were infected with SARS between November 2002 and July 2003. Of these 8,098 individuals, 774 died, meaning that SARS had a fatality rate of 9.6% (CDC “Global SARS Outbreak, 2003”). A second coronavirus was zoonotically transmitted from dromedary camels. This strain, named Middle East respiratory syndrome coronavirus or MERS-CoV, infected 2,496 individuals in 2012, killing 858 of those infected for a staggering fatality rate of approximately 34% (WHO “Middle East respiratory syndrome coronavirus (MERS-CoV)”). Researchers have also
detected coronaviruses endemic to animals which have the possibility to transmit to humans, such as several strains which are endemic to North American bats, in an attempt to prohibit zoonotic transmission (Dominguez et al. 2007). The severity of zoonotic coronaviruses has been known for decades and there is precedent for coronavirus outbreaks.

1.3 The COVID-19 outbreak

The COVID-19 pandemic began in December 2019 in Wuhan, China. It quickly spread out of China, infecting more than 1 million people in 171 countries by April 2, 2020 (Taylor 2021). COVID-19 spreads easily from person to person. It primarily spreads during close contact (approximately six feet or less) with an infected individual through airborne droplets which are produced through talking, breathing, coughing, etc. It can spread through airborne transmission when the virus lingers in the air for minutes or hours, even after the infected person has left the space, although this is less common than close contact spread (CDC 2020). So, the best way to avoid spreading COVID-19 is through wearing a mask, avoiding close contact with individuals outside of your household, and washing hands and surfaces frequently (CDC 2020). Public health campaigns from governments and other public agencies have been crucial in informing the public about the risk and spread, although results have, obviously, been limited (Zweig et al. 2021).

There has been a wide spectrum of reactions to the spread of COVID-19 at the governmental level. Some countries, like South Korea and New Zealand, quickly went into strict lockdowns and were able to manage viral spread within a few months (Baker et al. 2020; Yang 2021). In other countries, such as the United States and the United
Kingdom, the virus has been able to proliferate rather freely. There has been a similar spectrum of individual reactions to the pandemic. Some individuals have been hypercareful of the potential medical risks COVID-19 poses while others are reticent to lose their sense of normalcy and have continued as usual. Government issued mask mandates and social distancing guidelines have attempted to squash some of this individualism, but such orders are difficult to enforce and have even received backlash from groups who believe such public health ordinances infringe on individuals’ rights (Juliano et al. 2021). Regardless of the reception of these public health mandates and suggestions, it is undeniable that COVID-19 poses a risk to individuals’ health.

COVID-19 has a much lower fatality rate than its predecessors. As of December 1, 2021, COVID-19 had a 1.99% fatality rate meaning that 1.99% of all individuals who have been infected with COVID-19 have succumbed to the disease (CDC 2021). However, the lower fatality rate, as compared to SARS or MERS, has been offset by the much greater number of individuals who have been infected. There have been over 261,000,000 cases and 5,200,000 deaths reported to WHO as of November 29, 2021 (WHO 2021). COVID-19 infects through and primarily targets the respiratory system although those infected can present with a variety of symptoms or no symptoms at all (asymptomatic). The most common symptoms are cough, shortness of breath, chest pain, fatigue, fever, joint aches, and loss of taste and or smell (CDC 2020). There is a wide spectrum of severity and longevity for these symptoms. In most cases, if an individual has a symptomatic case of COVID-19, they will begin to experience symptoms in four or five days after they are exposed. Most of those with symptomatic cases will experience some symptoms within 11.5 days of exposure, but some may not experience symptoms until two weeks post-exposure (Lauer et al. 2020; CDC 2021).
With the extremely high number of infections, a 1.99 fatality rate, and the potential for long-term effects, which will be discussed in the following section, researchers have been working with urgency to produce a vaccination. According to the Johns Hopkins University of Medicine (2021) typically, creating, testing, and approving a vaccine takes anywhere from 2-10 years. Vaccine candidates must undergo preclinical testing in animal models and three phases of clinical trials. Producing a sufficiently large quantity of vaccinations and acquiring approval from the regulatory body, such as the Food and Drug Administration (FDA) in the United States can take several years. This timeline has been hastened as several successful COVID-19 vaccinations have been produced in the less than a year. The progress was sped up through combining clinical trial phases, emergency regulatory approval, and extra funding and other support from governments. There were more nearly 60 vaccinations at some phase of clinical trial during the summer of 2020, but as of December 1, 2021, only three vaccines have been approved by the FDA: Comirnaty and Pfizer-BioNTech, Moderna, and Janssen COVID19 vaccines (Knoll and Wonodi 2021; Mahase 2020; Schnirring 2021).

This research focuses on vaccine sentiment and skepticism about the safety of vaccinations, which has consistently been a major component of the anti-vaccination movement. This is in opposition to the data which indicates that any federally approved COVID-19 vaccinations are safe and thoroughly tested. The first approved vaccination was a messenger RNA (mRNA) vaccine produced by the company Pfizer. Its phase III trial showed that it was 95% effective, including in high-risk and elderly individuals (Pfizer 2020). mRNA sends messages to the cells in the body to make a protein that is found on COVID-19, enabling the immune system to safely learn how to recognize COVID-19 so it can quickly and efficiently eliminate it in the case of genuine infection.
mRNA vaccines do not use the live COVID-19 virus, so they cannot give the patient COVID-19 and the mRNA does not interact with or alter the patient’s DNA, it only causes long-term changes in the immune system (CDC 2020). The second approved vaccination was another mRNA vaccine made by Moderna which had over 94.5% efficacy across demographic groups (Mahase 2020).

One drawback of both the Pfizer and the Moderna vaccines, from a public health standpoint, is that they both require two doses of the inoculation and ongoing booster doses to reach full immunity. It can be challenging to get all patients to schedule and complete the second vaccination in the appropriate time frame, so promoting vaccinations which require two doses may slightly reduce the efficacy of vaccination programs. A third vaccination made by Oxford-AstraZeneca has recently been approved in the United Kingdom and by the European Union. The emergency approval of this vaccine has been slightly controversial for a few reasons. First, it has lower efficacy (~65%) and this efficacy is still based on patients receiving two doses of the vaccine. Second, the Oxford-AstraZeneca vaccine is a chimpanzee adenovirus vectored vaccine (Corum and Zimmer 2020). Some vaccine-skeptical individuals have taken issue with a component of the vaccine being sources from chimpanzees (Knapp and Rosenbaum 2020). However, adenoviral vectors are well-understood science and are considered a more “traditional” route of vaccine development (Doyle 2020).

Adenoviral vectors are, essentially, genetically engineered viral proteins to which the immune system can recognize and respond. Adenoviruses were first isolated in humans and are associated with respiratory and gastrointestinal diseases (Ghebremedhin 2014). Their genome is relatively compact, making it moderately easy to manipulate and modify (Doyle 2020). To produce an adenoviral vector for a vaccination, the first step is
modifying the viral genome of the adenovirus to produce the same surface protein as the virus which the vaccine aims to eliminate. The modified adenovirus is also unable to replicate, so once it is injected it is unable to reproduce, meaning it cannot cause an infection. Then, the modified adenovirus will enter the cells of the body which will begin to produce the surface protein. The immune system will respond to the surface proteins and create the antibodies to fight any future cells with the same surface proteins, meaning COVID-19 (Doyle 2020). So, the adenoviral vector is unable to infect the patient and the process of genetically modifying the adenovirus is a standard genetic procedure. The major concern amongst anti-vaccination groups about the Oxford-AstraZeneca is that the adenovirus is sourced from a chimpanzee (Knapp and Rosenbaum 2020). However, the source of the adenovirus has no impact on the functioning or safety of the vaccination, as the virus is fully isolated, modified, and rendered incapable of replicating.

1.4 Long-term effects of COVID-19

A study in Italy demonstrated that, with proper follow-up care, some symptoms can be detected in almost 90% of patients 2 months after their first COVID-19 diagnosis (Carfì, Bernabei, and Landi 2020). Other studies have demonstrated that there are potential severe long-term effects of COVID-19, even in those who did not experience severe symptoms (Chang, Lee, and Park 2020; Fan et al. 2020). A case study has even suggested that children seem to experience the same kinds and severity of long-term symptoms as adults do (Ludvigsson 2020). These symptoms include extreme fatigue, breathing difficulties, brain fog, and menstrual changes (Yong 2021). In such cases, which have been termed ‘Long COVID’, symptoms can persist or even intensify over for well over a year (Baig 2020; Yong 2021). Long COVID presents very similarly to autoimmune disease or chronic fatigue syndrome and, like both diseases, has been
underrecognized within the medical community leaving individuals experiencing Long COVID to form a community and advocate for more research about their affliction, as will be discussed in the following section. In other cases, COVID-19 does such significant damage to the cardiovascular system and other internal organs that it may cause lifelong medical issues for those afflicted (Becker 2020).

Many medical researchers are now focused on predicting what sorts of long-term effects COVID-19 may cause and how best to begin treating them. Long-term harm to the lungs is the most obvious risk associated with COVID-19 infection, as the respiratory system is the primary target of infection (Mason 2020). Accordingly, SARS-CoV-2 does seem to cause long-term lung inflammation and damage particularly in the alveoli and the squamous cells (e.g., Jain 2020; Kommos, Schwab, and Longerich 2020). It can also cause pulmonary thromboses, or clots in the arteries in the lungs (Kommos, Schwab, and Longerich 2020). These can be life threatening as they can hamper or entirely block the flow of blood to the lungs.

COVID-19 can cause damage to other organs as well. For instance, it is known that COVID-19 is associated with both ischemic and non-ischemic myocardial injury, or all causes of decreased heart functioning including blockages (Becker 2020; Becker 2020). This leaves individuals at risk for heart attacks and other cardiac issues long after their other COVID-19 symptoms have subsided. Other research has shown that COVID-19 can affect the kidneys, liver, intestines, and even the brain (Renu, Prasanna, and Gopalakrishnan 2020; Skok et al. 2020). This sort of organ damage typically accompanies more severe COVID-19 cases, but still makes COVID-19 infection significantly riskier. These additional risks, beyond the possibility of severe or lifethreatening symptoms, provide the impetus for an efficient, thorough, and
international vaccination program. Not all governments are willing or able to impose
strict regulations to slow the spread of COVID-19, as South Korea and New Zealand did,
so a mass vaccination program is the best path to managing COVID-19 and returning to a
sense of normalcy. Vaccination programs are not effective until a sufficient percentage of
the target population is inoculated to cause herd immunity.

Herd immunity is reached when a sufficiently large portion of a population is
immune to a certain disease that even susceptible individuals are protected from future
infection (Randolph and Barreiro 2020). An individual is considered immune to a
pathogen when they have either been infected with or been vaccinated for the pathogen,
making the immune system able to recognize and effectively eliminate it before it can
cause a notable infection, should the pathogen enter the individual’s body. The threshold
for reaching herd immunity varies between diseases because it is dependent on the basic
reproduction number or $R_0$ which refers to average number of infections a single infected
individual causes in a completely susceptible population. Researchers say that, since
COVID-19 has an $R_0$ of 3, approximately, about 2/3 of the population will need to have
immunity before the rate of infection begins to slow (Randolph and Barreiero 2020;
Fontanet and Cauchemez 2020). However, the effective reproduction number ($R_e$) may be
more descriptive of the variability of infection rates—it weighs the number of individuals
with immunity from vaccinations and the efficacy of the vaccinations—and as of August
2021 COVID-19 had an $R_e$ of about 1.6 (Murray 2021). Either way, pathogen spread is
considered uncontrolled until $R_e$ is less than or equal to 1.

As of December 2020, WHO (2020) estimated that in most countries less than
10% of the population had been infected with COVID-19, so reaching heard immunity
naturally would have devastating consequences. The presence of new variants of COVID-
19 increases this risk as vaccines may not be as effective against new variants, increasing the number of susceptible individuals and thus the $R_e$. One such strain of COVID-19, the Delta variant, arose in December 2020 and nearly twice as contagious as SARS-CoV-2, with an $R_e$ of 4.6, in the United States, and an $R_0$ between 3.2 and 8 worldwide (Murray 2021; Liu and Rocklöv 2021). The Delta variant was able to spread so quickly that it accounted for 99% of new COVID-19 cases in the United States by mid-November 2021 (Katella 2021). Another new variant, Omicron, was recognized by the World Health Organization (WHO) as a “variant of concern” on November 26, 2021, due to some mutations that are likely to make it more able to evade the immune system than previous strains (“Update on Omicron” 2021). These strains pose additional issues to containing COVID-19 to controlled spread.

Additionally, instances of COVID-19 reinfections where an individual tests positive, recovers, conclusively tests negative, then, after a period of time, becomes infected again are becoming increasingly common in both vaccinated and unvaccinated individuals (Iwaski 2021; Maragakis and Kelen 2021). These reinfections indicate that infection does not guarantee future immunity to COVID-19, which further complicates reaching herd immunity. This is partially the impetus for administering COVID-19 booster shots, or additional COVID-19 vaccinations which are given periodically to keep the immune system strengthened against the disease. Booster shots are common; measles, mumps, rubella, and chicken pox are required vaccinations which need boosters. These boosters are given several years after the first dose, unlike the COVID-19 vaccination where the booster is administered 6 months after the original shot (CDC 2021). The CDC approved booster shots for high-risk individuals in August 2021 and for all individuals in November 2021 (LaFraniere et al. 2021). Regardless of the timeframe, this demonstrates
that booster shots are precedented and are as safe as the first dosage of the vaccination. Although, the more doses of a vaccine that an individual must receive, the lower the percentage of individuals that receive all necessary doses (National Center for Health Statistics 2020). These statistics are based on vaccinations that are required to attend school in most states so, theoretically, there should be very few structural barriers keeping individuals from vaccination. This suggests that much of the equivocation is coming from other forms of resistance, such as perception-based hesitation, which this study aims to explore.

1.5 ‘Long COVID’: Twitter in action

This chapter will conclude with an analysis of a known social media narrative relating to COVID-19: community formation around the shared experience of ‘Long COVID’. This provides a relevant example of Mol’s (2008) discussion of patient choice in healthcare, which will be fully examined in the following chapter, and is an example of narrative spread on Twitter. The term Long COVID started as a hashtag on Twitter in late May by Elisa Perego (2020) and gained popularity, spreading outside of the United Kingdom in the following weeks. The term provided a succinct way for patients experiencing abnormally sustained COVID-19 symptoms, as discussed in the previous section, to describe their experience. Most medical narratives did not include sustained symptoms during the earlier months of the pandemic, so doctors and other medical professionals were ill-equipped to provide treatment for and guidance on long-term COVID-19 care (Lokugamage 2020). Some doctors and researchers were also prone to disregard patients’ symptoms, particularly if they had negative results from PCR and antibody tests which indicate if patients are currently or have been infected, respectively (Yong 2021).
The hashtag ‘long COVID’ highlighted an aspect of COVID-19 which needed more attention due to a lack of recognition within the medical field and limited research that reflected the most troubling symptoms individuals experienced (Yong 2021). Research has focused on potential predicting factors which could help medical professionals determine if a patient is more likely to experience sustained COVID-19 symptoms and methods of managing these symptoms (e.g., Nabavi 2020; Sudre et al. 2020). So, in this case, increased patient engagement, a component of patient choice, was essential for garnering proper care and aided overall understandings of COVID-19. Much like other chronic conditions, such as chronic fatigue syndrome or chronic neuropathy, it is difficult to diagnose and assess. Diagnosis will depend on patient input and long-term studies, neither of which have been particularly possible at this point due to constraints on medical care imposed by the longevity of the pandemic. However, it is worth mentioning here because of its connection to patient choice and illness narratives, which will be discussed in the following chapter.

The hashtag also helped individuals sharing similar experiences form a community. The Long COVID campaign has only grown in strength in the following months. Patients can find great comfort and useful resources from connecting with other individuals who are having the same experiences (Callard and Perego 2021). There is even a website, longcovid.org, which provides news, resources, stories, and opportunities for connection specifically for the thousands of patients who have COVID-19 symptoms for three weeks or longer. So, not only is there quantifiable evidence of a growing network formed over Twitter by users of the hashtag long COVID, but also this network has had a qualitative effect on patients’ lives. This provides a COVID-specific example of
the crux of this research: Twitter as a platform for patient-driven narratives and the implications of this sort of spread of information.
2.1 Vaccine narratives

There is a large body of anthropological research on vaccination covering everything from the efficacy and issues of vaccination campaigns in non-Industrialized countries to the process parents undergo when choosing to vaccinate their children and everything in between. This chapter will focus on a review of some of the literature pertaining to how individuals choose whether to vaccinate and which individuals are mostly likely to be vocal about their vaccine-related opinions. This provides the framework for the hypotheses and predictions for this research. Prior to this, it pertinent to define a few terms which are central to this research. First, the term vaccination narrative, the focus of this research, describes the factual, anecdotal, and emotional information which individuals encounter and experience during their decisions about vaccination. This can include demographically targeted public health campaigns, at-home “research”, and communication with other individuals who are either openly opinionated or are undertaking the same decision. For the purposes of this research, vaccination narrative will be treated as a counterpart to the concept of illness narrative.

The earliest vaccination research focused, primarily, on highlighting cultural barriers to inoculation with the goal of eliminating these barriers once they were well understood (Streefland, Chowdhury, and Ramos-Jimenez 1999). These barriers can be glossed as lack of access to healthcare, lack of insurance, and suspicion of the medical field due to previous negative experiences, such as discrimination. While these barriers have been an area of interest for decades, they are still very real for many communities. For example, Black Americans are getting vaccinated for COVID-19 at significantly
lower rates than white Americans (Recht and Weber 2021). Some experts credit the history of medical mistreatment of Black people in the United States for their ongoing resistance to vaccination and other public health interventions (eg. Recht and Weber 2021; Achter, Parrott, and Silk 2005). However, this sentiment is not unique to Black people in the United States. Studies have shown that public health initiatives universally have made marginalized groups, particularly people of color, distrustful of government agencies which seek to impose health regulations (eg. Laplante and Bruneau 2003; Closser et al. 2016). In addition to perceptual barriers to vaccination, there are still many communities, some of which overlap with the communities who are distrustful of governmental and medical agencies, who have poor medical access or are limited by financial constraints (Streefland, Chowdrury, and Ramos-Jimenez 1999; Reich 2014). These considerations encompass one aspect of previous vaccination research and can influence the decision-making process of individuals today as well as the vaccination narratives told.

The term illness narrative was coined by Arthur Kleinman (1988) to summarize the lived experience of patients with long-term symptoms and disabilities and the way that these patients describe their experience. Kleinman was concerned with reforming the way that medical professionals care for individuals with long-term illnesses and pinpointed the interpretation of patients’ descriptions of their symptoms and the difficulties symptoms cause in daily life as the area upon which to improve. The emphasis on the patient’s perceived quality of life tethers vaccination narrative to illness narrative. Both allow patients to be active contributors to their healthcare plans, whether they should be or not.
Another aspect of vaccination narratives comes from the growing impetus on the individual to be an active participant in their medical care and overall health. Patients are increasingly expected to be engaged and decisive in the health and treatment processes. This has been called patient choice and does not always lead to the expected or even the medically sound outcomes (Mol 2008; Zolkefli 2017). Individuals’ engagement with their health and health care has also been termed ‘patient activation’ (Hibbard and Cunningham 2008). This research will use the term ‘patient choice’ as it is more rooted in medical anthropology and incorporates more consideration of the individual’s quality of life.

2.2 Patient choice and the modern anti-vaccination movement

There was a stark increase in the number of medical malpractice claims in the 1960s which led to legislative intervention, updating the laws pertaining to medical malpractice, such as the New York Medical Malpractice Act of 1975 (Walter 1997; Kass and Rose 2016). This resulted in stricter punishments for malpractice as well as updates to the patient-doctor relationship, such as informed consent (Walter 1997). Informed consent is the communication between a patient and their provider that ensures that the patient fully understands any medical interventions before they decide to undergo a procedure. The provider is responsible for providing all necessary information and answering any questions the patient may have and asking the patient to complete proper documentation to confirm their informed consent (Shah et al. 2021). Informed consent is important because it better protects patient autonomy, or the right of competent adults to make decisions about their own care, which has been a fundamental component of institutional medicine since its beginning (Walter 1997). More recently, these ideas have presented themselves in the discourse about patient choice, which emphasizes the
patient’s desires even if they are contrary to the most medically sound decisions (Mol 2008; Tomlinson 2011; Will 2011; Zolkefli 2017). Patient choice allows individuals to opt out of receiving vaccinations and even remain unenrolled from any healthcare insurance.

Some scholars have raised doubts about unequivocally allowing for patient choice because even with informed consent procedures patients are not always the most knowledgeable and therefore may be unreliable at making the optimal choices for medical care (Quill and Brody 1996; Mol 2008; Zolkefli 2017). Mol (2008) claims that patient choice comes from ‘Western’ conceptualization of individual choice or the idea that every individual has the right to make all choices on their own based on their personal judgement of what is best. Mol (2008) also differentiates between care and cure. In her discussion of patient choice, the patients of interest are those with lifelong, incurable illnesses, so much of their ongoing health discussion focuses on how to best care for their symptoms and make their illness manageable day-to-day. The medical focus is on improving the patient’s quality of life. However, when discussing COVID-19, the goal is to cure the infection or to prevent needing a cure through vaccination and other preventative measures. The medical focus is on actionable steps toward ridding the patient of their affliction. Their quality of life matters for some aspects of care, but the medical goal is to reach an end point where the patient is no longer experiencing symptoms, is no longer infected with the disease, and needs no further care.

The distinction in the goals of the medical provider, care or cure, corresponds with differences in the patients’ roles in their healthcare. When the main goal is care, the quality of life for the patient must inform most, if not all healthcare choices, as treating their symptoms is the crux of the job. However, when patients have a curable affliction,
such as COVID-19, medical professionals aim to rid patients of the disease. This is not to say that patients’ lived experience becomes inconsequential, there are still many aspects of care provided, but the state of disease is transient and therefore moving past the period of disease is the goal. This is not to say that patient choice is always the key factor in care programs; sometimes people experience symptoms more harshly, clouding their judgment, or simply need help moderating their behaviors concerning their disease. There are some cases where it may be necessary for a medical professional to intervene and help the patient balance their desires with what is most medically sound (Quill and Brody 1996; Mol 2008; Zolkefli 2017). This calls into question modern movements, such as the anti-vaccination movement, which seeks to impose patient choice on scientifically tested methods of disease prevention.

The modern anti-vaccination movement has been growing since the 1990s but has been able to reach new populations through the Internet. Opposition to vaccines, however, has existed since the first vaccine, for smallpox, was created. The smallpox vaccine was developed by Edward Jenner in 1796 and was in widespread use by the early 1800s due to governmentally dictated vaccination programs in Europe and the United States (Stern and Markel 2005). There were a variety of objections to these vaccination programs particularly from parents who did not want their children vaccinated, individuals who felt forced vaccination infringed on personal freedoms, and religious leaders who were skeptical of the source of the vaccination, since it was isolated from cattle (Durbach 2000). Such skepticism was only heightened once governments and states put vaccination requirements into law. Anti-vaccination sentiment historically has increased every time a major new vaccination was approved, such as the Diphtheria,
Tetanus, and Pertussis vaccine or the Measles, Mumps, and Rubella (MMR) vaccine (Hussain et al. 2018). Although, the MMR vaccine caused controversy above and beyond what previous anti-vaccination movements had incurred (Kasik 2012; Rao and Andrade 2011).

After the MMR vaccination was approved, a British doctor, Andrew Wakefield, published a study in the Lancet that claimed that the MMR vaccine was inadequately tested and could cause autism in children. This study had a very small sample size, an uncontrolled design, and speculated somewhat wildly in the conclusions (Rao and Andrade 2011). Despite this, parents responded very strongly to these allegations causing a notable drop in vaccination rates and some areas in the United Kingdom even lost MMR herd immunity (Kasik 2012). The Lancet quickly retracted the study, published a statement saying that Wakefield’s work was unreliable and inaccurate, and Wakefield had his medical license revoked, but his work still heavily influences the modern antivaccination movement (Hussain, Ali, Ahmed, and Hussain 2018). This highlights the ongoing issue with the anti-vaccination movement: individuals have every right to be engaged with their healthcare, but there are instances where they may be getting unreliable information or simply do not have the necessary knowledge to make ‘good’ choices. Since Wakefield and his colleagues published their article suggesting a connection between the MMR vaccine and autism, there has been an influx of reports of other vaccine-related diseases and illnesses. The growing use of social media has only intensified these reports and has made it possible for individuals to share anecdotes about personal vaccine adverse events (VAEs). These stories of vaccine injury can spread quickly through Internet communities and may sway the opinion of individuals who use the Internet to research healthcare choices.
Briefly, it is relevant to characterize the modern anti-vaccination movement. There are too many perspectives and lenses to provide a complete review here, but research pertaining to the spread of anti-vaccination narratives on social media as well as any new narratives which have arisen pertaining specifically to the COVID-19 vaccination. Generally, the Internet has made scientific and medical information more easily accessible to the public. This has many positive implications for widespread understanding and acceptance of scientific concepts but also has opened new platforms for debates and doubts about the validity and safety of medical processes. Many antivaccinators are parents who are worried about the safety of their children (Garzio 2018; Hussain, Ali, Ahmed, and Hussain 2018). The two most common themes in antivaccination narratives are doubt in scientific and medical institutions, leading to skepticism about the safety of their products, and the benefits of more natural remedies and alternative medicine. This combined with the difficulties individuals face in determining accurate from false information online, results in people making ill-advised choices, despite their good intentions.

The first common narrative amongst anti-vaccination individuals is a distrust in the reliability of science reporting and the safety of ‘Big Pharma’ medical products. This is due, in large part, to an uptick in faux scientific reports, websites, and advocates who use frequently use skewed research or anecdotal reports to claim that vaccines are toxic and unnatural (Hussain, Ali, Ahmed, and Hussain 2018). Several studies have demonstrated that vaccine-related Internet searches, across platforms, are very likely to return some matches that contain vaccine misinformation (Keelan, Pavri-Garcia, Tomlinson, and Wilson 2007; Seeman, Ing, and Rizo 2010). These websites are typically very easy to understand and have strong emotional appeals, such as warnings about
vaccines permanently disabling children. Many highlight narratives of medical mistreatment or instances where medical professionals have ignored or disregarded the patient’s experience to encourage skepticism of the medical field. This serves to reinforces feelings of distrust vaccine-skeptical individuals likely already have toward their physicians. Many individuals, then, turn to alternative health advisors such as chiropractors and physicians who are outspoken against standardized health practices (Garzio 2018). These advisors often suggest natural medicinal remedies in lieu of vaccinations and other medical interventions.

The second major component of modern anti-vaccination movements is a preference for natural remedies and homeopathic medicine. This stems, in part, from fears about vaccine injury, harm caused by the vaccine post-injection, and coincides with growing doubts about mainstream medical practices (Harmsen et al. 2013; Nicoli and Appay 2017). Other individuals express beliefs that natural remedies are innately better than man-made medicines, in a form of nature worship (Aghapour 2020; Deleniv, Ariely, and Peters 2021). Either way, these individuals tend to preference natural “alternatives” to the COVID-19 vaccination.

Natural remedies and homeopathic medicine both use plant-based, mineral supplements, and certain animal products to treat or cure diseases and naturally build immunity. Individuals who prefer such health products likely believe in the strong, natural ability of the immune system to fight any pathogens and avoid severe infection. However, there is very little support for homeopathic medicine as the primary method of maintaining health (NCCIH 2018). Homeopathic medicine follows two rules, ‘like cures like’ and the ‘law of minimum dose’, neither of which agree with fundamental medical science (NCCIH 2018). These rules can be very appealing to skeptical individuals,
though, because they disavow intensive medical treatment and large doses of any sort of medicine. Despite preferencing natural remedies and very small or diluted doses, some studies have found that some homeopathic remedies have substantial amounts of active ingredients, certain heavy metals, or even alcohol, all of which can cause side effects, rendering the ‘natural’ remedies potentially more risky than clinically tested medicines and treatments (NCCIH 2018).

Homeopathic medicine works closely with the idea of patient choice because all the treatments are highly specialized to each individual’s symptoms and their illness or disease experience. This, then, requires that patients be continuously engaged with their care programs, as does patient choice, in its most activated form. However, most homeopathic programs disregard the counterpoint to patient choice: there are certain situations in which medical professionals are better equipped to make healthcare decisions than the patient, as discussed earlier in this section. Furthermore, because homeopathic treatments are so individualized there are hundreds of homeopathic remedies, there are thousands of possible treatments for every ailment (NCCIH 2018). Variability in treatment is not recommended by traditional medical standards, as it introduces potential error. So, homeopathic treatments and care programs are, not only, unrecommended by medical researchers but they also push the boundaries of patient choice past the degree which is ideal for ongoing care (NCCIH 2018). In this way, homeopathic remedies counter medical advice and overemphasize patient choice, which has negative implications for ongoing adherence to medical recommendations, such as vaccination programs.

Over the last decade, vaccination status has become an increasingly partisan issue in the United States (Baum 2011; Walter, Ophir, and Jamieson 2020; Weisel 2021).
Democrats are, generally, more willing to be vaccinated than Republicans and this divide has grown since the former President, Donald Trump, became outspoken in support of vaccine hesitancy (Hornsey et al. 2020). This divide is reflected in the general state of American politics which is facing excessive polarization where individuals are reticent to interact extensively with people who do not subscribe to their political ideology (Heltzel and Laurin 2020). Korn et al. (2020) propose that vaccination acts as a social contract, where individuals are more willing to act generously toward others who have complied with the contract by getting vaccinated or, conversely, act more generously toward others who share their hesitancy toward the vaccination. Since Republicans are more likely to be vaccine hesitant, this may further the divide between the parties as individuals who feel strongly affiliated with their party are likely already reluctant to interact with outsiders. Partisan effects will not be directly studied in this research, but the implications of the anti-vaccination movement aligning itself with the Republican party will be discussed in final chapter.
Chapter 3: Social-Media-Specific Vaccine Hesitancy

3.1 A general synthesis of vaccine hesitancy on Twitter

There are significant social media movements occurring in response to COVID19 and related events. For this reason, along with the public health implications, there has been a variety of research attempting to characterize and track the discourse around the COVID-19 vaccine. This section will detail some of these articles and situate this research within the corpus of work concerned with discourse about vaccination on social media including, but not limited to, research that focuses specifically on the discourse surrounding the COVID-19 vaccination. While research directly concerning COVID-19 is most relevant here, some studies have analyzed vaccine discourse on social media prior to the COVID-19 pandemic and, therefore, are notable as well. This research covers two general areas: the detection of communities creating and perpetuating narratives about vaccination and the characterization of the types of narratives. These articles are the direct predecessors to this research and are fundamental in methodology as well as conclusions.

Perhaps the most important question about vaccine hesitancy concerns who is perpetuating both pro- and anti-vaccine narratives and how individuals in both camps communicate with each other. Understanding this is key to combatting the antivaccination movement and proposing productive public health campaigns as the antivaccination movement relies more on anecdotes and narratives than scientific evidence. Within the United States, anti-vaccination supporters belong to a smaller, but closely knit groups in which individuals tend to produce less novel content but engage more in online discussions compared to pro-vaccination individuals (Germani and Biller-
Individuals who express anti-vaccine sentiments are more likely to promote conspiracy theories and politically identify as conservative or right-wing (Germani and Biller-Andorno 2021; Thelwall et al. 2021). They are also more likely to preference alternative news sources, particularly those that provide counternarratives to topical public health, medical, and political events (Blankenship et al. 2018).

Interestingly, an article by King et al. (2021) has suggested that individuals with PhDs are the most skeptical of all individuals who have finished college or some sort of further degree. However, individuals who did not finish college or high school were the most skeptical population (King et al. 2021). As suggested previously, these individuals cite fears about the safety of vaccinations, particularly those that are administered to children, and distrust of the medical and pharmaceutical institutions when tweeting about their rationale for vaccine hesitancy (Milani et al. 2020; Thelwall et al. 2021). Characterizing the individuals in each community only answers a portion of the question about vaccination discourse on social media, the methods of discussion and information production must also be addressed.

Some research has suggested that despite being a quantitatively smaller population, anti-vaccination narratives have managed to position themselves at the center of social media discourse (Blankenship et al. 2018; Johnson et al. 2020; Milani et al. 2020). This is, perhaps, due to differences in the way that content is produced by each group. Anti-vaccination individuals are less likely to produce novel content but are much more likely to reply to and retweet, adding a comment, existing tweets (Blankenship et al. 2018; Germani and Biller-Andorno 2021). This produces more active discourse between individuals belonging to the community than within the pro-vaccination faction, where individuals mainly produce novel tweets which are then liked or retweeted without
comment by agreeing individuals. There is limited interaction between pro- and antivaccine individuals due to an “echo-chamber” effect, in which individuals seek out those who share their views and opinions and build a community of likeminded individuals who are disinterested in hearing alternative opinions (Yuan and Crooks 2018; Germani and Biller-Andorno 2021; Thelwall et al. 2021). Twitter, and other social media sites, are spaces for building these echo-chambers.

Most previous studies have assessed anti-vaccination sentiment as a whole or in relation to a specific, relatively more isolated outbreak of some novel disease, such as SARs or MERs. Such events are likely to garner less discourse from a broad audience as they are from individuals who already have some interest in the use of vaccinations. The COVID-19 pandemic, in contrast, has greatly affected the public and has dominated world news for the better part of two years. This means that more people are engaging with public health measures and following scientific developments than before and are, subsequently, generating more online discourse, which can be a rich data pool, such as it is for this research.

3.2 Relevant social-media-based anthropology research

Anthropology has historically produced scholarship through going into the field and collecting data. It does not matter whether these collections are fossils, archaeological remains, or ethnographic narratives, anthropologists use these data to construct or reconstruct narratives about human life. Since anthropology is the study of human societies and cultures, it must be continually adapting to understand and describe new cultural phenomena. Social media has been a notable example of a new, worldwide phenomenon in the last three decades. Since the creation of the first, albeit short-lived
social media platform, Six Degrees in 1997, hundreds of other platforms have been conceived and, currently, nearly 4 billion people have an account on at least one social media platform (Dean 2021; Maryville University 2020). Since nearly half of the worldwide population uses social media to some extent, it obviously must be considered as a notable cultural happening and, as such, anthropologists have paid growing attention to ways to efficaciously gather and analyze social media data.

Anthropological attention was first paid to “mass media” in the early 1990s although there was some debate on how or if it falls within the cultural domain. Mass media includes radio, television, film, music, newspapers, magazines, and popular literature and analysis can be approached through a variety of lenses, making it a rich area for anthropological research (Spitulnik 1993). Social media provides a singular avenue for all these types of media so plays a similar role in society today, so offers the same variety of angles for research. A compelling component of social media for anthropological research is the process of virtual community formation and the qualities an individual must possess to belong to such virtual communities.

These virtual communities function similarly to “real life” communities in that they create ideas of the ‘other’, use shared symbolism, and can be conceptualized as a ‘whole’, much in the same way that Spitulnik (1993) describes mass media. Additionally, a recent ethnographic study has suggested that social media populations do not always fall victim to context collapse (Costa 2018). Context collapse is the phenomenon where relationships between individuals become strained or take on an abnormal character because too many distinct social groups exist in one space (Burgess 2019). So, Costa’s work demonstrates that social media studies can and frequently do provide significant
insight into lived social situations. Despite this, there has been some stigma against using data gathered through the Internet for anthropological research.

There have been several notable articles which are relevant to review here that use Twitter or other social media as their main data and their research falls within the scope of anthropology (or public health, which is closely related to this research). The previous section aimed to summarize what is broadly known about vaccine hesitancy and social media, as Sinnenberg et al. (2017) authored a review of over 130 health-related articles, published from 2010 to 2015, which use Twitter as a source of data for their research. They found that over half of the articles analyzed the content of tweets. The rest used Twitter to track the discourse about a particular topic, performed network analysis of users, and assessed the response of Twitter users with Twitter accounts and Tweets. The number of studies using Twitter approximately doubled every year in the study, suggesting a growing acceptance for the validity of using social media as a source of data. Notably, very few of these studies utilized demographic information and vaccinations were not a common research topic, setting this research apart. In their conclusions, Sinnenberg et al. (2017) explicitly suggest that the Twitter API and demographics are under-utilized in health research. This article provides useful background on recent usage of the Twitter API and highlights the gaps in research, which this study hopes to begin to fill.

There has been a surprisingly limited amount of research published by the time of writing that has taken an explicitly anthropological perspective. The most notable study of the sort assessed the reaction to the COVID-19 outbreak and subsequent lockdown on Danish Twitter (Breslin et al. 2020). This research tracked the flow of COVID-19 related
ideas between Danish citizens through targeted searches of COVID-19 terms and analyzed them to detect clusters of narratives and sentiment about COVID-19 and governmental response to the outbreak. They concluded that individuals were primarily concerned with health policies, economic policies, and civic morality. Their sentiment analysis utilized 8 tones, or emotions, to produce a more specific image of Danish sentiment than positive or negative tones, which is a more standard sentiment analysis. Most of the tones stayed relatively constant over the observed period but ‘trust’ dramatically increased, and ‘fear’ decreased once the first lockdown was announced. This study demonstrates that some individuals are forming communities on social media that correspond to shared emotional responses toward the pandemic.

Breslin et al. (2020) demonstrate that very detailed and useful anthropological research can be completed using the Twitter API as the data source. They attempt to answer the question ‘Why do we post?’ through assessing the trending terms and the sentiment behind Tweets. They provide an interesting anthropological interpretation of why individuals post that is worth summarizing. They propose that Tweeting is a form of imitative magic (Malinowski 1931; Frazer 1990) through which individuals seek to influence the issue, COVID-19, through performative behaviors. In this way, using a specific hashtag allows an individual to contribute to a large-scale discourse, belong to a type of community, and share their affective response with similarly affected individuals. This influence may have effects such as moral or normative group cohesion or emotional stabilization, such as the decrease in fear seen in the sentiment analysis. Through this type of analysis, social media data can squarely fit within anthropological research. However, they also note some important drawbacks to social media data. These will be discussed in more depth in the limitations section of Chapter 5 but can be summarized
here. Essentially, while Tweets are great data because they are quick and easy to aggregate, textual analyses can quickly lose context, rendering conclusions subject to doubt. Particularly for anthropological analyses, it is problematic to consider isolated words as emblematic of their original intended use (Costa 2018; Burgess 2019). However, these problems can be avoided through careful analysis, particularly through analyzing pairs of words or phrases instead of individual search terms (Costa 2018).

The final article worth mentioning here is written by a digital anthropologist, who offers some advice for conducting digital ethnographies (Góralska 2020). This demonstrates a growing acceptance of online research within anthropology. Góralska herself studies the production of health knowledge through digital ethnographies, meaning she applies ethnographic principles to online discourse and images to assess cultural phenomena on the Internet. The most notable section of her article details the lack of guidance and standardized practices for virtual fieldwork and ways to deal with potential ethical problems that uniquely arise during sources online posts. She emphasizes that it is important to avoid feeling overwhelmed with the quantity of data through careful cataloguing and detailed notetaking. It is also important to schedule regular work breaks and avoid becoming too immersed in recording the digital data, particularly while many individuals are still working from home, because it is advisable to avoid adding too much screen time to daily routines. It is unclear what the future holds for digital anthropology, but it seems likely that it will play a growing role, even if only as supporting evidence for more traditional fieldwork.

Góralska (2020) raises an interesting point about the ethics of sourcing data from the Internet. Essentially, using online sources allows the researcher to be entirely invisible while they are gathering their data: lurking and observing individuals’ ideas or
conversations does not raise the sort of alarm or require the certification, such as an IRB, that it would in a non-virtual setting. Much of the Internet is entirely public, so sourcing data does not raise much of an alarm. But the researcher may wish to enter private spaces, such as social media accounts which are private (meaning the owner approves all the individuals who can follow or see the account) or private groups on platforms like Facebook, which require a brief application and approval by one of the moderators of the group, to have access. In this case, it would be appropriate for the researcher to notify individuals with private pages that they will be used in whatever research. There are no ethical issues with using publicly available social media posts. This demonstrates that there are distinct benefits to online research. Such research can, essentially, be completed anywhere with little or no budget. It is, also, very easy to acquire hundreds of thousands or even millions of data points through social media APIs. In the 137 studies analyzed by Sinnenberg et al. (2017) over 5.1 billion Tweets, in sum, were analyzed, suggesting that the average study analyzed over 35 million Tweets. Hopefully it is now clear that social media can provide large, novel data sets which can be useful for a variety of forms of research.

3.3 Social media terminology

Prior to discussing social media discourse and individuals who belong to both the pro- and anti-vaccination movements, it is relevant to define some social media terms which will be used throughout the rest of this research. Users on Twitter have limited ways of producing and interacting with content. The basic way of creating new content is through posting a tweet. Tweets have a 280-character limit but there is no limit to the number of tweets an individual can post. So, despite being short-form, individuals can post multiple tweets about a subject if they have more complex thoughts they want to
communicate. In addition to posting stand-alone tweets, users can reply to existing tweets posted by other users either directly, using the ‘reply’ feature, or repost someone else’s tweet to their page, adding their own comment, which is called a quote tweet. In the latter case, the original tweet is shown directly below the added comment; in the former, the reply is added to a series of replies from other users below the original tweet. Quote tweets show up in users’ main ‘timeline’ whereas users need to click on a tweet to see any replies to it, so it takes an extra step to see the discourse occurring.

The timeline is composed of tweets and retweets from users that the individual is “following”, meaning the individual has opted into a direct link to a user’s content. It is also the main interface for Twitter: when a user accesses Twitter they will be brought to their timeline meaning they see tweets from individuals they follow before any other tweets. Individuals can also ‘like’ tweets. The number of ‘likes’ and ‘retweets’ are shown below the content of each tweet so users can easily see which tweets are more popular. Similarly, the number of followers each user has is displayed on their profile, making it public knowledge which users are more popular and have the potential to influence more individuals. If an individual has more followers, then there are more users that will see the individuals’ tweets in their timelines. Users are more likely to interact with tweets from individuals they follow because they are automatically shown the individual’s content instead of having to go searching for it. More followers, in turn, generates more likes and retweets.

Every user can select other users and topics, such as politics or video games, to ‘follow’. Content from followed users and topics comprises the main Timeline of the following user, meaning that when the user logs onto Twitter, they will see a series of Tweets from the users and topics they follow. Each user’s Timeline is customized to
them, as they choose who and what to follow, but the content is presented based on Twitter’s algorithm, so some tweets are more likely to be shown closer to the beginning of the Timeline than others. Twitter’s algorithm prioritizes tweets from individuals with whom users interact with more frequently, tweets with more engagement (measured in the number of likes, retweets, and replies), and tweets containing other types of media, such as pictures or videos (Johnson 2021). So, the tweets that a user sees are dependent on what users and topics an individual follows, the number of interactions a tweet garners, and the types of media in the tweet. This means that, contrary to the original design of the website, Twitter no longer shows tweets in a strictly chronological order in the Timeline. Users will see tweets with which they are more likely to interact, even if there are newer tweets from individuals or topics they follow but interact with less frequently (Cox 2021).

Social media platforms are increasingly specialized to interact with the interests of each user and provide content which aligns with these interests. While this is desirable, as users do not wish to see content in which they are uninterested, this can also intensify some of the echo-chamber effect, which will be discussed further in the following section. Generally, while selecting precisely what topics and users with which an individual interacts can make the user’s experience more enjoyable, it may also isolate them within insular communities that are susceptible to narrative spread, whether the information is reliable or not. On average, 500 million Tweets are sent per day, meaning that narratives can be spread very quickly and amongst individuals who would not interact otherwise (Sayce 2019). Social media is so worldwide and fast-moving that personal interests and feelings of relatability drive how and with what content individuals engage.
Chapter 4: Methods

4.1 Hypotheses

This research is driven by three hypotheses which aim to analyze the intersections of sentiment, narrative, and community belonging on Twitter during the early response to the emergency approval of the first COVID-19 vaccine on December 11, 2020. The hypotheses are as follows:

i. Average user sentiment, indicating sentiment in favor of or against the vaccine, is consistent across geographic regions.

ii. Twitter users belong to online communities within the site that are created by direct interactions between users and have different average user sentiment.

iii. These communities can be defined by common narratives.

4.2 Data aggregation and cleaning

Tweets were aggregated from the Twitter API through the website Stevesie, which acts as a search engine for individuals who want to mass source social media posts. Stevesie acts as an intermediary between the researcher and the Twitter API. The aggregated tweets were analyzed using Python code in the program Jupyter Notebook. Nearly 31,000 Tweets were sourced which were written between December 12 and December 14, 2020. December 12th was chosen as the earliest date from which Tweets were sourced as it is the day after the United States authorized the Pfizer/BioNTech vaccination (Gumbrect, Fox, and Mascarenhas 2020). The Pfizer/BioNTech vaccination was approved by the FDA late in the day on December 11th. Pfizer itself released the press statement about the approval after 11pm, so it is unlikely that any significant
amount of data was left out by collection starting on December 12\textsuperscript{th}. Russia and China both approved vaccines prior to December 11\textsuperscript{th}, but the approved vaccines had not yet completed the final stage of experimental trials (Ledford, Cyranoski, and Noorden 2020). Because these vaccines were not fully tested, the emergency approval was considered more controversial and therefore might have garnered more negative responses and skewed the data (Callaway 2020; Davidson 2020). Additionally, as the tweets sourced were all in English, the discourse was likely to be North America-centric, so it is reasonable to use the United States’ approval date.

Key words, which were used as the search term to aggregate the tweets, were intentionally chosen to be broad, as using uncommon or highly technical terms would also result in fewer available tweets. It was important to avoid sourcing tweets that were discussing some aspect of COVID-19 unrelated to the vaccine approval so, the search term ‘#covidvaccine’ was utilized. The usage of certain hashtags has been suggested to be a signal of community belonging, so it was reasonable to use a hashtag as a search term for this research (Davidson et al. 2019; Dai 2021). The search function used excludes non-exact matches but is not case-specific. Tweets were aggregated that included either search term but not both, as this allows comparison between users that utilize hashtags and those that do not. Search terms were decided to be the most likely to be used in the largest number of Tweets discussing the COVID-19 vaccine. Both the hashtag and the plain key word search were used for the same reason.

Prior to analysis, the text of the tweets was cleaned to remove content that would obfuscate analyses. Data cleansing code was acquired from Specht (2021) and E (2019). First, all URLs were removed, and any tweet that began with “RT”, which signifies that
the post is a simple re-post of another user’s tweet (a retweet), was removed. For the purposes of sentiment analysis, it is best to deal only with original tweets produced by users, as this is the only way of ascertaining a user’s exact sentiment toward the vaccine, as opposed to an individual’s endorsement of someone else’s opinions through a retweet. Then, all words were decapitalized, and a predesigned list of 182 most used words or stopwords, such as ‘a’, ‘and’, and ‘are’, were removed from all Tweets (Lungu 2019). While Tweet aggregation is not case-sensitive, analysis is, so decapitalizing ensures that the same words are not double-counted due to capitalization differences. Removing common terms which do not pertain to sentiment expedites sentiment analysis. The data set had to be further modified to enable network analysis through Gephi, which requires data to be input as nodes and edges. In this case, the individual users were the nodes of the network, and the edges are formed by instances of direct interaction between users such as a retweet, quote tweet, or a direct mention. Code for cleaning the data set for Gephi analysis was sourced from Sprecht (2021). Gephi analysis will be further discussed in the following section.

The benefit of sourcing data from the Twitter API, via a secondary site, is that a large quantity of data is returned, accompanying the text of the sourced tweet. Much of this is superfluous for these analyses, but some was retained as additional variables in the upcoming analyses. This includes the location, whether the user was verified, the number of followers the user had when the data was sourced, and when the user’s account was created. These variables can provide an estimate of how influential a user’s profile might be and enables geographic analyses. Tweets without an associated location were removed from the data set to enable geographic comparisons. Additionally, since retweets and quote tweets were included in the aggregated tweets, the same information about the user
who created the retweeted tweet was included when applicable. This allowed for network analyses of who was retweeting or directly mentioning whom. The number of likes, retweets, quote tweets, and replies for each sourced tweet were also retained to assess how much interaction tweets were garnering, on average.

4.3 Data analysis

The Natural Language Tool Kit (NLTK), a downloadable program designed for various types of textual analysis, was utilized to undertake sentiment analysis. Textual analysis aims to understand language in a way that allows the researcher to gain some sort of cultural or social understanding (Fürsich 2018). It is a qualitative form of analysis which attempts to understand the ideological and cultural underpinnings of the discourse. Textual analysis is a collective term for a variety of methods of content analysis such as ideological, narrative, rhetorical, gender, and discourse analyses. Typically, anthropological studies which utilize textual analysis do so in conjunction with other ethnographic methods because this deepens the context for drawing conclusions. However, textual analysis on its own can be sufficient, depending on the scale of the data and type of conclusions for which a study aims.

Textual analysis is unique from similar types of content analysis in that it first looks for patterns in the text then seeks to identify themes. In combination with geographic and network analyses, it can give a holistic description of what users are expressing what sentiment and what networks exist between users in the online and physical worlds, which is of anthropological interest. NLTK was trained on a scale from 2 to 2, like a Likert Scale, where -2 is strongly opposed to the vaccine, 2 is strongly in favor, and 0 reflects a neutral tone. The training set was built through manually scoring
randomly sampled tweets until there were 40 tweets of each sentiment score—2, 1, 0, etc.—then, the scored tweets can be read into NLTK to be interpreted by a sentiment analysis algorithm (see Appendix A for samples of tweets with each manually assigned sentiment score). Once the sentiment analysis algorithm had been exposed to the training set, the same sentiment scale could be applied to the entire set of tweets. Code for inputting the teaching set and confirming its accuracy was taken from Goyal (2021). Finally, measurements of how many tweets were expressing what sentiment were output. The code for this process was modified from Lungu (2019). Sentiment was determined by the number of words in each tweet the algorithm had determined were expressly positive or negative based on the emotional weight of the words in the text. This differentiates between positive and negative sentiment as well as degrees of sentiment so ‘horrific’ carries more negative sentiment than ‘bad’, for example.

The algorithmically scored data set was further analyzed. First, an average sentiment score was assigned to each sampled user based on the average sentiment from all tweets the user had produced within the data set. This average sentiment score provides a generalization of whether the user had non-neutral sentiment toward the vaccine. Then, because all users in the data set have an associated location, sentiment score could be compared across regions in the United States. Users were manually assigned to one of 5 regions—northeast, southeast, midwest, southwest, and west—based on their reported location (O’Connor 2012). The correlation between average user sentiment score and geographic region was tested through an ANOVA and represented as a box plot using code from De-Yu (2021).

Finally, the program Gephi was used to create a network describing user interactions via the set of sourced tweets. Hammer (2021) and Specht (2021) were used as
guides for how to use Gephi most effectively to create networks of Twitter users. Prior to analysis, the tweet sets were cleaned and reconfigured to distinguish between users who posted original tweets and users who retweeted existing tweets. For these networks, individuals who posted original content were designated as the nodes of the networks and edges, or connections between the nodes, were formed by instances where individuals retweeted, quote tweeted, or directly mentioned (by tweeting out another individual’s username) each other. The tweet sourcing method compiled user information about the creator of the retweeted or quote-retweeted tweets, whereas it did not include information about other interactions, prohibiting creating a network based on ‘likes’, for example. The network analysis returned 7 distinct communities which were further analyzed for narrative themes. Gephi indicated that there were 8 communities, but one was disregarded because it only accounted for about 15 users, which is a negligible percent of the total users. A tweet from the most followed user and the most ‘liked’ tweet from any user in the community were compiled to determine what narratives were likely seen by the greatest number of individuals within a community (Appendix B). These analyses in combination allow for assessment of the types of users, the types and frequency of pro- and anti-vaccine sentiment, and networks formed by users discussing the COVID-19 vaccination.
Chapter 5: Results

5.1 Sentiment score, regional division, and community belonging

The only personal information collected from users was the geographic location from which they were tweeting. This allowed for comparison of users’ average sentiment score across regions. Average sentiment scores were distributed nearly identically across regions (Fig. 1). There was no significant variation between regions (Table 1). The averages did skew slightly positive, with all regional averages falling between 0.11 and 0.17. The southeast had the most outliers, particularly users expressing strongly positive sentiment, although there were also the most users tweeting from the southeast (1,100 users) which likely explains this observation. The west had the lowest average sentiment (0.1135) and the second lowest number of users (692) tweeting from a location in the region. The Midwest (768 users) and Northeast (994 users) had the highest average user sentiment, 0.1699 and 0.1619, respectively. There were no significant differences between regions and the overall average sentiment was very weakly positive toward the vaccine (0.1423).
Fig. 1: Distribution of Sentiment Analysis by Region. Sentiment score ranges from -2 to 2. The boxes are centered around the mean sentiment score, marked as ‘x’, for users within the region. The box itself represents the interquartile range and the upper and lower fences indicate the bounds of the 25th and 75th percentile. Any users whose sentiment fell outside the fences is considered an outlier and is represented as a dot. There is no significant difference in users’ average sentiment scores between regions (p=0.2054).

Table 1: ANOVA of Average Sentiment Score by Region
Network analysis produced 7 significant clusters of users, representing communities of individuals who were either retweeting each other or mentioning each other by username in tweets. Some individuals from the original data set were not mentioned in any tweets and did not retweet or mention any individuals in their tweets, so were not represented in the network analysis. Therefore, the network analysis accounted for 1,511 Twitter users (Fig. 2A). The 7 significant clusters represented about 21% of the total users included in the network analysis or 311 users. Of these clusters, the purple and the light green were the largest, including 72 and 61 users respectively (Fig. 2B). The composition of these communities and some predominant narratives within them will be explored further in the following section.
Fig. 2: Network of Users Tweeting about #Covidvaccine. A) Network analysis of 1,511 Twitter users. Each color represents one community of users. Colored lines show connections between individuals within a community. The color choice for each community is arbitrary. All grey dots are users who are connected to at least one other user but not to a larger community. B) The same network with the outlying users removed to provide a clearer look at the 7 notable communities.
There was more variation between the average sentiment of these communities than there was between regions however, there was still minimal difference across communities (Fig. 3). The average sentiment score fell between 0.10 and 0.30 for all but two communities. The largest community had the lowest average score (0.084) whereas the dark brown community had the highest average score (0.451) with 36 community members. This average was nearly double the average of any other community and is nearly six times as positive in sentiment as the community with the lowest average score. The quantitatively smallest community, dark green, contained only 19 individuals and had the second-lowest average sentiment (0.119). All communities had non-neutral and non-negative average user sentiment, which is concordant with previous analyses.

Fig. 3: Average sentiment score within each community. Each bar represents the average sentiment score for all users in the communities detected through network analysis. The color of the bar corresponds to the color of the community in the network.
5.2 Narrative themes within communities

The most-liked tweet and a tweet from the user with the most followers from each community were identified and manually examined to detect common narrative themes and other similarities between the communities (Appendix B). In some cases, further tweets were assessed from a community to determine if the narrative was addressed by most of the members, in others, there were no strong narrative trends, so no further tweets were assessed. Communities will be addressed from highest average sentiment to lowest average sentiment, for continuity with the previous section. The dark brown community (DB) had the highest average user sentiment and the top tweet and top user reflected that. Both tweets were strongly in support of the vaccine and healthcare workers, who were the first to receive and administer the vaccinations. There were no notable trends outside of the most positive average sentiment toward the vaccine. The top user and tweet from the orange community (O) were both expressly in favor of the vaccine, as expected from the average sentiment score. Both the top tweets were from New York City news outlets and, over 85% of the individuals belonging to this community were tweeting from a geographic location in New York. This is the only community to have a discernable tie to a geographic region.

The pink (PK) and blue (B) communities both had primarily factual, news reportstyle tweets as their top contributors. PK seemed to include more content from reporters and B included more healthcare professionals, including the United States Surgeon General, who was the most followed individual in B and of all users in the network. The top two tweets from the light green community (LG) expressed strong emotional responses—tears from happiness—to photos and videos of vaccines being produced and distributed. This sort of emotional response was expressed by a few other
users in the community, as well, and is not apparent in tweets from the previously discussed communities, suggesting a unique narrative about vaccine distribution in the light green community. The darker green community (DG) seemed to be particularly polarized about the vaccine, as the user with the most followers in the community presents staunchly antivaccine ideas and the most-liked tweet was from a series of tweets describing the efficacy and safety of the new vaccine, based on FDA documents. Many individuals in this were responding directly to the CDC or some other government agency as well as other users in their tweet, which explains how such polarization can fit within one community.

The purple community (P) was the only one where both top tweets expressed concern over the safety and necessity of the COVID-19 vaccine. The most liked tweet in P came from a user who identifies as a ‘vaccine choice advocate’ on her Twitter profile page. The other tweet came from a freelance journalist who was supporting claims about vaccine unreliability from a far-right magazine, The Federalist. While there was more obvious anti-vaccine sentiment expressed within this community, there were also many users who were in support of the COVID-19 vaccine but were displeased with the work done by President Trump leading up to the approval. Nearly all the tweets sampled from P directly mentioned @realDonaldTrump, along with other users, suggesting that direct reference to the former President is a unique narrative trend within this community. There do seem to be some narrative trends that differ between some communities detected through network analysis. These distinctions will be discussed in the following chapter.

Chapter 6: Conclusions
6.1 Discussion

This research addressed three main hypotheses:

i. Average user sentiment, indicating sentiment in favor of or against the vaccine, is consistent across geographic regions.

ii. Twitter users belong to online communities within the site that are created by direct interactions between users and have different average user sentiment.

iii. These communities can be defined by common narratives.

Overall, this research produced limited statistically significant results but detected some interesting trends in the prominent narratives within communities. The hypotheses will be addressed in order before moving to discuss the limitations of this research and some directions future studies may take. User sentiment was, approximately, equally variable between all regions meaning there was no difference in sentiment toward the vaccine across geographic space (Fig. 1; Table 1). This is concordant with previous research that has demonstrated that communities formed around a single topic are likely to be nonlocalized and that vaccination-specific twitter communities are randomly geographically dispersed across the United States (DeJohn et al. 2018; Bello-Orgaz, Hernandez-Castro, and Camacho 2017). Additionally, social media is widely accepted as a space that allows interactions between individuals who would not otherwise interact, in part because they are spatially isolated (Sayce 2019). This supports the first hypothesis: average user sentiment was not geographically localized.

Average user sentiment is slightly more variable between communities than between geographic regions, between which it is nearly identical. Network analysis successfully defined 7 notable communities, based on direct interactions between users,
that accounted for just over 20% of the users included in the network (Fig. 2A). All but two communities (DG and P) had higher average user sentiment than the average across geographic regions (Fig. 3). There was more variation in average user sentiment between populations, as well, but still not a significant amount of variation. P was also the largest community and was relatively unconnected to the other communities. Communities are connected if an individual retweets, quote tweets, or mentions an individual in a different community (Fig. 2B). These connections are instances where narratives may be shared between communities if there are distinct narratives to share. LG was the second largest community but had the most connections and was only completely disconnected from P. These connections suggest that narratives are unlikely to be confined to one community, as individuals are frequently interacting with users from other communities. This is, more or less, the case for 3 communities: DB, PK, and B. While DB had much higher average sentiment, tweets from all three communities simply expressed different degrees of support for the COVID-19 vaccine, the researchers responsible for producing the vaccine, and the healthcare workers who will administer it (Appendix B). The composition of the users varied a bit, B had more medical professional and PK had more journalists, but it is reasonable to assume that individuals are likely to engage with others in their profession about relevant issues, so this is not particularly groundbreaking.

LG and O had notable results, but they are not related to anti-vaccination narratives. The sampled tweets from LG both expressed the sentiment that the user was so happy about the vaccine that they could cry. It seemed unlikely that both the top tweets from LG would mention ‘happy tears’, so the rest of the tweets from the community were examined and there were a few more that expressed the same feeling. LG was middling in terms of average user sentiment, so these strong emotions were a bit surprising. It is
possible that these tweets were balanced out by the number of tweets with completely neutral sentiment, obscuring the few users who expressed the same, strong emotional response to the vaccine approval. O was notably geographically localized: over 85% of the users in the community sent their tweets using #covidvaccine from New York. This is likely because the very first dose of the vaccine administered in the United States was given in a New York City hospital, so there was a flurry of local news coverage on it, in addition to the national reports (Saplakoglu 2020).

There were some more relevant trends, as well. First, this research did not produce evidence of any echo chamber effects within the communities (Yuan and Crooks 2018; Germani and Biller-Andorno 2021; Thelwall et al. 2021). There were plenty of instances where users within the community expressed polarized ideas. For instance, the individual in DG with the most followers had used #covidvaccine in a tweet with a poll asking which vaccine users thought was going to kill the most people. The tweet with the most favorites from DG was part one of a chain of 8 tweets from a doctor, describing the FDA’s rationale for giving the Pfizer vaccine emergency approval. The rest of DG was similarly polarized to the top two tweets, with tweets from self-identified anti-vaccine advocates as well as medical professions.

Users who expressed anti-vaccination sentiments and identified as vaccine hesitant, or some similar term, on their Twitter profile were present in every community. This presence supports previous research that suggests that anti-vaccine individuals are pervasive across social media despite being a minority viewpoint (Blankenship et al. 2018; Johnson et al. 2020; Milani et al. 2020). They were only particularly prominent in DG, as discussed above, and in P. The sampled tweets from P promoted a vaccine conspiracy narrative and directly mentioned former President Trump to ask him to
prohibit a vaccine mandate, which had not been formally proposed at the time. Interestingly, many of the tweets in P directly mentioned Trump, suggesting that including Trump in the discourse about the new vaccine was a unique narrative feature to the community. These mentions are not always vaccine hesitant, as the sampled tweets are. Some tweets mention Trump to criticize his promotion of vaccine hesitancy or his lack of actions to mitigate COVID-19 through the approval of the first vaccine.

This emphasizes the growing partisanship of the anti-vaccination movement, led by Trump (Hornsey et al. 2020). He is the only individual to be directly mentioned a notable number of times in the collected tweets and the community which mentions him most frequently is also the only community where the most followed user and most liked tweet are vaccine hesitant. It is crucial to understand the current popular anti-vaccination narratives to be better able to address the base causes of vaccine hesitancy, particularly in a public health crisis such as the COVID-19 pandemic.

6.2 Limitations

This study was not without certain limitations. Demographic data about the users was limited and hard to obtain because of Twitter’s privacy policy and the dearth of information that Twitter requires users to submit when making an account. Additionally, Twitter has, historically, had an issue with a high quantity of ‘bot’ or fake accounts that are frequently created for a malicious purpose such as influencing elections or propagating misinformation (Russell 2018). While Twitter has worked to eliminate bot accounts, there may still be some that have evaded detection, particularly if the purpose of the bot is just to retweet news articles which is somewhat common, but this could potentially contribute to the highly matter-of-fact tone that most of the collected tweets
take. Similarly, Twitter has been attentive to the rise in false information and negative sentiment about the COVID-19 vaccines and has taken steps to remove such content (Twitter Safety 2020). This potentially suppressed narratives from individuals who are skeptical about the new vaccinations. Additionally, recent research has suggested that anti-vaccination sentiment may be quantitatively the minority of the social media posts sampled but has a demonstrably greater effect (Johnson et al. 2020). Quantifying such impact was outside of the scope of this research but poses a possibility for future research as will be discussed in the following section.

Aside from limitations that innately arise from using Twitter as the social media platform for this research, some anthropologists have questioned the efficacy and validity of using keyword searches to do research on sentiment and textual trends (Breslin et al. 2020). However, taking a comparative lens by also sourcing through the hashtag search adds sufficient complexity to alleviate at least some of this worry. Additionally, there is only a certain subset of the population that uses social media, so it is possible that the sample unequally represents certain groups. Pew Research Center (2019) determined that a greater percent of Twitter users are under 50 years old, are college graduates, and make greater than $75,000 a year than the average of the public in the United States. Twitter users do roughly match the United States general public composition when assessing gender and racial categorization. The Twitter API does not directly provide demographic data beyond the city or region from which the Tweet was sent (and some users have location services disabled, so region is not provided) so demographic data had to be acquired through secondary processes. This introduces more potential error than would be present in a data set acquired from a social media platform that requires individuals to provide more demographic information, like Facebook. These limitations do pose some
issues, but this research has attempted to control for and be cognizant of such issues during analysis and discussion.

6.3 Future directions

Previous research has demonstrated that sentiment on Twitter tends to skew more negative than on other social media sites such as Facebook or Reddit (Hussain et al. 2021). It could be useful to do a comparative study on the response to the COVID-19 vaccine on multiple social media platforms. Additionally, adding more time depth to such a study may reveal more complex sentiment trends, such as growing negative sentiment toward the vaccine over time and different narrative trends. A follow-up to this research could explore the change in sentiment toward the vaccine over time by sampling recent tweets about vaccination from the users included here. Sentiment score and community belonging could be assessed in the same way and then compared to the results from December 2020, providing a glimpse at the changes in sentiment and narrative over the first year of COVID-19 vaccine roll-out.

While somewhat outside the scope of this research, there is significant evidence that the COVID-19 vaccine has become a political issue (King et al. 2021; Schmelz and Bowles 2021; Sharfstein et al. 2021). There is potential for some useful interdisciplinary research between political science and anthropology, investigating the rise of politicization of public health measures, particularly vaccines. As previously mentioned, the anti-vaccination movement has been accepted by the Republican party, exposing more individuals to unreliable narratives about the safety and efficacy of vaccines. It would be useful to incorporate political party affiliation in a future study about the sentiments toward vaccines within online communities to determine the strength of the
correlation between partisanship, sentiment, and community belonging. Public health measures are only effective if a significant majority of the population is enthusiastically participating, so it is of the utmost importance to figure out a way to depoliticize such measures and increase participation.

6.4 In summary

There remains considerable work to be done to understand the spread of vaccination narratives on social media. This study looked at a small sample of tweets from the 48 hours after the Pfizer/BioNTech COVID-19 vaccine was given emergency clearance by the FDA. This study indicated that Twitter users who joined the discourse about the COVID-19 vaccination were not geographically localized and on average, felt weakly in favor of the novel vaccine. The differences in predominant narrative between some of the communities suggests that some individuals are being exposed to different content based on the groups to which they belong, even if the overall sentiment toward the vaccine is similar across communities. Although, as a full year has passed since the first vaccines were approved worldwide, and vaccine hesitancy is increasingly politicized in the United States, it is likely that sentiment would be less neutral and more variable now.

This highlights the issue at hand: social media spreads narratives quickly, some online communities are likely to become insular and promote a certain set of narratives, and these narratives change quickly over time. When such narratives involve individuals’ intent to comply with public health initiatives, namely receiving the COVID-19 vaccine, a real problem can occur. This is the case with the growing resistance, spearheaded by the anti-vaccination movement, to receiving the vaccine. COVID-19 will pose a serious risk until viral transmission is more strictly managed, which will not occur until a large
majority of the population receives the vaccine. As suggested by previous research and supported here, vaccine hesitant users can be found across online communities, even communities which are not overtly opposed to the vaccine. Careful analysis of the prevalent narratives within online communities, as undertaken here, can highlight narratives that may be obscured by general assessments of vaccine sentiment. This, in turn, may prove useful for determining more effective ways to persuade individuals that vaccines are safe and necessary, which is crucial to ensuring enough vaccines are distributed to control COVID-19 transmission and evolution.
Bibliography


Harlow, S. (2013). It was a “Facebook revolution”: Exploring the meme-like spread of narratives during the Egyptian protests. Revista de Comunicación, 12, 59–82.


https://doi.org/10.7326/M20-0504

https://doi.org/10.1038/d41586-020-03441-8

Liu, Y. and Rocklöv, J. (2021). The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *Journal of Travel Medicine, 28*(7), taab124. https://doi.org/10.1093/ijtm/taab124

https://blogs.bmj.com/bmj/2020/07/10/patientsexperiences-of-longcovid-are-missing-from-the-nhs-narrative/


https://medium.com/analytics-vidhya/exploring-twitter-data-using-pythonaf1287ee65f1

http://dx.doi.org/10.1136/bmj.m4471


https://online.maryville.edu/blog/evolution-social-media/


https://doi.org/10.1183/13993003.006072020


WHO. (2020). Coronavirus disease (COVID-19): Herd immunity, lockdowns and


Appendix A:
Table 2. Sample Tweets for Each Sentiment Score

<table>
<thead>
<tr>
<th>Sentiment Score</th>
<th>Example Tweet</th>
</tr>
</thead>
</table>
| -2              | "I will ensure that I and my family - wife and daughter is NOT GOING TO TAKE THIS rapidfire #CovidVaccine solution. It's a Jhumla even as with each passing day Variants of #Covid19 appear and twist everything. This is a desperate and unscientific solution to pacify DEFEAT."
|                 | (From @terence_fdes) |
| -1              | "You trust him (the enemy) After all that he has done to destroy us... how could you allow him to stick a needle in you..” #Farrakhan on #CovidVaccine (From @WritetheeFuture) |
| 0               | "Pfizer began shipping its COVID-19 vaccine around the nation today. Healthcare workers and nursing home residents are expected to be among the first to receive the vaccine. #CBSNews #CovidVaccine (From @ashley_godwin) |
| 1               | "First Coronavirus vaccines being given the the US. The beginning of the end! #3pmthanks #COVID19 #CovidVaccine (From @3OClockWakeUp) |
| 2               | "Just opted-in to receive the #CovidVaccine! This is a big step in the right direction. #WearAMaskSaveALife #gethevaccine (From @dkalnow) |
Appendix B:
Table 3. Sampled Tweets for Community Narrative Analysis

<table>
<thead>
<tr>
<th>Community (by color)</th>
<th>Example Tweet</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@DrFNA / thank U for your courage and for this thread about your experience as a volunteer for the #CovidVaccine. I believe it will save some lives.</td>
<td>From Dr. Quinn Capers, IV who is a cardiologist and professor at UT Southwestern (21,282 followers, 0 likes).</td>
</tr>
<tr>
<td></td>
<td>This should be declared National I Love Nurses Week #CovidVaccine <a href="https://t.co/JRGj2fCAhZ">https://t.co/JRGj2fCAhZ</a></td>
<td>From Joseph L. Mills, MD (8,370 followers, 11 favorites). The link is to an MSNBC video showing nurses receiving and administering doses of the vaccine.</td>
</tr>
<tr>
<td></td>
<td>BREAKING: Sandra Lindsay, ICU nurse and the director of patient care nursing at @NorthwellHealth on LI about to be among first Americans to get #CovidVaccine #OperationWarpSpeed @realDonaldTrump</td>
<td>From TalkRadio 77 WABC, a New York City-based news and talk radio station (19,932 followers, 2 likes).</td>
</tr>
<tr>
<td></td>
<td>@NYGovCuomo with a quick thank you for Sandra Lindsay, the nurse about to get the first #CovidVaccine in U.S. #nbc4ny</td>
<td>From Andrew Siff, a reporter from NBC NY Channel 4 (8,427 followers, 47 favorites). This tweet was accompanied by a video of former Governor Cuomo during a press conference.</td>
</tr>
<tr>
<td></td>
<td>. @MSNBC @maddow @amjoyshow @TiffanyDCross please have @ScientistSwanda on to describe the #CovidVaccine.</td>
<td>From Adam L. Stanley (12,547 followers, 1 favorite). He seems to be a lifestyle coach with a neoliberal ideology. He recently changed his @ and cleaned old tweets off his page, so this tweet is not actually accessible anymore.</td>
</tr>
<tr>
<td>More coverage here about how Michiganders helped pave the way for the #Pfizer #COVID-19 vaccine. Some interesting people involved in ways big and small that brought the nation to this point: <a href="https://freep.com/story/news/health/">https://freep.com/story/news/health/</a></td>
<td>From Kristen Shamus, a reporter at the Detroit Free Press (4,624 followers, 44 likes). This tweet was part of a series of tweets about the vaccine shipment from the Pfizer plant in Portage, Michigan.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>2020/12/13/how-michigan-helpedpave-way-first-coronavirusvaccine/6521469002/… @freep #CoronavirusVaccine #CovidVaccine</td>
<td>Today is a historic day. The #CovidVaccine is being delivered across the country to save lives and bring hope to millions. From Dr. Vivek Murthy, the U.S. Surgeon General and the top followed individual in the network (936,565 followers, 115 favorites).</td>
<td></td>
</tr>
<tr>
<td>ICYMI: Our Practice Advisory provides an overview of currently available #COVID19 vaccines and guidance on use in pregnant and lactating patients. As #COVIDvaccine development and regulatory approval are rapidly progressing, info and recs will evolve w/more data: <a href="http://bit.ly/3qSOCaJ">http://bit.ly/3qSOCaJ</a></td>
<td>From ACOG, The American College of Obstetricians and Gynecologists (44,019 followers, 26 favorites).</td>
<td></td>
</tr>
<tr>
<td>Gotta say, these images of #HeroHealthCareWorkers rolling up their sleeves for #CovidVaccine shots making me teary. Science might just survive this.</td>
<td>From Katherine Eban, an investigative journalist working for Vanity Fair (18,536 followers, 58 favorites). She published an investigation in June proposing that COVID-19 came from a lab-leak. This is a quote tweet of a report about the first dose of the vaccine being given in Florida.</td>
<td></td>
</tr>
<tr>
<td>I unexpectedly broke down sobbing while watching images of the @pfizer #CovidVaccine distribution center on @ThisWeekABC today. Thank you, science.</td>
<td>From Michael A. Gisondi, MD, a medical education researcher at Stanford (5,058 followers, 74 favorites).</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Which #CovidVaccine will kill more people in 2021? #COVID19 #Covid_19</td>
<td>From @AaronDayAtlas (45,283 followers, 0 likes) an account that currently focuses on cryptocurrency. This example is missing the second part of this tweet which was a poll that allows other users to respond to this question. 21 users responded.</td>
<td></td>
</tr>
<tr>
<td>Howdy Folks! I attended a discussion w/ @CDCgov medical officers on the Pfizer-BioNTech #CovidVaccine Here are a few answers to some common q'ns: 1. Should patients receive special info before getting the vaccine? --&gt; FDA Fact Sheet for Recipients and Caregivers (1/8)</td>
<td>From Dr. Lipi Roy who is currently a public speaker and the medical director for Housing Works, a foundation working to end homelessness and AIDS (18,118 followers, 171 favorites). This tweet was the first tweet in a series of 8 tweets Dr. Roy released, all discussing this topic.</td>
<td></td>
</tr>
<tr>
<td>This is a good list of reporters who gave their OPINION that Pres. @realDonaldTrump was falsely promising #CovidVaccine would be done by end of 2020. Will any of them admit they were wrong? Will they learn from this to report just the facts?</td>
<td>From Emily Miller, who runs her own journalism blog and is verified on Twitter (62,346 followers, 31 favorites). This is a quote tweet responding favorably to an article from The Federalist, a far-right news source.</td>
<td></td>
</tr>
<tr>
<td>@realDonaldTrump Please keep in mind over half the country intends to refuse this vaccine. let us be free. No #CovidVaccine Mandates</td>
<td>From Rita Palma who identifies as a vaccine choice advocate (4,239 followers, 33 favorites).</td>
<td></td>
</tr>
</tbody>
</table>