Title

Public Perceptions of COVID-19 Vaccine Information: Quantitative Analyses of Survey and Twitter Data in the United States

Author

Venzon, Aldreen

Publication Date

2022

Public Perceptions of COVID-19 Vaccine Information:
Quantitative Analyses of Survey and Twitter Data in the United States

By

ALDREEN VENZON
DISSERTATION

Submitted in partial satisfaction for the requirements of the degree of

DOCTOR OF PHILOSOPHY

in

Nursing Science and Health Care Leadership

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____
Katherine K. Kim, Chair

____
Jill. G. Joseph

____
Jingwen Zhang

____
Nick Anderson

Committee in Charge

2022

Table of Contents

ABSTRACT

Objectives: We aimed to understand the public perception of COVID-19 vaccines using survey and Twitter data. For the survey study, we focused on examining the COVID-19 vaccine perspectives of the rural population in the Central Valley of California, which was predominantly Latinx. Specifically, we looked at the level of trust in the source and content of the vaccine information they received, their view of the safety and effectiveness of vaccines, and their accessibility to vaccines and information at the time when vaccines were readily available to the public. For the Twitter study, we focused on metropolitan and nonmetropolitan communities in the United States and examined tweet sentiment and emotion scores in the early stage of the pandemic and through the public release of the first COVID-19 vaccines.

Methods: For the survey data, a total of 900 survey responses were collected in two rural counties in the State of California from March 30 to April 25, 2021. The survey was offered via web and phone in English, Spanish, Punjabi, and Hmong. The respondents were asked about their perceptions of COVID-19 vaccines, messaging, and sources of information. For the Twitter data, we used 127,648 tweets for the analysis after data cleaning, reverse geocoding of tweets, and assigning geographical designations to compare public perception between metropolitan and nonmetropolitan areas. We quantified public perception using the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon to calculate sentiment scores and the NRCLex (National Research Council Canada Lexicon) to calculate emotions scores for the tweets. Next, we explored patterns in public perception of COVID-19 vaccines from March 11, 2020, to September 12, 2021. Then, we compared public perception between two separate periods (i.e., before and after December 11, 2020, when the Food and Drug Administration issued an emergency use authorization of Pfizer, the first COVID-19 vaccine).

Results: In the survey approach, 41% of respondents were Latinx. The most frequent concerns noted for COVID-19 vaccine hesitancy were lack of confidence in the vaccine and the state and federal government (46-56%). However, complacency about the seriousness of the COVID-19 vaccines and disease (35%) and convenience or issues in access, travel time, and cost of vaccines (20%) were not associated with decisions regarding COVID-19 vaccination. In the Twitter approach, we found that public sentiment and emotion varied by geography though our findings did not significantly differ for metropolitan and nonmetropolitan residents. Fear was prevalent in the early times when COVID-19 was announced as a pandemic. However, this was quickly taken over by the emotion of trust later as the breakthroughs in COVID-19 vaccines were announced. Specifically, trust peaked on November 9, 2020, when Pfizer announced its vaccine was 90% effective. Then, around December 11, 2020, positive and negative tweet sentiments started diverging more clearly than the extreme sentiment fluctuations before this period.

Conclusions: For urban or rural and metropolitan or nonmetropolitan communities, news and social media are potent outlets for health information and can significantly change the public's perceptions about COVID-19 vaccines. The survey data shows rural residents in the Central Valley of California, predominantly Latinx, have high confidence or trust in healthcare providers, and the county public health department. However, approximately 40% of these rural residents were still unlikely to get vaccinated, similar to rural populations throughout the country. Recommendations to combat COVID-19 vaccine hesitancy amongst Latinx rural residents include leveraging trusted sources such as local doctors, family/friends, and local public health departments to encourage vaccination amongst this population. In addition, Twitter data shows that announcements from the public media or private institutions appear associated with the public's perception of COVID-19 vaccines, such as the first news of the effectiveness of

the Pfizer COVID-19 vaccine or the notice of blood clot issues caused by the Johnson & Johnson COVID-19 vaccine. Also, there was no significant difference in the mean sentiment or emotion scores between geographical distributions from March 2020 to September 2021. Overall, COVID-19 vaccine news appears to penetrate the public whether people are in urban or rural and metropolitan or nonmetropolitan communities.

CHAPTER 1. Introduction

Introduction

Vaccine hesitancy has been one of the top ten global health threats since 2019 deemed by the World Health Organization (WHO). [1] Vaccine-preventable disease, not including COVID-19, has killed an estimated 3 million people per year worldwide over the past decade. [1] If more people had been vaccinated, 1.5 million more deaths per year worldwide could have been prevented. [2] The COVID-19 pandemic has resulted in numerous deaths and affected lives worldwide in unprecedented ways. There have been more than a total of 6.34 million deaths worldwide and around 1.01 million in the United States as of July 1, 2022. [3] Moreover, it has been estimated that COVID-19 vaccines could have prevented an additional 235,000 deaths in the United States from December 1, 2020, to September 30, 2021. [4]

Vaccine hesitancy is measured on a continuum ranging from full vaccine acceptance to full vaccine rejection, with some acceptance, delay, and refusal in between. [5] People continue to challenge the effectiveness of vaccinations, [6] seek alternative vaccination schedules, [7] and may refuse vaccination for themselves and their children. [8] In the United States, the skepticism about the legitimacy of health information in the context of the COVID-19 pandemic has also been linked to geographical location.[9,10] Early studies have seen United States counties in rural areas appear to be more skeptical about the legitimacy of COVID-19 health-related information. [9,10] Despite the overwhelming evidence of the benefits of vaccines, [11–14] there is still a growing public vaccine hesitancy in these areas. [15,16]

Vaccine hesitancy has heightened attention to people's doubts and concerns, mainly as expressed in social media, regarding vaccine safety.[17] Other evidence indicates that negative perceptions contribute to vaccine hesitancy. [5,18] Emotions and sentiments can partially reflect perceptions and are only part of people's complex psychological states. Emotions can often be categorized into eight universal primary states, anger, anticipation, disgust, fear, joy, sadness, surprise, or trust. [19] Sentiments, on the other hand, are defined as a positive, neutral, or negative mental attitude that can capture complex social states. [19] Both emotions and sentiments may lead to or reflect underlying beliefs in lack of confidence in vaccines, strong negative attitudes towards vaccines, mistrust in authorities, and complacency or overall perception that vaccines are unnecessary. [18]

This dissertation focuses on this evidence of geographical factors, such as rural or nonmetropolitan designation, and perception factors, such as emotionality and sentimentality, that may shape public attitude and behaviors towards COVID-19 vaccination. This research specifically examines communities in the United States and their perception of COVID-19 vaccines as expressed in both geographical and digital space via surveys and via Twitter during the year 2020-2021.

Summary of the Literature

Topic 1: Vaccine Hesitancy and Sources

Vaccine Hesitancy. The World Health Organization Strategic Advisory Group of Experts on immunization (SAGE) Working Group model [5] describes vaccine hesitancy as a continuum ranging from full vaccine acceptance to full vaccine rejection, with some acceptance, delay, and some refusal in between. Their model attributes vaccine hesitancy to the “3 C’s”: confidence, complacency, and convenience.[5] Lack of confidence is related to mistrust in authorities and strong negative attitudes towards vaccines, as exhibited by those in the anti-vaccination movement. [18] High complacency is associated with the perception that vaccines are not necessary and diseases are low risk. [18] Low convenience is related to the physical and geographical accessibility, travel time, and costs of vaccines. [18] Authors such as Razai et al. have also proposed in their “5 C’s” of vaccine hesitancy on additionally tackling the issue of communications, or the misinformation about vaccines that drives fear and anxiety in getting vaccinated, and context, or people’s sociodemographic, religion, ethnicity, an occupation which appear to be factors that may lead to vaccination uptake or lack thereof. [20]

The general population receives health and medical information from various sources. These sources have their own perceived credibility and subsequently may affect decisions about their healthcare. They can either be beneficial when accurate information is presented to the public or detrimental when sensationalism or extravagant claims about medical findings are made. [21] Nevertheless, to keep the public accurately informed and the community healthy, a key factor is identifying sources most likely to affect health behavior and decisions, such as vaccination. Given that health professionals continue to be the most trusted source of vaccine information and discussion of vaccination acceptance or hesitancy remains to be an important routine task for health care professionals, [22–26] primary care physicians and other health care providers must have accurate information regarding vaccination and tools to communicate with parents. Dorell et al. [24] used surveys to determine the factors influencing parents to vaccinate their adolescents. The study found that health care provider recommendations were the most influential factor for vaccinating. Other sources of influence reported were school requirements (46.4%), news coverage (31.3%), family (31.0%), friends (27.9%), magazines (25.6%), internet (17.2%), television (15.7%), and drug companies (12.0%). Individual vaccination coverage was higher for parents who reported receiving recommendations from a health care provider. Also, compared with parents who had unvaccinated teens, a higher proportion of parents with vaccinated teens reported that their doctors recommended vaccines and were key to making their vaccination decisions.

Topic 2. Geographical and Digital Space

In the United States, it has been recognized that nonmetropolitan communities have a higher incidence of COVID-19 cases and mortalities, and lower COVID-19 vaccine coverage. [27] This discrepancy appears related to nonmetropolitan residents' skepticism of the COVID-19 vaccines. [9,10] A large-scale survey conducted by Larson et al.[28] examined the state of public trust in vaccinations and studied individual perceptions of vaccines' importance, safety and effectiveness. Their findings suggest that groups with high rates of negative vaccine sentiments appear related to individuals with more access to health services, specifically in the metropolitan regions in the United States. This finding is related to Coronado's [29] study where their analysis shows that densely populated areas, such as metropolitan regions in the United States, such as Los Angeles County and Sacramento County, served as the main hubs for vaccine information and sentiment spread. Vaccine hesitancy models, such as Razai et al.'s [20] and Perretti-Watel's, [30] extend the WHO model to further explore the great structures that appear connected to an individual's vaccine behaviors and sentiments. [15] These models help emphasize the importance of considering individual's social determinants of health in understanding vaccine behaviors and geographical differences are one of these great structures recommended to be explored. [15] It is important to understand geographical differences including metropolitan and nonmetropolitan communities' perception of COVID-19 vaccines in the literature, especially as the discrepancies between these geographical designations remain quite apparent. [27]

The platform for health information from different media, including online websites, television, and social media, is essential to look at as well. These platforms of vaccine information dissemination can determine the mass distribution of accurate or harmful information. This is important as evidence shows people rely on the internet when deciding whether to vaccinate themselves or their children. [26,31] Social media platforms contain discussions from parents claiming vaccines have harmed their children, more so than online news websites. [32] Tang et al. [33] conducted a systematic review of the literature on public opinion, the accuracy of the information, and the methodological use of social media to study emerging infectious diseases, including vaccine-preventable ones. The authors show that Twitter was the most scrutinized and most often used platform as a search strategy for researching infectious diseases (40%). This was followed by YouTube (14%), Facebook (14%), blogs (14%), forums (7%), and other social media sources (11%). Before the COVID-19 pandemic, the most researched diseases were H1N1 or swine flu, Ebola virus, and H7N9 or avian flu/bird flu 2018. The authors also highlighted how commercial corporations and governmental organizations, such as the CDC, WHO, and United States Department of Health & Human Services, used Twitter as a means of one-way communication of disease updates, policies, prevention, official action, general information, and scientific research to the public. Thus, information-oriented social media, such as Twitter, should not be ignored as they are sometimes the first line of a news source for the general public. [34] The research findings from this platform are vital in providing healthcare professionals the tools needed to understand public views and accurate responses to concerns and doubts about vaccinations in real-time. [35] Twitter offers a unique feature, among other social media platforms, in that it allows for real-time data streaming on health attitudes and opinions. As the data are being streamed, they can rapidly be transferred into the data cleaning and preparation process, such as emotion and sentiment capture, and immediately be ready for analysis. [36]

Topic 3: Emotions and Sentiments

Emotions play an important role in communication, specifically in persuasion. For example, authorities, such as healthcare professionals and government officials, can enhance their credibility when using emotionally intense language. In contrast, less trusted sources can enhance their appeal by using low intensity in their messages. [37] The language expectancy theory (LET) posits that the use of language is all based on a cultural norm and can directly affect a person's behavior. Specifically, persuasion with greater emotional intensity in words appears associated with the receivers' perceptions and decisions. [37] In conjunction with the emotional intensity of the words, the perceived credibility of the source affects the trust and persuasiveness of the message, according to LET. Below are instances where LET has provided a useful frame for exploring how emotionality can affect the spread of information during the COVID-19 pandemic and persuade vaccination attitudes in social media.

Emotions and sentiments in online content can appear connected to the propagation of information at scale. [38–41] Li et al. [38], Himelboim et al. [39], and Chen and Dredze showed how the emotionality and sentimentality of a post's content could be an important predictor of how likely the tweet will spread.[39,40] Li et al. [38] found that other factors, along with emotions and sentiments, are responsible for disseminating information related to COVID-19. Along with the number of followers and who the users followed, the emotions of the content posted had the most considerable relation to the virality of the post. [41] If the user resided in a developed city, they would likely have more social influence online and more social capital. [38] Also, posts containing criticisms or casting doubt were more likely to be reposted if the users used fewer negative words ($b=-0.01$, $p<0.05$) or reside in urban cities in the United States ($b=0.023$, $p<0.05$). [38] The longer the length of the content and the less negative emotion is included, the more likely it is to be reposted. [38]

Emotion states, such as trust, appear associated with compliance with preventative behaviors in public. Brodeur et al. [9] wanted to better understand this phenomenon by studying the relationship of various factors, such as trust and social capital, on compliance with the shelter-in-place order in the United States. The authors relied on public smartphone data measuring average distances traveled as a proxy for shelter-in-place compliance. Brodeur et al. [9] also used data from the 2000-2014 General Social Survey, including American confidence, societal morality, trust, and social capital variables. Specifically, the authors focused on questions related to geography and trust in governmental institutions, medicine, science, and the press. Engle et al. [10] studied the same concept of trust by exploring public smartphone data points measuring average distances traveled as a proxy for shelter-in-place compliance from February 24, 2020, to March 25, 2020. [9,10] Overall, Brodeur et al. [9] and Engle et al. [10] found that shelter-in-place limits the mobility of the public, but areas with higher trust in authorities were less mobile than areas with less trust in authorities. Further, as shown from the results of the General Social Survey, the impact of lockdowns revealed that the public rated the survey placing more trust in the press and media compared to trust in healthcare and government authorities ($b=-0.12$, $p<0.05$). [9]

Both Xu and Guo [42] and Broniatowski et al. [43] studied how professional sources of vaccine information, including healthcare professionals and government officials, can focus on using positive emotions in their messages to drive long-term gains in vaccination to the public. They can also help identify rumors about vaccination that spread slowly but remain prevalent. [42] Xu and Guo [42] and Broniatowski et al. [43] found that, compared to pro-vaccine messages, anti-vaccine messages had higher use of emotionality and negative sentiments, specifically fear, anger, and sadness. [42] Messages that contain more negative emotions and distrust in authorities were found to be less favored (mean=57) compared to messages with positive emotions (mean=230.5). [42] Both studies also showed how emotions and sentiments, spread by unverified and untrustworthy sources, appear associated with the public's perception of vaccinations. [42,43] These implications are important as they can guide how the public responds to vaccine messages and how authorities in vaccine information can properly establish strategies to counter vaccine hesitancy and drive positive behaviors. In addition, characterizing perceptions in vaccine messages can help health authorities to identify successful approaches to promote positive preventative behaviors.

Study

There is evidence that perceptions, reflected by emotionality and sentimentality in messages, appear related to driving vaccine hesitancy behaviors in public.[9,10] Studies have explored the importance of emotions like trust and positive sentiments as they appear associated with individual preventative health behaviors during the COVID-19 pandemic. [9,44] There is also evidence that perceptions, in the context of emotions and sentiments, appear associated with information propagation at scale, even during the COVID-19 pandemic. [38,39,45] Several studies have shown that emotions in the content shared online appear connected to the extent and magnitude of the spread of the message to the public. [38–40] Specific emotion states such as trust, fear, anger, anxiety, awe, sadness, and positivity also appear related to the propagation of vaccine-related information. [38–40,42] Exploring perceptions, as reflected by emotions and sentiments, may not be enough as the literature suggests the importance of also taking into account powerful structures, including geographic location, in holistically understanding individual health decision-making. [5,20,30] At a large scale, there is still a need to understand geographical differences in preventative health behaviors such as the metropolitan and nonmetropolitan communities' perception of COVID-19 vaccines as health discrepancies in these communities can be observed. [27]

Overall, previous studies have analyzed public perceptions about vaccines as reflected by emotions and sentiments via surveys and Twitter. [40,43] However, to the best of our knowledge, there is little to no evidence on how these perceptions translate to specific preventive behaviors, such as vaccination, and what the effects of geographical location are. Specifically, how emotions and sentiments shape the public's attitudes about vaccination in the COVID-19 pandemic in metropolitan and nonmetropolitan areas remains unknown.

CHAPTER 2. Rural Consumers' Perspectives on COVID-19 Information

Background

Around 97 percent of the United States land mass is rural, and 20 percent of the population, or approximately 60 million people, reside there.[46] Most residents of rural counties in the United States (50-75%) are not fully vaccinated and are more likely to be exposed to the COVID-19 virus.[15] COVID-19 infection rates have been disproportionately higher in the Latinx population, or the Hispanic-Americans, in the United States, compared to the general population.[47,48] There is concern that the factors contributing to vaccine hesitancy in this population in rural areas in the United States are also not well characterized or understood.[15,16,49]

The World Health Organization (WHO) vaccine hesitancy model[5] breaks down vaccine hesitancy into "3 C's": confidence, complacency, and convenience.[5] Lack of confidence relates to mistrust in authorities and strong negative attitudes towards vaccines.[18] High complacency is the perception that vaccines are unnecessary and diseases are low risk.[18] Low convenience involves the physical and geographical accessibility, travel time, and costs of vaccines.[18] While the WHO vaccine hesitancy helps explain factors for an individual's vaccine hesitancy, other vaccine hesitancy models, such as Razai et al. (2021) "5 C's" of vaccine hesitancy [20] and Perretti-Watel's (2015) model [30] expands the WHO model to further understand the structures around an individual that may affect their hesitancy to be vaccinated. These factors include culture, religion, economic status, and community, such as rural or urban, which appear to be associated with individuals' decisions for vaccination. [15]

The Latinx community has been reported to have a higher rate of lack of confidence in healthcare and medical institutions secondary to immigration status and English proficiency to navigate the healthcare system.[16,49] This, in turn, reportedly has adversely affected health outcomes for this community with increased hospitalizations and emergency department visits.[50]

High complacency has also been observable for under-represented populations such as the Latinx community in rural areas due to language barriers, lower education levels, and lower-income levels. Health literacy and education are strong determinants in a person's complacency or refusal of vaccinations.[18] Existing literature has shown that lower knowledge about health information correlates with a higher likelihood of spreading misinformation and lower vaccination uptake.[51] Convenience and access to healthcare services are issues for the Latinx community in rural areas due to socioeconomic and geographical factors.[16,52] Latinx groups make up a substantial proportion of critical front-line and low-wage occupations such as healthcare support and food services, construction and maintenance, and production and transportation occupations.[53] Even before the pandemic, few in the Latinx community held jobs that allowed them to socially distance themselves and work from home.[54] Additionally, the Latinx community, compared to non-Hispanic whites, have lower income levels, a stable source of jobs, and healthcare insurance coverage.[55] The precarious work and financial conditions predispose the Latinx community to increased COVID-19 exposure and extreme inconvenience or access to healthcare services.

This topic is particularly urgent as new coronavirus variants have recently emerged and are already spreading rapidly.[56] The new variants are more transmissible and have increased immune system evasion, especially for unvaccinated people.[56] Experts suggest that populations with low vaccination rates will likely see larger outbreaks, putting the rural areas at a higher risk for the new variants if residents remain unvaccinated.[56] This is especially troubling as rural residents, compared to urban residents, tend to be less willing to get the COVID-19 vaccine, and the Latinx community appears to experience increased levels of COVID-19 exposure.[15,16] There has been little literature on this topic with rural and under-represented groups such as the

Latinx population, who disproportionately experience health and socioeconomic disparities.[57] What is still unknown are the views of rural residents in California around COVID-19 vaccines, precisely the level of trust in the source and content of the vaccine information they received, their view of the safety and effectiveness of vaccines, and their accessibility to vaccines and information at the time when vaccines have been readily available to the public. The purpose of the current study was to examine the COVID-19 vaccine perspectives of the rural population in the Central Valley of California, which is predominantly Latinx and includes smaller yet important minority groups that are poorly represented in the existing literature.

Methods

A quantitative survey was conducted from March 30 to April 25, 2021, with residents from two specific rural counties in the State of California and adults 18 years and older. The help of a market research and data analytics firm was employed to conduct the survey. The research firm used automated methods to dial 44,102 phone numbers, send 31,406 invitation link emails, and send 19,684 text invitation links yielding a total of 900 respondents.

The survey was administered anonymously via a mix of hybrid email-to-web, text-to-web, and phone. For the email-to-web method, a random sample of United States voters with email data was used and was sent an email invitation link to take the survey online. For the text-to-web method, a random sample of United States voters with mobile phone data was used and was sent a text invitation link to take the survey online. The phone survey was also used, specifically live phone surveys were conducted by trained, professional interviewers via landlines and mobile phones. The survey was offered in different languages, with English (85%) being the most used, followed by Spanish (15%). Punjabi and Hmong languages were also offered because of the prevalence of these languages in the two counties. The email-to-web and text-to-web methods had a toggle on the screen for the language of preference, while the phone method had bilingual interviewers available for the language preference.

The survey was validated by pre-testing the questions with a few sample residents (See Appendix II for complete questions). Questions to examine confidence include asking respondents what are their most and least trusted sources of COVID-19 vaccine information and if they are convinced that COVID-19 vaccines are safe, effective, and have dangerous side effects. Questions to examine complacency include asking the respondents if they believe that COVID-19 vaccines are unnecessary if they already had COVID-19, if they are convinced that the COVID-19 vaccine is needed for further protection from the virus, COVID -19 vaccines prevent them from getting dangerously sick with COVID-19, and if they get COVID-19, they could spread the disease to friends, family, and others around them. Questions to examine convenience include asking respondents if they have physical or geographical accessibility to get COVID-19 vaccines if they are aware COVID-19 vaccines are free for everyone, if they have no issues taking time off work or other responsibilities to get the COVID-19 vaccine, and if they would see a doctor more often if they could do it by phone or video conference.

For questions about most and least trusted sources, a list of people and organizations was read, and respondents had to answer if they trust that source as a reliable source of COVID-19 information on a scale of a great deal, a fair amount, not too much, or not at all. For questions about COVID-19 vaccines safety, effectiveness, side-effects, and necessity, a list of statements about COVID-19 vaccines were read and respondents had to answer how convincing the statements in a scale of very convincing, somewhat convincing, not too convincing, or not at all convincing. For questions about COVID-19 vaccines accessibility and platform for seeing a doctor, a list of statements about COVID-19 vaccines accessibility was read and respondents had to answer if they agree with the statements on a scale of strongly agree, somewhat agree, somewhat disagree, or strongly disagree. Before the analysis of data, these scales were further simplified from four categories into two categories of trust or do not trust, very or not at all convincing, and total agree or disagree.

Sixty-one close-ended questions were asked over a span of 15-minutes, including questions about the respondent's demographics, concerns about COVID-19, perception of the safety and effectiveness of COVID-19 vaccines, and their preferred platform and source for COVID-19 information. For the data analysis, we used descriptive statistical measures, including frequencies and distributions, which were then used to examine the resident's views of the COVID-19 vaccines.

Results

There was a total of 900 respondents, with 41% identifying as Latinx (See Table 1). The population was mostly adults 18-54 (63%). There were more Democrats (37%) than Republicans (20%) amongst the respondents, and about one-third identified as neither (33%). Most graduated high school or completed their undergraduate studies (84%). A similar proportion of respondents had a household income below $61,000 (29%) and at or above $61,000 (25%).

The participants responded to a variety of questions (See Appendix II for complete questions). 53% reported being not vaccinated or will not get vaccinated (See Table 2). Among unvaccinated respondents, 58% were more likely to get the vaccine when it became available, while 40% were unlikely to get it. The most common reasons reported by those that were unvaccinated and unlikely to get the vaccine included: the vaccine is too new or not tested enough (18%), they do not want it or do not take vaccines (16%), they are healthy (15%), they concerned about side-effects (13%), and they do not trust vaccines (9%). Those that were unvaccinated and likely to get the vaccine reported that they would get the vaccine at their doctor's office (30%), the pharmacy (18%), the hospital (13%), or at a public vaccination site (11%). Most respondents perceived COVID-19 vaccines as safe (73%), effective (75%), and free for everyone (84%) (See Table 3).

Most respondents (76%) agreed that it is easy to find information about when they can get a COVID-19 vaccine (See Table 4). The most trusted sources of reliable information about COVID-19 amongst the respondents were their doctor or healthcare provider (83%), followed by friends or family that have already received the vaccine (77%), and the County Public Health Departments similarly (71%). The most minor trusted sources of reliable information about COVID-19 included their employers (58%), the Biden administration (56%), State government officials (51%), and religious or faith leaders (44%). The most preferred platform for receiving COVID-19 testing and vaccine information is email (31%), the Department of Public health website (27%), text message (19%), and social media (17%). Many (57%) would see a doctor more often if they could do it by phone or video conference.

Discussion

COVID-19 Vaccine Confidence

Lack of confidence or mistrust in authorities and strong negative attitudes toward the COVID-19 vaccine are apparent among the vaccine-hesitant. California was the first state in the country to issue stay-at-home and quarantine orders at the beginning of the pandemic. [58] Throughout the past two years, the state had removed stay-at-home orders when COVID-19 cases started to decline and re-issued the orders when cases began to spike again several times over. [58] As of January 2022, the Latinx population accounts for 49.3% of the state's COVID-19 cases and 45.3% of the state's COVID-19 deaths. [59]

This study showed a supermajority agreeing that the vaccines are generally safe (73%) and effective (75%). This finding parallels the messaging most respondents found convincing, including the message that COVID-19 vaccines are safe and work, and dangerous side effects are extremely rare (76%). Respondents also found the message that vaccines were not rushed and were based on reliable science (72%) very convincing. Despite the resounding agreement about the safety and effectiveness of these vaccines in this study, the COVID-19 vaccination status of the residents of these rural counties tells a different story, with only less than half of the population (48%) being vaccinated or partially vaccinated as of April 2021. Looking back at the bigger picture, around two out of ten rural residents in these California counties, were unlikely to get vaccinated. This is comparable to existing studies where three out of ten rural residents in the country say that they will definitely not get vaccinated or only if required [15] or one-third of the population in the country are unsure of getting vaccinated as of April 2021.[60] Similar questions have been asked in the 2021 Kaiser Family Foundation (KFF) survey. However, in the KFF survey, the population of rural residents in the US was far less convinced in the messaging that COVID-19 vaccines are safe and effective (19%), the long-term impact of COVID-19 could be worse (16%), and scientists have been working on the technology used in vaccines for the past decades (24%).[15]

Different factors contribute to residents' confidence in the COVID-19 vaccine and the population's vaccination status. In this study, half were not vaccinated (50%), and a few said they would not get vaccinated (3%). In this subpopulation of people who were not vaccinated or will not get vaccinated, less than half (40%) were unlikely to get vaccinated because the vaccine is too new/not tested enough/not FDA approved (18%), or does not trust vaccine/conspiracy (9%). These findings generally reflect similar results of the KFF survey, where one of five rural adults in the US say they will not get vaccinated because it is too new/not enough information (19%) or do not trust vaccines (12%).[15]

Confidence in vaccine information can also be based on where the information is coming from. For example, respondents say that for COVID-19 vaccine information, they trust their own doctor or healthcare provider (83%), followed by friends or family that have already received the vaccine (77%), the Centers for Disease Control and Prevention (CDC), and the County Public Health Departments similarly (71%). Similarly, the respondents of this survey said they were less likely to trust COVID-19 information from the Biden Administration (56%), and a poll of US residents trusted the federal government the least (46%). These findings were generally aligned with results from polling US residents, where health officials (84%) and local departments (72%) were the most trusted source of health information, and the federal government (46%) was the least trusted source.[61] Local doctors, family/friends, the CDC, and Local Public Health Departments will be key to sharing messages to improve vaccination rates among the residents. In addition, these trusted groups are important in stressing the shared responsibility of disseminating timely and accurate information about the safety and efficacy of vaccines. [26,28] While federal/state politicians, religious leaders, and employers are the least trusted groups in getting vaccine information, thus lower confidence should be placed in using these groups to disseminate information about COVID-19 vaccines.

COVID-19 Vaccine Complacency

High complacency, or the perception that vaccines are unnecessary, was not as evident in this study. There is varying consensus amongst the respondents about the necessity and risks of the COVID-19 vaccines. Around one-third of the population believed the COVID-19 vaccine was unnecessary if they already had COVID-19 (35%). However, a majority of the same population were also convinced that the COVID-19 vaccine is needed for further protection from the virus and to prevent the spread to people close to them and COVID-19 vaccines prevent them from getting dangerously sick with COVID-19 (76%), and if they get COVID-19, they could spread the disease to friends, family, and others around you, but being vaccinated can protect those around them too (79%). These findings are quite different from similar questions from rural residents in the whole country, where less than half of the population state that The seriousness of coronavirus in the news has generally been exaggerated (44%) and that They are worried they or a family member will get sick from the coronavirus (40%).[15]

The respondents are also split on the gravity of the side effects of the vaccine, with almost half agreeing that COVID-19 vaccines have dangerous side effects (39%) while the other half (56%) do not agree. Most of the respondents of this survey, however, think that the long-term health impacts of COVID-19 are much more serious than the chance of having side effects from the vaccine (74%). These findings were dissimilar from the findings of other studies. For example, Kirzinger et al. (2021) found that more than half of the population think they might experience serious side effects (64%), while a small minority think while the long-term effects of the vaccine may be unknown, the long-term effects of COVID-19 could be worse (16%).[15] Of importance is that Kirzinger et al. (2021) study interviewed around 1,000 residents in rural areas across the United States, while this study focused on the state of California, and their sample had a smaller representation of the Latinx population (15%) compared to this study (41%). [15] Looking back at the sub-population of residents interviewed in this study, they were unlikely to get vaccinated as a whole as a few said they are healthy/do not need it/give it to others first (15%), does not work/ineffective/can still get COVID (6%), or COVID-19 is not that serious/has high survival rate (3%). This is a similar finding from the Kirzinger et al. (2021) survey, where one of five rural adults in the US say they will not get vaccinated because they believe the vaccine is not effective against COVID-19 (8%), do not need it (5%), or they are not concerned about COVID-19/getting sick (5%).[15]

Most of the respondents (76%) in this study agree that it is easy for them to find information about when they can get a COVID-19 vaccine, and their preferred platforms include email (31%), the Department of Public health website (27%), text message (19%), and social media (17%). In addition to trusted groups of people sharing COVID-19 information, specifically healthcare professionals and government officials, it will also be important to share information that is evidence-based through online platforms. A combination of these strategies can be especially effective for minority groups like the Latinx population, who have historically had a lack of trust in healthcare establishments and have language barriers, shown to be related to lower health literacy and high complacency with COVID-19 vaccine uptake.[16]

COVID-19 Vaccine Convenience

Low convenience or physical and geographical accessibility, travel time, and costs of vaccines are among the factors of vaccine hesitancy. A majority of the respondents of this study are aware that COVID-19 vaccines are free for everyone (84%) and have no issues taking time off work or other responsibilities to get the COVID-19 vaccine (80%). Of the people in rural California that are not vaccinated (50%), more than half of them (58%), or around (29%) of the total population, are likely to get the vaccine when it is available to them. This finding is close to the proportion of rural residents in the US who state they will get the vaccine as soon as possible (16%).[15] Regarding physical access, most say it is easy for them to get to a place where they can get a COVID-19 vaccine (82%). The top location where they will go to get the COVID-19 vaccine is at the doctor’s office (30%), a pharmacy (18%), a hospital (13%), or a public vaccination site (11%). These same locations have been the main preference for the Latinx population in other studies as well, with perceived access being higher with doctors and healthcare providers (71%), vaccine locations (67%), or hospitals (66%).[15]

In this study, geographical accessibility, travel time, or costs of vaccines were not factors that prevented residents from getting the COVID-19 vaccine. Yet, almost half of the respondents (46%) have put off getting medical care because of COVID-19. This is similarly reflected nationwide, where around (41%) of US adults have put off getting medical care, including both emergent and routine care, during the pandemic.[62] This is where digital access may be more important, with many stating that they (57%) would see a doctor more often if they could do it by phone or video conference, and most (54%) have used a telehealth appointment to see a doctor or get medical advice before. This preference for phone or video conferences during the COVID-19 pandemic is consistent with the existing literature that shows increasing trends in interest in telehealth and telemedicine during the pandemic in the US.[63] Among the platforms to get information, the public has had an increasing preference for using telehealth or phone/video conference, which can further increase the convenience of getting vaccine information for rural residents and further reduce temporary disruptions in accessing medical care.

Limitations

This survey was conducted between March 30 to April 25, 2021, while vaccinations were only fully opened to all Californians ages 16 and above starting April 15, 2021. Caution must be applied when interpreting the ~53% non-vaccination rate, as access to vaccines at this time may have played a part in the respondent’s vaccination status. The survey was conducted in only two counties in rural California which may not be fully representative of rural California or rural America. Our sample had a representation of Latinx residents (41.1%) similar to the population of the State of California (39.4%)[64] but not representative at the national level (18.5%). However, our sample did include mostly adults 18-54 (63%), close to the national level (55.3%). Respondents were mostly male (49.7%) which is similar to the national level (49.5%). Most respondents graduated high school or completed their undergraduate studies (84%) also comparable to the national level (88.5%). A proportion of respondents had a household income below $35,999 (15.7%) which is close to the national level (11.4%) as of July 2021.[65]

Conclusion

This study explored the World Health Organization (WHO) vaccine hesitancy model[5] framework to better understand the rural perspectives about COVID-19 vaccines in early 2021 in the United States. There was high confidence or trust in COVID-19 vaccines, healthcare providers, and county public health departments amongst respondents in this study. However, two out of ten rural respondents in these counties were still unlikely to get vaccinated, which was aligned with findings of the general rural residents in the country. Complacency or perception COVID-19 vaccines are needed were not apparent in this study. Most of the rural respondents in this study believe the seriousness of the effects of COVID-19 vaccines in preventing sickness amongst themselves and those around them, which differed from the general rural population in the country who believe otherwise. Similarly, convenience or physical and geographical accessibility to COVID-19 vaccines was not a factor in this study, as most rural respondents were aware the vaccines were free or had no issues taking time off to get the vaccine.

In most of the rural and minority populations in California, healthcare providers and family members or friends play an important role in sharing positive messages about COVID-19 vaccines and stressing the importance of being vaccinated to protect themselves and those around them. The Public Health Departments at the county level are key to providing proper and accurate guidelines to ensure the public's safety from the COVID-19 pandemic. The continued use of online platforms and direct communications such as email, the Department of Public health website, text messages, social media, and telehealth approaches appear related to improving confidence and decreasing complacency about COVID-19 vaccines. The focus on leveraging key groups and improving digital health resources are important and appear connected to improving COVID-19 vaccination rates in this pandemic.

Appendix I. Tables and Figures

Table 1. Respondent Characteristics.

Characteristic	n (%)
Gender	
Male	447 (49.7%)
Female	425 (47.2%)
Prefer not to respond	28 (3.1%)
Age	
18-34	254 (28.2%)
35-54	312 (34.7%)
55-64	136 (15.1%)
65 or over	187 (20.8%)
Prefer not to respond	11 (1.2%)
Political Affiliation	
Democrat	329 (36.6%)
Republican	183 (20.3%)
Neither one	294 (32.7%)
Prefer not to respond	94 (10.4%)
Ethnicity	
White or Caucasian	511 (56.8%)
African American or Black	26 (2.9%)
Asian or Pacific Islander	33 (3.7%)
Biracial or Multiracial	56 (6.2%)
Something else	164 (18.2%)
Prefer not to respond	110 (12.2%)
Hispanic	
Yes	370 (41.1%)
No	475 (52.8%)
Prefer not to respond	55 (6.1%)
Education	
Some grade school or high school	58 (6.5%)
Graduated high school or Technical/Vocational school	220 (24.5%)
Some college or Less than 4-year degree, Graduated college/4-year degree (BA, Bachelor)	481 (53.4%)
Graduate/Professional (MA, Master's, PhD, MBA, Doctorate)	103 (11.4%)
Prefer not to respond	38 (4.2%)
Household Income	
Below $35,999	141 (15.7%)
From $36,000 to $60,999	122 (13.6%)
At or Above $61,000	228 (25.3%)

Prefer not to respond	409 (45.4%)

Table 2. COVID-19 vaccination status.

Items	n (%)
COVID-19 Vaccination Status	
Not vaccinated	450 (50%)
Partially vaccinated	171 (19%)
Fully vaccinated	234 (26%)
Will not get vaccinated	27 (3%)
Have appointment scheduled	27 (3%)
Prefer not to respond	9 (1%)
(If Not vaccinated) Are you very likely, somewhat likely, somewhat unlikely or very unlikely to get the vaccine when it's available to you?	
Very likely	324 (36%)
Somewhat likely	198 (22%)
Somewhat unlikely	90 (10%)
Very unlikely	270 (30%)
Don't know/Refused	18 (2%)
(If Unlikely to get vaccine) Why are you unlikely to get the vaccine when it's available to you?	
Vaccine is too new / Not tested enough / Not FDA approved	162 (18%)
Does not want it / Does not take vaccines	144 (16%)
Healthy / Doesn't need it / Give it to others first / Unnecessary	135 (15%)
Side effects / Allergic reactions / Dangerous / Long-term effects 13	117 (13%)
Does not trust vaccine / Conspiracy	81 (9%)
Already has or had COVID-19	72 (8%)
Doesn't work / Ineffective / Can still get COVID	54 (6%)
COVID is not that serious / High survival rate	27 (3%)
Does not trust government / Politics	9 (1%)
Ineligible / Does not meet requirements	9 (1%)
Ingredients / Technology	9 (1%)
Other/Don't know/No opinion	81 (9%)
(If Not vaccinated and Likely to get vaccinated) Where do you think you will go to get your COVID-19 vaccination	
At your doctor's office	270 (30%)
A pharmacy	162 (18%)
A hospital	117 (13%)

A public vaccination site	99 (11%)
Other/Don't know/Refused	81 (9%)
Your work or place of employment	63. (7%)
A mobile vaccination site	54 (6%)
A County health clinic	45 (5%)
A Community Health Center	9 (1%)
A supermarket	9 (1%)

Table 3. Perspectives on COVID-19 vaccine safety, efficacy, and impact.

Items	**n (%)**
COVID-19 vaccines are safe	
Total Agree	659 (73.2%)
Total Disagree	214 (23.8%)
Don’t know/refused	27 (3.0%)
COVID-19 vaccines are effective	
Total Agree	679 (75.4%)
Total Disagree	182 (20.2%)
Don’t know/refused	39 (4.3%)
COVID-19 vaccines have dangerous side effects	
Total Agree	351 (39.0%)
Total Disagree	505 (56.1%)
Don’t know/refused	44 (4.9%)
COVID-19 vaccines are free for everyone	
Total Agree	751 (83.5%)
Total Disagree	114 (12.7%)
Don’t know/refused	35 (3.9%)
COVID-19 vaccines are not needed if you already had COVID-19	
Total Agree	310 (34.5%)
Total Disagree	569 (63.2%)
Don’t know/refused	21 (2.3%)
[SAFE] Hundreds of thousands of people were involved in the tests of the approved COVID-19 vaccines to help make sure they are safe and that they work. Since then, millions of Americans have been vaccinated and dangerous side effects are extremely rare.	
Very convincing	686 (76.2%)
Not at all convincing	202 (22.5%)
Don’t know/refused	12 (1.3%)
[NOT RUSHED] Although the vaccine seemed to be created really quickly, it is actually based on science that has been developed over more than 20 years. Scientists simply modified existing vaccines designed for other viruses and made them fight COVID-19 instead.	

Very convincing	645 (71.7%)
Not at all convincing	239 (26.6%)
Don't know/refused	16 (1.8%)
[PROTECT - INDIVIDUAL] COVID-19 can cause serious illness or death, and there is no way to know how the virus will affect you. Vaccines prevent you from getting dangerously sick with COVID-19.	
Very convincing	684 (76.0%)
Not at all convincing	211 (23.4%)
Don't know/refused	5 (0.6%)
[PROTECT - COMMUNITY] If you get COVID-19, you could spread the disease to friends, family, and others around you. Getting vaccinated is not just to protect you, but also to protect our loved ones.	
Very convincing	713 (79.2%)
Not at all convincing	179 (19.9%)
Don't know/refused	8 (0.9%)
[LONG TERM IMPACT] The health impacts of COVID-19 are much more serious than the chance of having side-effects from the vaccine. That's why scientists and doctors encourage everyone to get the vaccine, because the benefits are far bigger than the risks.	
Very convincing	669 (74.3%)
Not at all convincing	218 (24.2%)
Don't know/refused	13 (1.4%)

Table 4. Perspectives on COVID-19 vaccine information.

Item	n (%)
I am able to take time from work or other responsibilities to go get a COVID-19 vaccine	
Agree	715 (79.5%)
Disagree	159 (17.7%)
Don't know/refused	26 (2.9%)
It is easy to find information about when I can get a COVID-19 vaccine	
Agree	687 (76.3%)
Disagree	202 (22.4%)
Don't know/refused	11 (1.2%)
It is easy for me to get to a place where I can get a COVID-19 vaccine	
Agree	738 (82.0%)
Disagree	147 (16.3%)
Don't know/refused	15 (1.7%)
I have put off getting medical care because of COVID-19	
Agree	409 (45.5%)
Disagree	486 (54.0%)

Don’t know/refused	5 (0.6%)
I would see a doctor more often if I could do it by phone or video conference	
Agree	514 (57.1%)
Disagree	375 (41.7%)
Don’t know/refused	11 (1.2%)
And have you or anyone in your household ever utilized a tele-health appointment to see a doctor or get medical care or advice?	
Yes	489 (54.3%)
No	385 (42.8%)
Don’t know/refused	26 (2.9%)
<u>Trust as a source of reliable information about COVID-19</u>	
The Biden administration	
Trust	506 (56.2%)
Do Not Trust	380 (42.2%)
Don't know/Refused	14 (1.6%)
The US Centers for Disease Control and Prevention, or CDC	
Trust	635 (70.6%)
Do Not Trust	259 (28.8%)
Don't know/Refused	6 (0.7%)
State government officials	
Trust	458 (50.9%)
Do Not Trust	429 (47.7%)
Don't know/Refused	13 (1.4%)
County Department of Public Health #1 (name has been blinded)	
Trust	641 (71.2%)
Do Not Trust	240 (26.7%)
Don't know/Refused	19 (2.1%)
County Department of Public Health #2 (name has been blinded)	
Trust	516 (57.3%)
Do Not Trust	297 (33.0%)
Don't know/Refused	87 (9.7%)
Your own doctor or healthcare provider	
Trust	750 (83.3%)
Do Not Trust	143 (15.9%)
Don't know/Refused	7 (0.8%)
Religious or faith leaders	
Trust	394 (43.8%)
Do Not Trust	485 (53.9%)
Don't know/Refused	21 (2.3%)
Your employer	

Trust	524 (58.2%)
Do Not Trust	308 (34.2%)
Don't know/Refused	68 (7.6%)
Friends or family that have already received the vaccine	
Trust	695 (77.2%)
Do Not Trust	191 (21.2%)
Don't know/Refused	14 (1.6%)
Best way to receive information about COVID-19 testing and vaccines	
Email	281 (31.2%)
Department of Public health website	241 (26.8%)
Text message	170 (18.9%)
Social media	152 (16.9%)
Your city website	97 (10.8%)
Regular mail	73 (8.1%)
Phone call	67 (7.4%)
Something else	30 (3.3%)
Don't know/Refused	84 (9.3%)

Appendix II. Full set of questions

Survey of Residents
Stanislaus and Merced Counties
Hybrid Email-to-Web/Text-to-Web/Live Telephone Survey
15 minutes; n=900
FINAL 03/24/21

GREETING: Hello, my name is ________, may I speak with **(NAME ON LIST)**?

INTERVIEWER: ONLY

INTRO: Hello, my name is ________, and I'm conducting a survey for __________ to find out how people feel about issues in California's Central Valley. We are not trying to sell anything and are collecting this information on a scientific and completely confidential basis.

LANGUAGE OF INTERVIEW
1. English
2. Spanish
3. Punjabi
4. Hmong

1. GENDER **(OBSERVED)**
   1. Male
   2. Female

2. First, for confirmation purposes only, which county do you live in?
   1. Merced County
   2. Stanislaus County
   3. Other/(Don't know/Refused) **TERMINATE**

3. Do you feel that things in **[COUNTY FROM Q2]** are generally going in the right direction or do you feel things have gotten pretty seriously off on the wrong track?
   1. Right direction
   2. Wrong track
   3. (Don't know/Refused)

4. What do you think is the most important problem facing **[COUNTY FROM Q2]** today? **(OPEN END, RECORD VERBATIM RESPONSE, ACCEPT UP TO THREE RESPONSE ONLY)**

5INT. Thinking about you and your family, how concerned are you about each of the following? Please use a scale of 1 to 10, where 1 is not at all concerned and 10 is extremely concerned.

(PROMPT IF NECESSARY: How concerned are you about that item, on a scale of 1 to 10, where 1 is not at all concerned and 10 is extremely concerned?**)**
SCALE:

1. 1 – Not at all concerned
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
9. 9
10. 10 – Extremely concerned
11. (Don't know/Refused)

(RANDOMIZE)

5. Personally getting COVID-19
6. Family members getting COVID-19
7. Not being able to get a COVID-19 vaccine

(END RANDOMIZE)

8. When you are around people you don't live with, do you always, mostly, sometimes, rarely or never **wear a mask**?
    1. Always
    2. Mostly
    3. Sometimes
    4. Rarely
    5. Never
    6. (Does not apply)
    7. (Don't know/Refused)

9INT. Whether you personally do these things or not, how safe or unsafe do you think the following actions are right now: very safe, somewhat safe, somewhat unsafe, or very unsafe? **(PROMPT IF NEEDED:** Would you say that is very safe, somewhat safe, somewhat unsafe, very unsafe?**)**

SCALE:

1. Very safe
2. Somewhat safe
3. Somewhat unsafe
4. Very unsafe
5. (Don't know/Refused)

(RANDOMIZE)

9. Spending time <u>indoors</u> with people you don't live with
10. Eating at a restaurant
11. Traveling by public transportation or rideshare
12. Attending a social gathering with people from 3 or more households

(END RANDOMIZE)

13INT. And now, regardless of whether or not you have received the COVID-19 vaccine, please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree that each of the following describes the COVID-19 vaccines.
(PROMPT IF NECESSARY: Would you say you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with that statement?**)**

SCALE:

1. Strongly agree
2. Somewhat agree
3. Somewhat disagree
4. Strongly disagree
5. (Don't know/Refused)

(BEFORE EACH: The COVID-19 vaccines…)

(RANDOMIZE)

13. Are safe
14. Are effective
15. Have dangerous side effects
16. Are free for everyone
17. Are not needed if you already had COVID-19

(END RANDOMIZE)

18INT. Please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with each of the following. **(PROMPT IF NECESSARY:** Would you say you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with that statement?**)**

SCALE:

1. Strongly agree
2. Somewhat agree
3. Somewhat disagree
4. Strongly disagree
5. (Don't know/Refused)

(RANDOMIZE)

18. I am able to take time from work or other responsibilities to go get a COVID-19 vaccine

19. it is easy to find information about when I can get a COVID-19 vaccine

20. It is easy for me to get to a place where I can get a COVID-19 vaccine

21. It is easy to get tested for COVID-19

22. People don't need to be tested for COVID-19 if they've already been vaccinated

23. I have put off getting medical care because of COVID-19

24. I would see a doctor more often if I could do it by phone or video conference

(END RANDOMIZE)

25. Which of the following is your current COVID-19 vaccination status? **(READ OPTIONS 1 THROUGH 3)**
    1. I have not been vaccinated
    2. I have received one dose of a vaccine that needs two doses, or
    3. I am fully vaccinated
    4. (I will not get vaccinated)
    5. (Have appointment scheduled)
    6. (Don't know/Refused)

(ASK IF Q25 = 1, NOT VACCINATED)

26. Are you very likely, somewhat likely, somewhat unlikely or very unlikely to get the vaccine when it's available to you?
    1. Very likely
    2. Somewhat likely
    3. Somewhat unlikely
    4. Very unlikely
    5. (Don't know/Refused)

(IF Q26 = 3 OR 4, ASK Q27)

27. Why are you unlikely to get the vaccine when it's available to you? **(OPEN ENDED, RECORD VERBATIM)**

(RESUME ASKING ALL)

(IF Q25 = 1, NOT VACCINATED AND IF Q26= 1 OR 2, LIKELY TO GET VACCINATED, ASK Q28 AND Q29)

28. Where do you think you will go to receive your COVID-19 vaccination? **(READ LIST IF NECESSARY)**
    1. At your doctor's office
    2. A hospital
    3. The Livingston Community Health Center

4. A mobile vaccination site
5. A pharmacy
6. A supermarket
7. A public vaccination site
8. A County health clinic
9. Your work or place of employment
10. (Other) **(please specify)**
11. (Don't know/Refused)

29. When a COVID-19 vaccine is available to you, will you… **(READ LIST)**
    1. Get any vaccine available as soon as you can, or
    2. Wait until you can get the brand of vaccine you want, or
    3. Something else
    4. (Will not get the vaccine)
    5. (Don't know/Refused)

(RESUME ASKING ALL)

30INT. Next I will read you some statements related to getting a COVID-19 vaccine. After each one, please tell me how convincing that statement is as a reason to get a COVID-19 vaccine– very convincing, somewhat convincing, not too convincing, or not at all convincing. **(PROMPT IF NECESSARY:** Is that a very convincing, somewhat convincing, not too convincing or not at all convincing reason to get a COVID-19 vaccine?**)**

SCALE:

1. Very convincing
2. Somewhat convincing
3. Not too convincing
4. Not at all convincing
5. (Don't know/Refused)

(RANDOMIZE)

30. **[PROTECT - INDIVIDUAL]** COVID-19 can cause serious illness or death, and there is no way to know how the virus will affect you. Vaccines prevent you from getting dangerously sick with COVID-19.

31. **[PROTECT - COMMUNITY]** If you get COVID-19, you could spread the disease to friends, family, and others around you. Getting vaccinated is not just to protect you, but also to protect our loved ones.

32. **[LONG TERM IMPACT]** The health impacts of COVID-19 are much more serious than the chance of having side-effects from the vaccine. That's why scientists and doctors encourage everyone to get the vaccine, because the benefits are far bigger than the risks.

33. **[BACK TO WORK]** Our economy relies on local businesses being open, and people being able to go to work. The safest and quickest way to get our economy moving and local businesses up and running is to get everyone vaccinated as soon as possible.

34. **[SAFE]** Hundreds of thousands of people were involved in the tests of the approved COVID-19 vaccines to help make sure they are safe and that they work. Since

then, millions of Americans have been vaccinated and dangerous side effects are extremely rare.

35. **[NOT RUSHED]** Although the vaccine seemed to be created really quickly, it is actually based on science that has been developed over more than 20 years. Scientists simply modified existing vaccines designed for other viruses and made them fight COVID-19 instead.

36. **[LIVINGSTON MOBILE]** Soon, the Livingston Community Health Center will provide COVID-19 vaccinations at mobile sites in Stanislaus and Merced counties, making it easy for adults throughout our area to get vaccinated.

(END RANDOMIZE)

(ASK IF Q25 = 1, NOT VACCINATED)

37. Given what you've heard, are you very likely, somewhat likely, somewhat unlikely or very unlikely to get the vaccine when it's available to you?
    1. Very likely
    2. Somewhat likely
    3. Somewhat unlikely
    4. Very unlikely
    5. (Don't know/Refused)

(RESUME ASKING ALL)

38INT. Next I'm going to read you a list of people and organizations. For each, please tell me how much you trust that person or organization to provide reliable information about COVID-19: A great deal, a fair amount, not too much, or not at all.
(PROMPT IF NECESSARY: Do you trust **(QX)** a great deal, a fair amount, not too much, or not at all to provide reliable information about COVID-19?**)**

SCALE:

1. A great deal
2. A fair amount
3. Not too much
4. Not at all
5. (Don't know/Refused)

(RANDOMIZE)

38. The Biden administration

39. The US Centers for Disease Control and Prevention, or CDC

40. State government officials

41. **[COUNTY FROM Q2]** Department of Public Health

42. The Livingston Community Health Center

43. Your own doctor or healthcare provider

44. Religious or faith leaders

45. Your employer

46. Friends or family that have already received the vaccine

(END RANDOMIZE)

47. What would be the best way for you to receive information about COVID-19 testing and the vaccines – on the **[COUNTY FROM Q2]** Department of Health website, by email, by regular mail, by text message, by phone call, your city website, social media or something else? **(IF SOMETHING ELSE)** What would be the best way to receive communications? **(RECORD UP TO 2)**

1. **[COUNTY FROM Q2]** Department of Public Health website
2. Email
3. Regular mail
4. Text message
5. Phone call
6. Your city website
7. Social media
8. Something else **(Please specify)**
9. (Don't know/Refused)

48. Have you been tested for COVID-19? If yes, how many times?

1. Yes, once
2. Yes, twice
3. Yes, three or more times
4. No, never
5. (Don't know/Refused)

49. Have you personally had COVID-19?

1. Yes
2. No
3. (Don't know/Refused)

50. And have you ever heard of the Livingston Community Health Center? **(IF YES)** Have you or a member of your household ever visited the Livingston Community Health Center for medical care?

1. Yes, I have visited the Livingston Community Health Center
2. Yes, a family member has visited the Livingston Community Health Center
3. Have heard of the Livingston Community Health Center, but have not visited
4. Have never heard of the Livingston Community Health Center
5. (Don't know/Refused)

DEMOS. My last questions are for statistical purposes only. As a reminder, your responses are completely confidential.

51. In what year were you born? **(DO NOT READ CATEGORIES; CODE AS APPROPRIATE)**

1. 1946 or earlier
2. 1947-1951
3. 1952-1956
4. 1957-1961
5. 1962-1966
6. 1967-1971
7. 1972-1976
8. 1977-1981
9. 1982-1986
10. 1987-1991
11. 1992-1996
12. 1997-2003
13. (Refused)

51B. **[AGE RANGE - CODE FROM PREVIOUS QUESTION]**

[IF Q51=1997 thru 2003 Q51B=1]

[IF Q51=1987 thru 1996 Q51B=2]

[IF Q51=1977 thru 1986 Q51B=3]

[IF Q51=1967 thru 1976 Q51B=4]

[IF Q51=1966 thru 1957 Q51B=5]
[IF Q51=1956 thru 1947 Q51B=6]
[IF Q53 = 1946 or earlier Q53B =7]
(IF REFUSED THEN ASK FOLLOWUP: "Which age group are you in? **(READ LIST)**…"**)**

1. 18-24
2. 25-34
3. 35-44
4. 45-54
5. 55-64
6. 65-74
7. 75 or over
8. (Refused)

52. Do you generally think of yourself as a Democrat, a Republican or neither one?

1. Democrat
2. Republican
3. Neither one
4. (Don't know/Refused)

53A. Do you consider yourself to be of Hispanic, Latino, or Spanish origin?

1. Yes

2. No
3. (Don't know/Refused)

53B. Do you consider yourself to be white or Caucasian, African American or Black, Asian or Pacific Islander, biracial, multiracial or something else?
1. White or Caucasian
2. African American or Black
3. Asian or Pacific Islander
4. Biracial or Multiracial
5. Something else
6. (Refused)

53. **[COMBINED VARIABLE FROM Q53A AND Q53B]**

IF Q53A=1 THEN Q53=1

IF Q53A=2-3 AND Q53B=1, THEN Q53=2

IF Q53A=2-3 AND Q53B=2, THEN Q53=3

IF Q53A=2-3 AND Q53B=3, THEN Q53=4

IF Q53A=2-3 AND Q53B=4, THEN Q53=5

IF Q53A=2-3 AND Q53B=5, THEN Q53=6

IF Q53A=2-3 AND Q53B=6, THEN Q53=7
1. Hispanic/Latino
2. Non-Hispanic White
3. Non-Hispanic African American/Black
4. Non-Hispanic Asian/Pacific Islander
5. Non-Hispanic Biracial/Multiracial
6. Non-Hispanic something else
7. (Refused)

54. What is the last grade you completed in school? **(READ LIST IF NECESSARY)**
1. Some grade school
2. Some high school
3. Graduated high school
4. Technical/Vocational school
5. Some college/Less than 4-year degree
6. Graduated college/4-year degree (B-A, Bachelor)
7. Graduate/Professional (M-A, Master's, P-h-D, M-B-A, Doctorate)
8. (Don't know/Refused)

55. Do you own or rent your apartment or home? **(READ LIST IF NECESSARY)**
1. Own/buying
2. Rent/leasing
3. (Don't know/Refused)

56. Including yourself, how many people live in your immediate household, including any children under the age of 18? **(TEXT BOX)**

57. And for statistical purposes only, what was your household's total income in 2020? **(TEXT BOX) (DK/REFUSE = 9999)**

(IF Q57 = 9999, ASK Q58)

58. Well, what was your household's approximate income in the month of February 2021? **(TEXT BOX) (DK/REFUSE = 9999)**

(RESUME ASKING ALL)

59. Are you covered by any form of health insurance or health plan? This would include any private insurance through an employer or any plan you purchased yourself, as well as a government program like Medicare or Medi-Cal/Medicaid.
1. Yes, I am covered by a health plan
2. No, I am not covered by a health plan
3. I'm not sure/Prefer not to respond

(IF Q59 = 1, ASK Q60)

60. Of the following, which best describes your MAIN source of health insurance coverage. **(READ LIST)**
1. My own employer
2. My spouse, partner, or other family member's employer
3. Purchased it myself
4. Medicare
5. Medi-Cal/Medicaid
6. Some other government program (Champus, TriCare, etc.)
7. Somewhere else
8. Prefer not to respond

(RESUME ASKING ALL)

61. And have you or anyone in your household ever utilized a tele-health appointment to see a doctor or get medical care or advice? **(IF YES)** Was that you, or someone in your household?
1. Yes, myself
2. Yes, someone in my household
3. Yes, both myself and someone in my household
4. No
5. (Don't know/Refused)

THANK YOU!

CHAPTER 3. COVID-19 Vaccine Sentiments and Emotions on Twitter: An observational study to characterize the U.S. public's perceptions

Introduction

COVID-19 Vaccines

In the early phase of COVID-19 in the United States (U.S.), when vaccines were not available, the primary public concern was to contain the virus in specific areas, especially in the COVID-19 hotspot states.[66] As we know, this has proven ineffective as COVID-19 spread in the country, with each state using its guidelines to address the pandemic.[67] The spread of vaccine hesitancy, moreover, has contributed to the spread and devastation of COVID-19 in the United States. [4] Vaccinating against COVID-19 is an important strategy as the current Center for Disease Control and Prevention (CDC) recommendations state that people should be vaccinated regardless if they already have the virus or not and to protect themselves and others.[68] There is also official information from the CDC regarding the side effects of the COVID-19 vaccines and minimal long-term impacts, yet there are still those who are skeptical about getting the vaccine. [68] COVID-19 vaccine hesitancy can be explored using the World Health Organization's model of lack of confidence or trust, high complacency or perception of vaccines as unnecessary, and low convenience or low geographical accessibility to vaccines.[5]

Sentiment and Emotions & Twitter

At the center of this hesitancy to be vaccinated were people's perceptions of COVID-19 vaccines. [15] Published studies have successfully used sentiment and emotion analysis to observe patterns in public perceptions of COVID-19 vaccines. Some important findings that are known regarding public perceptions and COVID-19 vaccines are that emotions and sentiments in online content appear associated with the spread of information at scale, connected to compliance with preventative behaviors such as vaccinations, and play an important role in communication, specifically in the persuasion of positive health behaviors. [38–41]

Emotion and sentiment analysis is an existing approach to studying the opinions and feelings expressed in textual format.[69] The focus of this paper was on emotions and sentiments as they can partially reflect public perceptions. Although emotions and sentiments are only part of people's complex psychological states, they are important to explore as they may also reflect underlying beliefs in lack of confidence or strong negative opinions towards vaccination. [18] Health researchers can use this type of analysis to understand how attitudes change throughout a polarizing event, such as the official use and recommendation of COVID-19 vaccines to the general public. [67]

To better quantify people's perception of COVID-19 vaccines, previous studies have effectively used online platforms, such as Twitter, and applied sentiment and emotion scores to people's tweets. [70–74] These studies have recognized that geographical data is an important component in understanding the public's perception of COVID-19 vaccines. [70,72,73] Mislove et al. (2021) explored the Twitter data distribution geographically and found that areas densely populated, such as metropolitan areas, were overly represented. Thus, geographical data was an important consideration when working with tweets.

Purpose

Previous studies that examined COVID-19 vaccine public perceptions, in the context of emotions and sentiments, on Twitter have not looked at emotion and sentiment scores with geographic information from early in the pandemic in 2020 through late 2021. [71,74] Those that looked at public perceptions, in the context of emotions and sentiments, and used geographic information only used country or state-level data, and without the precision and distinction between heavily and lightly populated areas. [71,74] Those with the county- or local-level tweet data precision do not explore public emotions or sentiments in their study at all. [75]

Current literature recommends the importance of considering powerful structures, including geographic information, in fully understanding individual health preventative behaviors and decision-making. [5,20,30] There is currently a gap in the literature that does not examine public perceptions using emotion or sentiment scores on Twitter, recent full-time-series data from early 2020 through late 2021, and geographical information. We addressed this gap in the literature by looking at what the public perceptions are about COVID-19 vaccines in the United States from early 2020 through late 2021. Specifically, we looked at public perception using quantitative emotion and sentiment scores on Twitter. In addition, we tried to address the geographical differences of the tweets (i.e., metropolitan and nonmetropolitan areas) in our study.

Methods

Data Extraction

This study was a quantitative analysis of Twitter posts about COVID-19 vaccines in English in the U.S. from March 11, 2020, to September 12, 2021. The dataset did not contain any retweets, and we only collected tweets with geocodes. We queried tweets that include keywords that mention variations of COVID-19 (e.g., covid, corona, coronavirus, etc.) or mentions of companies that produce COVID-19 vaccines (e.g., Pfizer, Moderna, Johnson & Johnson, etc.). We combined these with mentions of vaccinations (e.g., vax OR vacc'd OR vaccination, etc.). We excluded any tweets that mention non-COVID-19 vaccines (e.g., DPT, influenza, TDAP, etc.). More details of the data extraction, cleaning, and processing results are shown in Figure 1.

Twitter Application Program Interface (API) v2 Academic Research [76] was used for collecting tweet data. Twitter API v2 provided variables of interest, including the tweet date and time, the full text of the tweet, and the geographical bounding box of where the tweet was posted (i.e., longitude and latitude). The U.S. Department of Agriculture (USDA) Rural-Urban Continuum Codes, [77] which contains the metropolitan or nonmetropolitan designations for U.S zip codes, was used for classifying the tweet locations. We processed the tweets through standard data cleaning practices for sentiment and emotion analysis. We used the tweet's city, county, and zip code to add the USDA rural-urban continuum codes in each tweet, specifically the metropolitan and nonmetropolitan designation of each geotagged tweet.

Overall, the initial Twitter API v2 data extraction generated tweets using the inclusion and exclusion criteria in Table 2. The data cleaning and reverse geocoding process was used to reduce the extracted data to tweets that contain location data. Finally, merging the Twitter and USDA rural-urban continuum codes yielded tweets that contain geographical designations.

Data Processing

For the unit of analysis, we aggregated the tweets per day. For sentiment analysis of tweets, we used the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon, a well-known rule-based lexicon for calculating sentiment scores. [78] This lexicon has been generated and validated through crowd-sourced validation methods via Amazon Mechanical Turk and performs well with online social conversations such as on Twitter. [78] According to the VADER lexicon, a sentiment can be described as people's opinions, evaluations, and attitudes, and their lexical features, such as words, can be labeled according to their orientation of positive, neutral, or negative. [78] The VADER lexicon produces a compound score, which was calculated by summing the sentiment score of each word and yields an output score for each tweet between -1 (extreme negative sentiment score) to 1 (extreme positive sentiment score). [78] We then created a category, sentiment, which outputs negative (tweet compound score <= -0.05), neutral (tweet compound score between >-0.05 and <0.05), and positive (tweet compound score >= 0.05).

For emotion analysis of tweets, we used the National Research Council Canada Lexicon (NRCLex) [79] a dictionary containing 14,128 rows of unique English words that are similarly validated via crowdsourcing methods. This lexicon scores each column with matching negative sentiment, positive sentiment, and eight emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). Each word contains a corresponding score of 0 to 4 for each sentiment. The approach was to match each word of the tweets to one or more emotions and their scores from the lexicon. We then produce a compound score by summing the emotion score of each word to yield an output score for each tweet between 0 to 4 for each of the emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust).

Data Analysis

We conducted our analyses in two phases. For phase 1, we analyzed the whole time-series data (i.e., March 11, 2020, to September 12, 2021) using events in the CDC COVID-19 museum timeline (see Appendix A – Methods and Table 4 for more details). [67] For phase 2, we analyzed the time-series data stratified into two separate periods (i.e., March 11, 2020, to December 11, 2020, and December 12, 2020, to September 12, 2021) and two geographical designations (metropolitan and nonmetropolitan). The first time-period, dating from March 11, 2020, to December 11, 2020, represents when the public knew that COVID-19 was a pandemic and was anticipating the development of vaccines. The second time-period, dating from December 12, 2020, to September 12, 2021, represents when the first COVID-19 vaccines were authorized and available to the public. We used the Welch's t-test, or unequal variances t-test, specifically in phase 2, to determine if there are statistically significant differences between these two time periods and between the two geographical designations (see Appendix A – Methods for more details).

Results

Twitter Dataset

The initial Twitter API v2 data extraction generated 215,826 tweets, then reduced to 181,686 tweets that contain location data, then further reduced to yield a total of 127,648 tweets that contain geographical designations. There were 27,513 (21.6%) tweets from March 11, 2020, to December 11, 2020, and 100,135 (78.4%) from December 12, 2020, to September 12, 2021. There were 97,815 (76.6%) tweets in metropolitan areas and 29,833 (23.4%) tweets in nonmetropolitan areas. For sentiments, there were 37,118 (29.1%) tweets with negative sentiment, 33,861 (26.5%) tweets with neutral sentiment, and 56,669 (44.4%) tweets with positive sentiments. For emotions, all eight categories of anger, anticipation, disgust, fear, joy, sadness, surprise, and trust were scored in all 127,648 tweets.

Sentiment Analysis

Certain COVID-19-related events and news appear associated with the public's sentiment score mean and sum. For example, COVID-19-related events most strongly associated with tweet sentiment scores include November 9, 2020, when Pfizer announced their first Covid-19 vaccine as 90% effective in their latest trials (mean=0.24, max=272); December 11, 2020, when the FDA issued an Emergency Use Authorization (EUA) the first COVID-19 vaccine (mean=0.07, max=40.8); December 15, 2020, when the FDA announced that the Moderna COVID-19 vaccine was 94% effective and safe (mean=0.19, max=179); April 13, 2021, when CDC put out a recommendation pausing the Johnson and Johnson COVID-19 vaccine due to blood clot issues (mean=0.04, max=40); May 7, 2021, when Pfizer requested FDA to have full approval of their COVID-19 (mean=0.11, max=110); and August 23, 2021, when the FDA approved Pfizer COVID-19 vaccine for official use (mean=0.23, max=239).

Phase 1 – Time-series Analysis

Figure 2 shows how the mean sentiment about COVID-19 vaccines changes per day over 18 months for each geographical designation. For example, from the announcement of COVID-19 as a pandemic on March 11, 2020, to December 11, 2020, the means of positive and negative tweet sentiments were sporadic, with a broad sentiment means range for both the metropolitan (mean=0.76 to -0.28) and nonmetropolitan (mean=0.91 to -0.61) areas. However, after December 11, 2020, when the FDA issued an Emergency Use Authorization (EUA) for the first COVID-19 vaccine, there was a pattern of individual tweets having less mixed sentiments with more narrow sentiment mean range for both the metropolitan (mean=0.57 to -0.39) and nonmetropolitan (mean=0.61 to -0.23) areas.

Figure 3 shows the sentiment scores sum or how the total sentiments about COVID-19 vaccines add up per day over 18 months for each geographical designation. For example, from March 11, 2020, to November 8, 2020, there were no changes or fluctuations in the sum of tweet sentiments for both metropolitan (max=61) and nonmetropolitan (max=12). However, on November 9, 2020, when Pfizer announced their first COVID-19 vaccine as 90% effective in their latest trials, there was a drastic rise in the sum of tweet sentiments for the metropolitan (max=292) and nonmetropolitan (max=80) areas. We found similar spikes in the sum of tweet sentiments on December 15, 2020, for the metropolitan (max=214) and nonmetropolitan (max=45) areas when the FDA announced that the Moderna COVID-19 vaccine was 94% effective and safe. Similarly, we found spikes in the sum of tweet sentiments on May 7, 2021, for the metropolitan (max=123) and nonmetropolitan (max=23) areas, when Pfizer requested the FDA to have full approval of their COVID-19, and on August 23, 2021, for the metropolitan (max=272) and nonmetropolitan (max=75) areas, when the FDA approved Pfizer COVID-19 vaccine for official use.

Phase 2 – Time-series and Geographical Designation Comparison

We explored the sentiment score mean of the first time-period (March 11, 2020-Dec 11, 2020) versus the second time-period (December 12, 2020-Sep 12, 2021). Overall, there was a significant difference in the mean sentiment scores between these two time periods with a mean difference of -0.0124, tweet sentiments were negative on average after December 11, 2020 ($p<0.0001$) (see Table 5).

We also explored the sentiment score mean of each geographical distribution (metropolitan and nonmetropolitan). Overall, we found no overall significant difference in the mean sentiment scores between geographical distributions (metropolitan and nonmetropolitan) with a mean difference of -0.0046 ($p=0.1363$) (see Table 6).

<u>Emotion Analysis</u>

Phase 1 – Time-series Analysis

Figure 4 shows how the mean emotion scores of COVID-19 vaccines for each of the eight emotions (i.e., anger, anticipation, disgust, fear, joy, sadness, surprise, trust) changes per day over 18 months for each geographical designation. For example, from the announcement of COVID-19 as a pandemic on March 11, 2020, to November 8, 2020, the mean tweet emotions scores across the eight emotions were sporadic and had broad emotion score mean ranging from trust (mean=2) to disgust (mean=0.03) for metropolitan areas, and anger (mean=3.52) to disgust (0.00) for nonmetropolitan areas.

After November 9, 2020, when Pfizer announced their first Covid-19 vaccine as 90% effective in their latest trials, the mean emotion scores of trust (mean=0.85) and anticipation (mean=0.80) had a wider range than other emotions of surprise (mean=0.52), anger (mean=0.46), joy (mean=0.42), sadness (mean=0.39), fear (mean=0.30), and disgust (mean=0.22) for metropolitan areas. Similarly, the emotions of trust (mean=0.81) and anticipation (mean=0.77) had a wider range than other emotions of surprise (mean=0.43), anger (mean=0.36), joy (mean=0.34), sadness (mean=0.36), fear (mean=0.27), and disgust (mean=0.18) for nonmetropolitan areas.

After around May 7, 2021, when Pfizer requested FDA provide full approval of their COVID-19 vaccine, all the eight emotions started to merge back into the same emotion scores range, with anticipation (mean=0.63) taking over the trust (mean=0.32), then joy (mean=0.22), fear (mean=0.10), sadness (mean=0.09), anger (mean=0.09), surprise (mean=0.10) and disgust (mean=0.07) for metropolitan areas. Similarly, the same emotional scores range with anticipation (mean=0.55) taking over the trust (mean=0.28), then joy (mean=0.09), fear (mean=0.16), sadness (mean=0.14), anger (mean=0.12), surprise (mean=0.07) and disgust (mean=0.10) for nonmetropolitan areas.

Figure 5 shows the emotion scores sum or how the total emotions about COVID-19 vaccines for each of the eight emotions add up per day over 18 months for each geographical designation. For example, from March 11, 2020, to November 8, 2020, there were no changes or fluctuations in the sum of tweet emotions, with the three highest emotions in this period being trust (max=148), anticipation (max=114), and joy (max=82) for metropolitan areas. In comparison, the three highest emotions are anger (max=56), sadness (max=45), and anticipation (max=29) for nonmetropolitan areas.

Phase 2 – Time-series and Geographical Designation Comparison

We explored the eight emotions' score mean for the first time-period (March 11, 2020-Dec 11, 2020) versus the second time-period (December 12, 2020-Sep 12, 2021). Overall, the mean emotion scores had a significant difference between the first time-period (Mar 11, 2020-Dec 11, 2020) versus the second time-period (Dec 12, 2020-Sep 12, 2021) only for emotions of anger (mean difference=-0.0305, $p<0.0001$), anticipation (mean difference=-0.1094, $p<0.0001$), disgust (mean difference=-0.0161, $p<0.0001$), joy (mean difference=-0.0203, $p<0.0001$), surprise (mean difference=-0.1402, $p<0.0001$), and trust (mean difference=-0.0874, $p<0.0001$) (see Table 5).

We also explored the emotion score mean of each geographical distribution (metropolitan vs. nonmetropolitan) for each of the eight emotions. Overall, there was no significant difference in the mean emotion scores between geographical distributions for all the eight emotions ($p>0.05$) (see Table 6).

Discussion

Principal Findings

We aimed to study the U.S. public perception of COVID-19 vaccines perception using quantitative sentiment and emotions scores on Twitter we extracted from March 11, 2020, to September 12, 2021. We focused on looking at the mean and sum of the public's tweet sentiment scores and various emotions. In phase 1, we explored the data from March 11, 2020, to September 12, 2021. In phase 2, we compared two separate periods (i.e., before and after December 11, 2020) and geographical designations (i.e., metropolitan and nonmetropolitan areas).

In phase 1, specific COVID-19-related announcements from the public media or private institutions appear associated with the public's perception of COVID-19 vaccines. We observed changes in the public's sentiment and emotion patterns on days that announced the effectiveness, safety, and official use of COVID-19 vaccines. For example, there were extreme fluctuations in sentiment and emotion scores from the announcement of COVID-19 as a pandemic on March 11, 2020, through December 11, 2020. The public perception of COVID-19 vaccines varied greatly in the early development stages of the vaccine. However, around December 11, 2020, when the FDA issued an Emergency Use Authorization (EUA) for Pfizer, the first COVID-19 vaccine, the pattern shifted. Positive and negative tweet sentiments started more clearly diverging from one another. The emotions of trust and anticipation began taking over other types of emotions. Of importance, is that the content of the tweets may differ before and after the release of the first COVID-19 vaccines. Generally, as an example, before December 11, 2020, some tweets were discussing the necessity, potential benefits, or detriments of vaccination, but after December 11, 2020, some tweets focused on discussing the safety, effectiveness, or side-effects of specific COVID-19 vaccines.

Similarly, for sentiment and emotions sum, from the announcement of COVID-19 as a pandemic on March 11, 2020, through November 8, 2020, there were no changes or fluctuations in the amount of tweet sentiment or emotions. It is also important to note that in the early development stages of the vaccine, there were few recorded opinions or tweets (21.6%) about the COVID-19 vaccines compared to after the release of the first COVID-19 vaccine (78.4%). This changed on November 9, 2020, when Pfizer announced their first COVID-19 vaccine as 90% effective in their latest trials. There was a dramatic rise in the sum of tweet sentiments. There was also a greater increase in the emotions of trust, anticipation, and surprise compared to other emotions. We observed a pattern of high public sentiment and emotion scores on days when COVID-19 breakthrough news was released. Specifically, on December 11, 2020, when the FDA issued an Emergency Use Authorization (EUA) of Pfizer, the first COVID-19 vaccine; on December 15, 2020, when the FDA announced that the Moderna COVID-19 vaccine was 94% effective and safe; on April 13, 2021, when CDC put out a recommendation pausing the Johnson and Johnson COVID-19 vaccine due to blood clot issues; on May 7, 2021, when Pfizer requested FDA to have full approval of their COVID-19; and on August 23, 2021, when the FDA approved Pfizer COVID-19 vaccine for official use.

In phase 2, we observed a significant difference in the mean sentiment scores between the first time-period (March 11, 2020-Dec 11, 2020) versus the second time-period (December 12, 2020-Sep 12, 2021). After the FDA issued a EUA of Pfizer, the first COVID-19 vaccine, the public sentiment was more negative on average during the second time-period compared to the first time-period. We also observed that the sum of tweets increased this same day, as more people took this news to Twitter to share their opinions. Additionally, we found no difference in the mean sentiment scores between geographical distributions. The population in metropolitan areas did not have a different opinion about COVID-19 vaccines than in nonmetropolitan areas.

We observed that the mean emotion scores in tweets for each of the eight emotions decreased after December 11, 2020, except for the emotions of fear and sadness, which we found not significant. Of significance are emotions of anger, anticipation, disgust, joy, surprise, and trust decreasing after December 11, 2020. Anticipation and surprise understandably declined as the news of the Pfizer COVID-19 vaccine being 90% effective had been released one month prior. The public expressed a combination of less anger, disgust, joy, and trust as the COVID-19 vaccine was released to the public, and its effects are being observed in the real world.

Our findings support other literature and related work where fear was prevalent in the early times when COVID-19 was announced as a pandemic. [72,73] However, this was quickly taken over by the emotion of trust later as the COVID-19 vaccines were released. Specifically, trust peaked on November 9, 2020, when Pfizer announced its vaccine was 90% effective [70,74] which was one of the events we found when the public trust was at its highest. Similarly, we found that public sentiment and emotion did not significantly differ at the geographical designation-level of metropolitan and nonmetropolitan areas.

Overall, in the context of the WHO vaccine hesitancy model, [18] lack of confidence was apparent during the first few months of the pandemic in 2020 as negative emotions of fear were high. This shifted during and after the news of the effectiveness of the first COVID-19 vaccine as positive emotions of trust, anticipation, and surprise compared to other emotions peaked during this time. High complacency, or the perception that COVID-19 vaccines are unnecessary, may be observed after the COVID-19 vaccine was publicly released, where the public began to express a combination of less anger, disgust, joy, and trust as fewer people are taking to Twitter to express their opinions. Additionally, this time frame is when mean emotion scores in tweets for each of the eight emotions began to decrease except for the emotions of fear and sadness. Although we did not see a significant difference in public sentiment and emotions between metropolitan and nonmetropolitan areas, this may not be indicative of high convenience or physical and geographical accessibility to COVID-19 vaccines.

Implications

This study relied on the geographical designation of the USDA Rural-Urban continuum codes to categorize the tweet location. Future work can use the designation of rural and urban, in addition to metropolitan and nonmetropolitan, to further explore the effects of a geographic designation on the public's perception of COVID-19 vaccines. Our findings reveal specific events such as the COVID-19 vaccine developments and breakthroughs that are associated with when people take the discussion and express their feelings online in each geographic designation. Identifying timeframes when the public when there was peak public interest in COVID-19 vaccines can prove helpful in combating vaccine hesitancy, misinformation, and disinformation. [70–72]

This study relied on the most relevant COVID-19 news on specific dates to find associations with the tweet sentiment and emotions scores. Future work can also leverage probabilistic clustering or topic modeling to identify emerging topics from tweets on specific dates. Topic modeling, complemented by sentiment and emotion analysis, has been a conventional approach used by previous works. [70–72,75] This approach has effectively organized tweets into different themes and patterns. As a result, topic modeling can provide a clearer view of the topics associated with tweets and, in turn, can offer tailored COVID-19 vaccine messaging to the public.

Limitations

The first limitation was the use of ecological data in this study. This study uses grouped Twitter data to estimate the characteristics of individuals. Specifically, we used groups' mean and sum emotion and sentiment scores as a proxy to explain individual perceptions about COVID-19 vaccines. The use of aggregate emotion and sentiment scores may not infer perception towards vaccination specifically, as Tweets may also contain perceptions about politics, policies, events, or other topics. Thus, caution must be taken to avoid ecological fallacy and make assumptions that the public's emotions and sentiments about COVID-19 vaccines apply to the individual level and only specifically to vaccines as their own entity. Furthermore, according to Twitter's developer and agreement policy, we cannot use the data to segment individuals based on sensitive personal information, including health beliefs, vaccination, or political affiliation. [80] Thus, to comply with Twitter's policy, the data had to be grouped instead of tracking individual users' emotions and sentiments.

The second limitation was using the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon [78] and the National Research Council Canada Lexicon (NRCLex) [79] for sentiment and emotion score calculation. Each dictionary was created and validated via open crowdsourcing or the authors themselves. Although the words are in the English language, there might be different meanings for the words in the context of this study. Thus, caution must still be applied in the final sentiment and emotion score interpretation. These lexicons were validated for a different purpose other than reading tweet text for sentiments on vaccines. The lexicons only look at unigrams or single words and do not consider any qualifiers before the word. Phrases such as "not bad" and "not amazing" may be misinterpreted as negative and positive. Negations and sarcasm are not considered in the scoring. The sum of sentiment scores can have a different influence on a text with a large amount of positive and negative words and have a mean score of zero.

The third limitation was the use of Twitter as a sample. As a result, the sample will only consist of individuals who use Twitter and not any other demographics. This, in turn, was not representative of the entire U.S. population, specifically not those who do not use this platform to express their opinions and concerns. Nevertheless, as described above, Twitter has a unique advantage for this study.

Conclusion

To the best of our understanding, this study was the first to explore public sentiments and emotions about COVID-19 vaccines at the geographical precision of metropolitan and nonmetropolitan levels. Our study contributes knowledge to the field of nursing, public health, and informatics in a few ways. What we now know was that first, public sentiment and emotions appeared associated with COVID-19 vaccine news and breakthroughs from March 2020 to September 2021. Second, the news and social media are potent outlets for health information and can potentially affect the public's perceptions about COVID-19 vaccines in both metropolitan and nonmetropolitan areas. Third, there was no significant difference in metropolitan and nonmetropolitan areas regarding the public sentiment and emotions about COVID-19 vaccines. COVID-19 vaccine news appears to penetrate the public no matter their geographical location.

What is still unknown is, first, the effects of rural and urban classification, in addition to metropolitan and nonmetropolitan, on the public's perception of COVID-19 vaccines. Second, we still do not know what specific topics and themes lie within this sample and their significance in each geographic designation. Third, misinformation and disinformation were not the focus of this study, so their effects on public sentiment and emotions in each geographic designation are also still unknown. Thus, complementing this study with approaches such as using the classification of rural and urban to explore other effects of geographic designation, leveraging probabilistic clustering like topic modeling to identify emerging themes, and identifying misinformation and disinformation to explore how it affects the public perception of COVID-19 vaccines can help further the knowledge we have in COVID-19 vaccine perceptions.

Figure 1. Data cleaning and processing results.

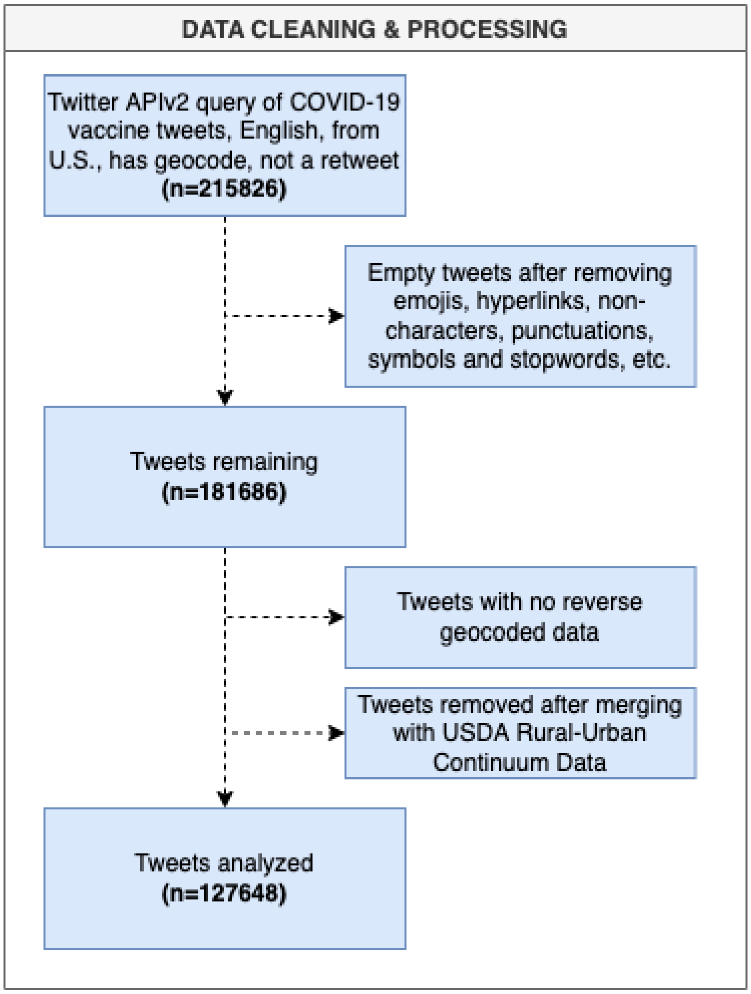
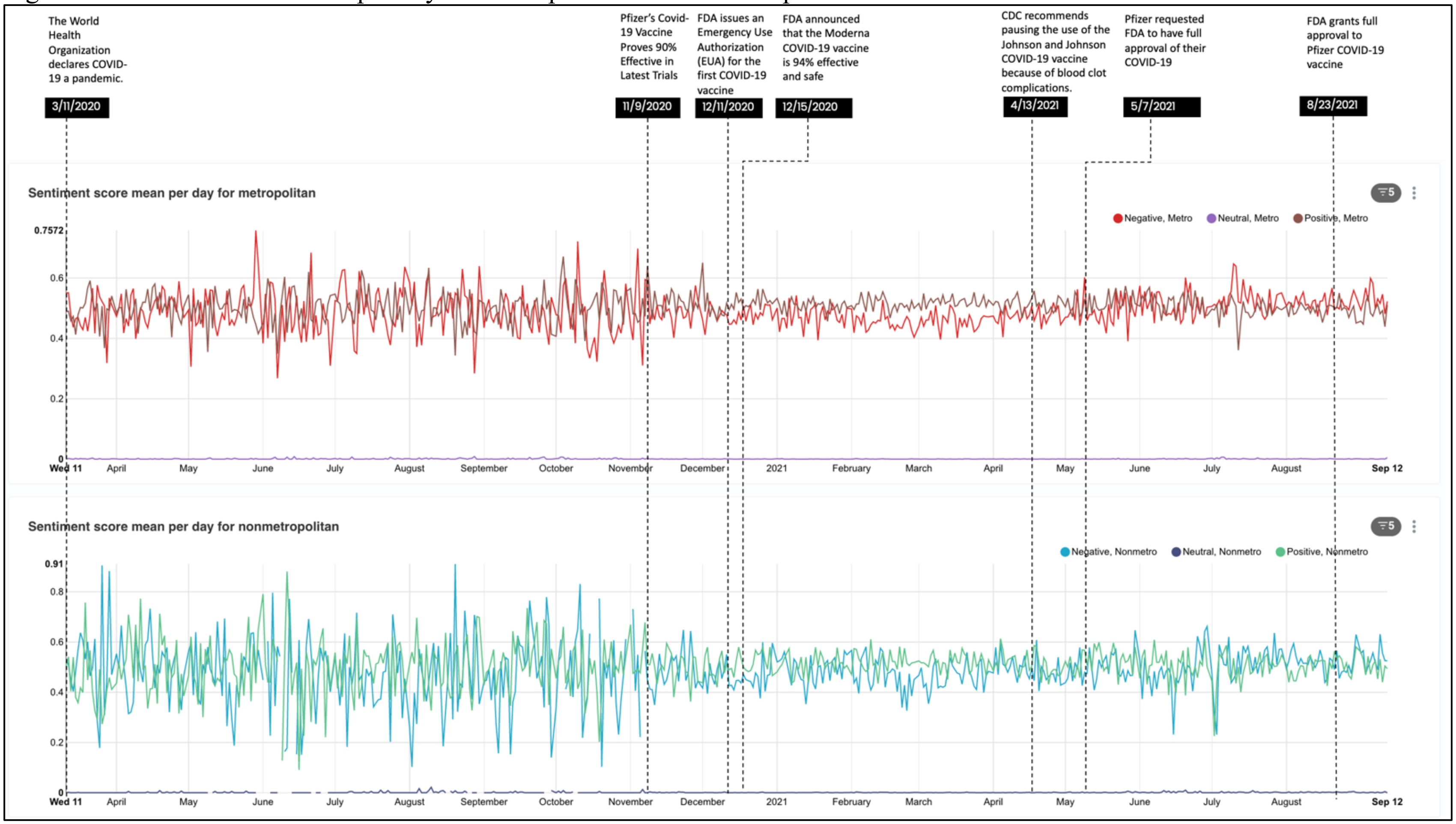
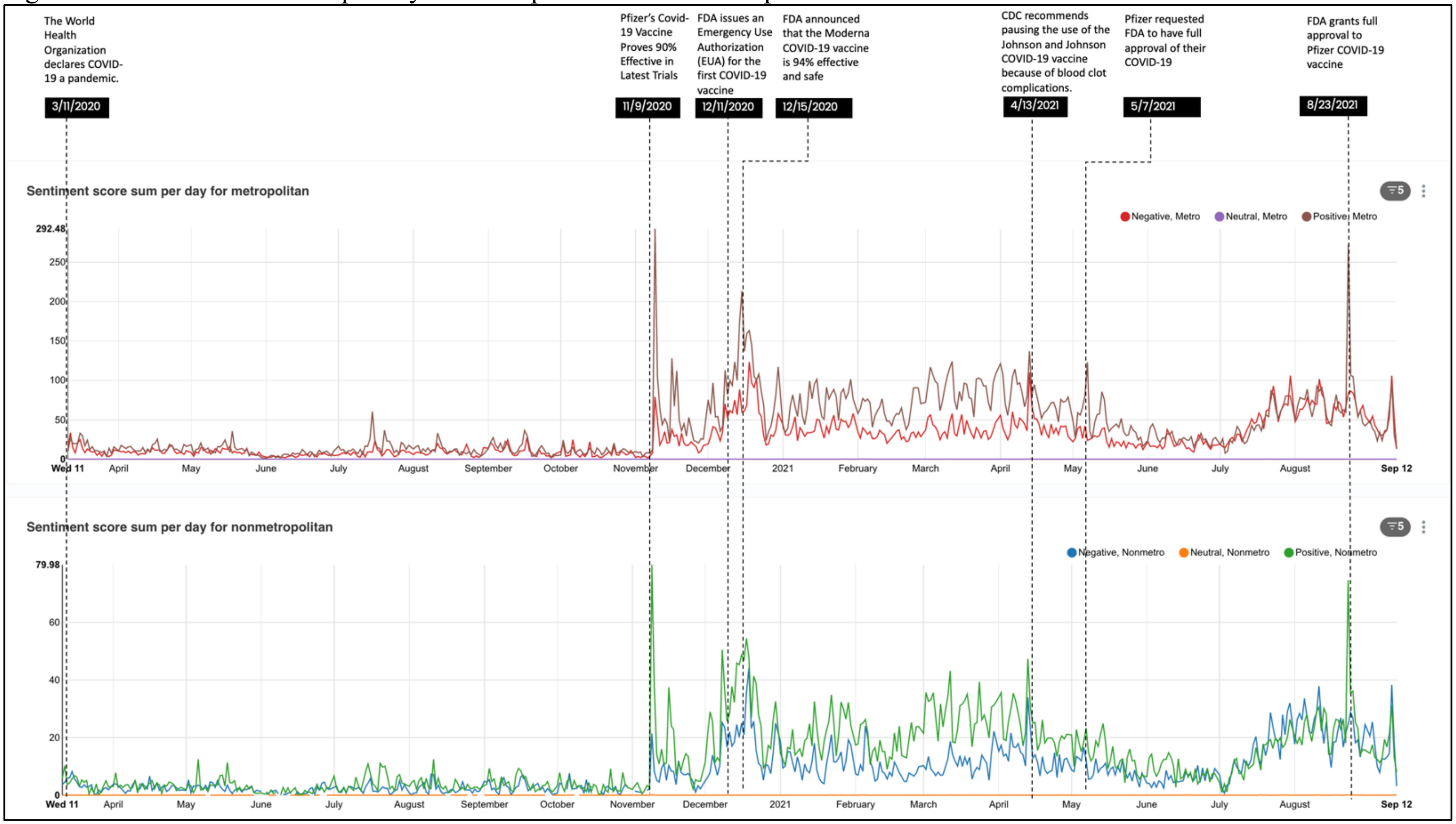


Figure 2. Sentiment score mean per day for metropolitan and nonmetropolitan.



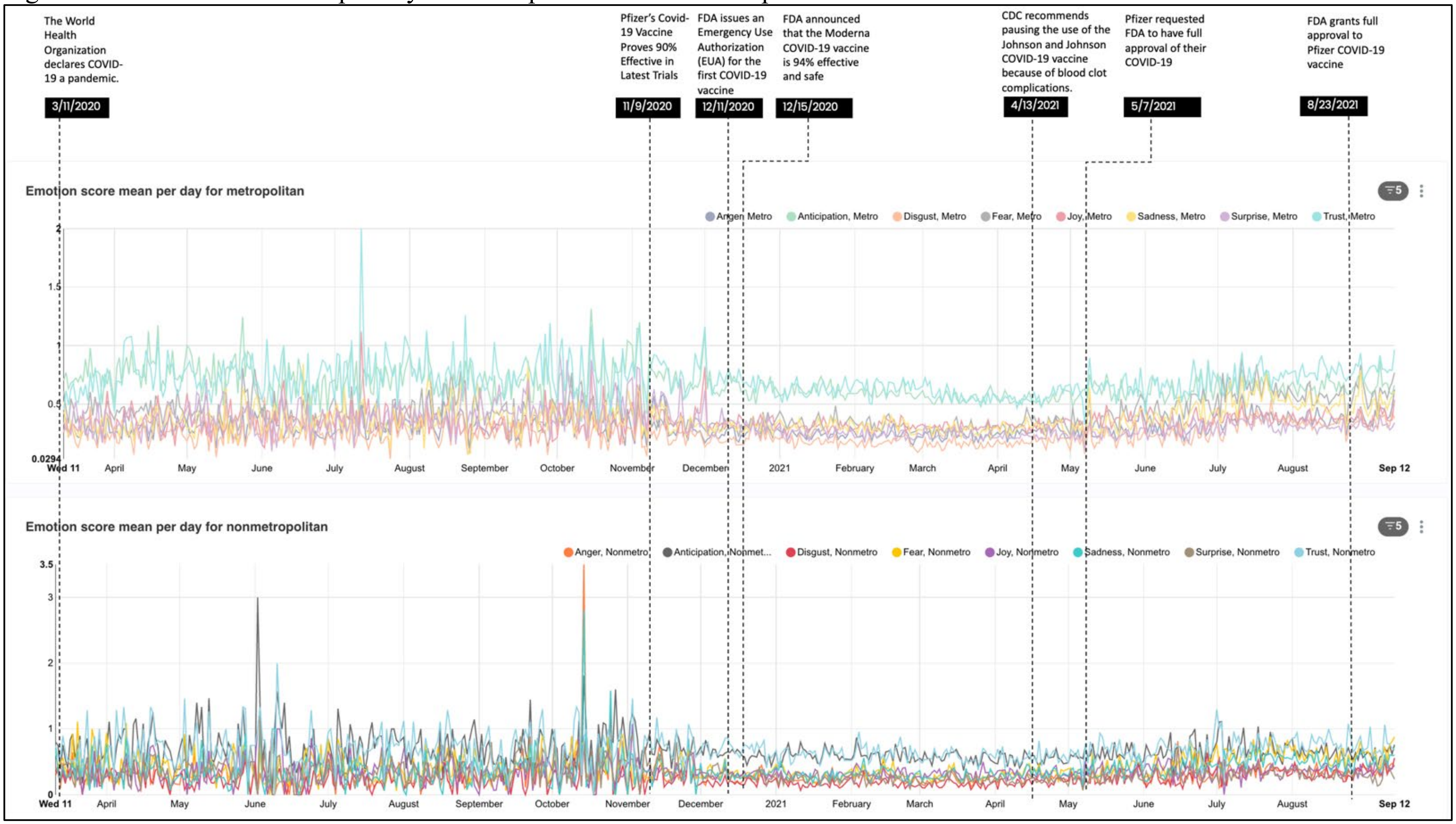
Note. Line charts of the mean sentiment scores (absolute value (-1 to 1)) of COVID-19 Vaccine Tweets, grouped by Sentiment Categories (Negative, Neutral, Positive) per day from Mar-11-2020 to Sep-12-2021 for each geographical designation (metropolitan and nonmetropolitan).

Figure 3. Sentiment score sum per day for metropolitan and nonmetropolitan.



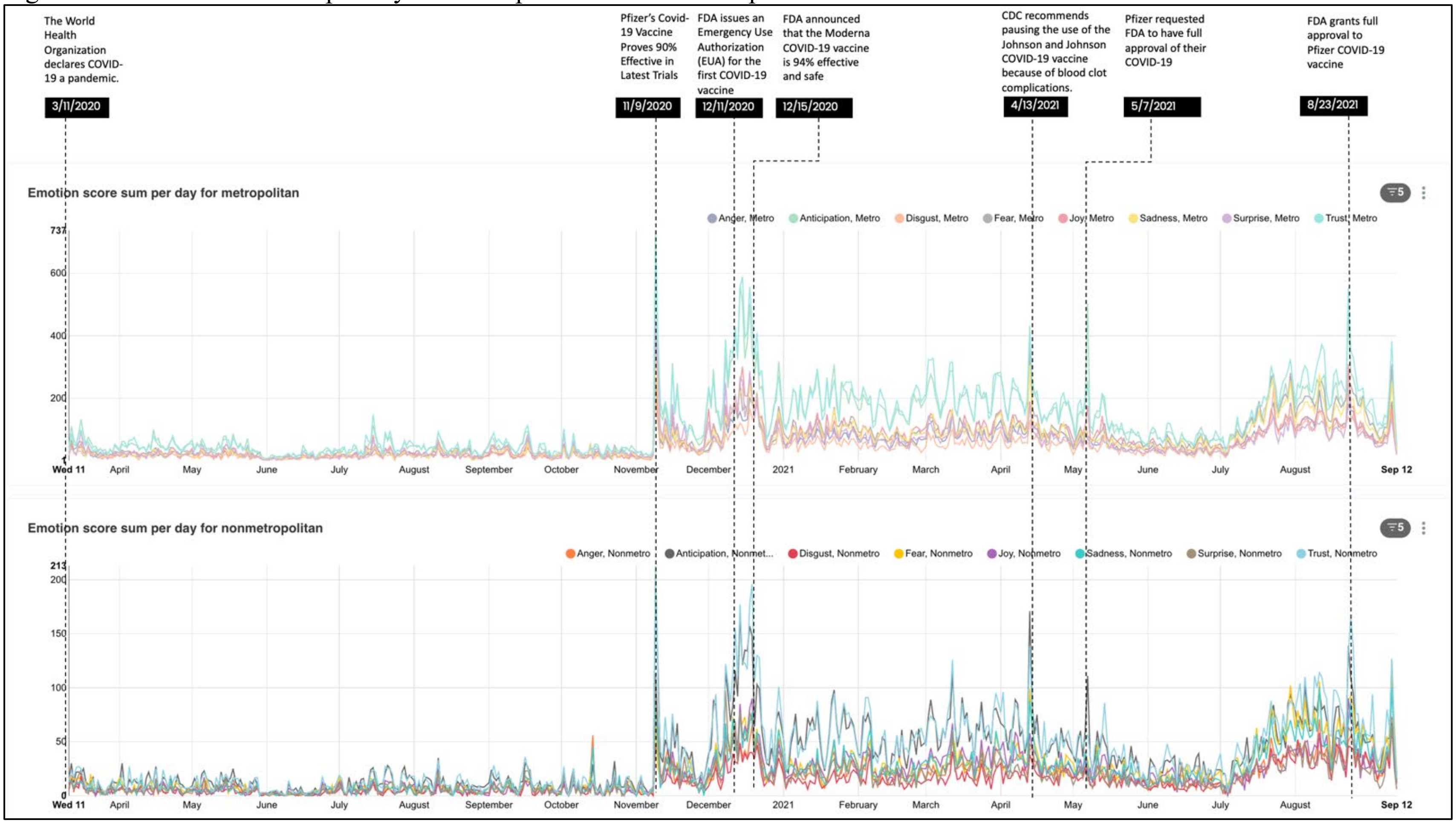
Note. Line chart of the sum of sentiment scores (absolute value (-1 to 1)) of COVID-19 Vaccine Tweets, grouped by Sentiment Categories (Negative, Neutral, Positive) per day from Mar-11-2020 to Sep-12-2021 for each geographical designation (metropolitan and nonmetropolitan).

Figure 4. Emotion score mean per day for metropolitan and nonmetropolitan.



Note. Line chart of the mean emotion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) per day from Mar-11-2020 to Sep-12-2021 for each geographical designation (metropolitan and nonmetropolitan).

Figure 5. Emotion score sum per day for metropolitan and nonmetropolitan.



Note. Line chart of the sum of motion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) per day from Mar-11-2020 to Sep-12-2021 for each geographical designation (metropolitan and nonmetropolitan).

Appendix A – Methods

Figure 6. Overview of methods and analysis.

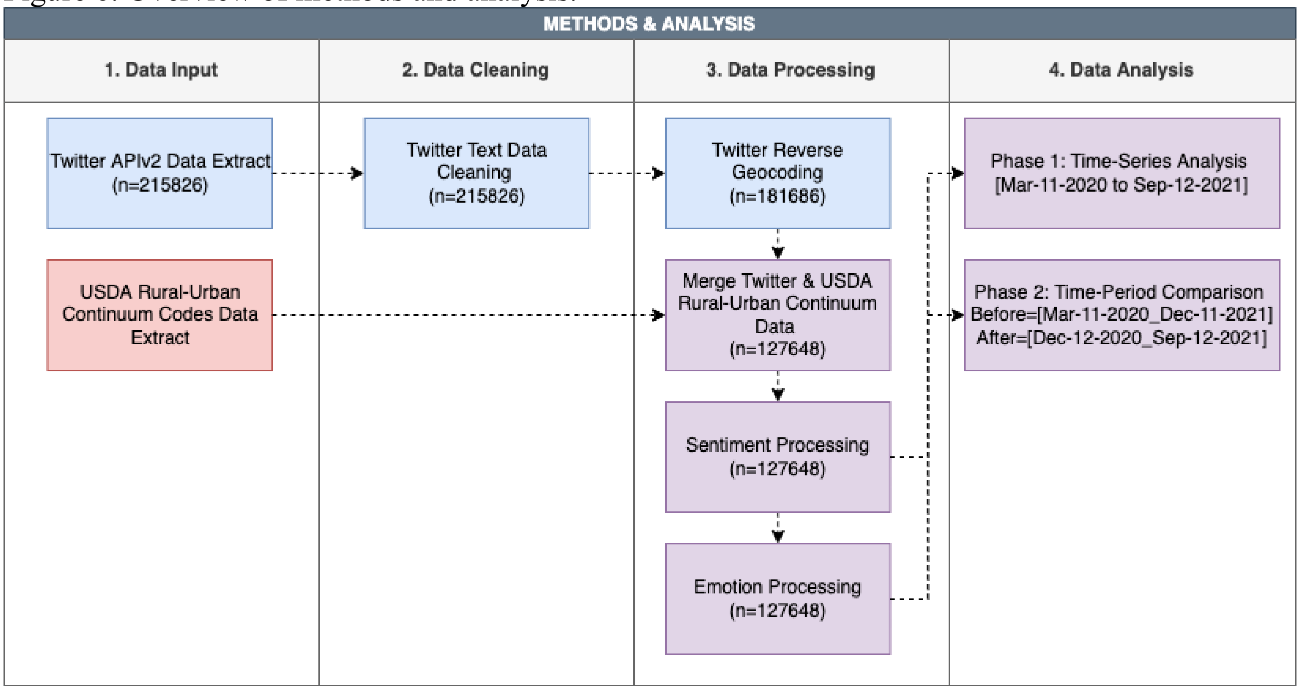


Table 1. Tweet Characteristics.

	count	**sum**	**mean**	**min**	**max**
By Tweets					
Total Tweets	127648	-	-	-	-
By Time Period					
3-11-2020 to 12-11-2020	27513	-	-	-	-
12-12-2020 to 9-12-2021	100135	-	-	-	-
By Geographical Designation					
Metro	97815	-	-	-	-
Nonmetro	29833	-	-	-	-
By Emotions (Emotion Scores)					
Anger	127648	36368	0.28	0	9
Anticipation	127648	80060	0.63	0	9
Disgust	127648	28669	0.22	0	9
Fear	127648	50338	0.39	0	10
Joy	127648	43187	0.34	0	8
Sadness	127648	45398	0.36	0	9

Surprise	127648	37518	0.29	0	6
Trust	127648	85923	0.67	0	10
By Sentiments (Sentiment Scores)					
Negative	37118	-18183	-0.49	-0.99	-0.05
Neutral	33861	6	0	-0.05	0.05
Positive	56669	28901	0.51	0.05	0.99

Table 2. Inclusion and exclusion criteria.

Inclusions	**Exclusions**
Global Filters:	*Global Filters:*
has:geo (i.e., has geographic data in the tweet)	-is: retweet (i.e., Not a Retweet)
place_country: U.S. (i.e., tweet in the U.S.)	
lang:en (i.e., tweet is in English)	
start_date: 2019-12-01 (i.e. tweet after Dec. 1, 2019)	
end_date: 2021-12-31 (i.e. tweet before Dec. 31, 2021)	
Keyword Filters:	*Keyword Filters:*
pfizer* AND (vacc* OR vax*), biontech* AND (vacc* OR vax*), sinopharm* AND (vacc* OR vax*), sinovac* AND (vacc* OR vax*), moderna* AND (vacc* OR vax*), oxford* AND (vacc* OR vax*), astrazeneca* AND (vacc* OR vax*), covaxin* AND (vacc* OR vax*), sputnik* AND (vacc* OR vax*), johnson* AND (vacc* OR vax*), covid* AND (vacc* OR vax*), corona* AND (vacc* OR vax*), sars-cov-2* AND (vacc* OR vax*)	-acute flaccid, -Adenovirus, -afm, -AIDS, -Anthrax, -anti-flu, -BCG, -Bird flu, -cancer, -chickenpox, -cholera, -Diphtheria, -Diptheria, -DPT, -DTap, -dtp, -DTPa, -DTwP, -Encephalitis, -flu, -gardasil, -guardasil, -H1N1, -H7n9, -human papilloma, -HAV, -HBV, -hepatitis, -HEV, -Hexavalent, -Hib, -HIV, -HPV, -influenza, -IPV, -lyme, -Malaria, -measel, -measle, -measles mumps, -meesel, -meningitis, -meningococcal, -MMR, -mumps, -NOS, -OPV, -Papillomavirus, -Pentavalent, -pertussis, -Pneumococcal, -polio, -pox, -Rotavirus, -rubel, -shingle, -SIDS, -smallpox, -sudden infant, -swine flu, -TB, -Td, -TDaP, -tetanus, -tuberculosis, -typhoid, -varicella, -whooping cough, -yellow fever, -zika

Table 3. Data sources and tools used.

Tool Name	Type	Used For
Python v3.10.3	Tool	Data Input, Cleaning & Processing
Google Colab (3/10/22)	Tool	Data Input
Amazon Web Services S3 (3/18/22)	Storage	Data Storage
Ikigai (3/18/22)	Tool	Data Visualization
Data Source Name	**Type**	**Used For**
Twitter API v2	API	Twitter Data
Geopy v2.2.0	API	Reverse Geocoding Twitter Geo Data
U.S. Department of Agriculture (USDA) Rural-Urban Continuum Area Codes 2013	CSV	Metropolitan vs. Nonmetropolitan Classification
Natural Language Toolkit (NLTK) 3.7	Lexicon	Text Data Cleaning
National Research Council Canada Lexicon (NRCLex) v3.0.0	Lexicon	Emotions Data Analysis
Vader Lexicon (3/15/2021)	Lexicon	Sentiment Data Analysis

Table 4. COVID-19 events based on the CDC Museum Timeline. [67]

Date	COVID-19 Events
3/11/2020	COVID-19 officially declared a pandemic by World Health Organization (WHO)
3/13/2020	Due to COVID-19, the president of the US issues a national emergency
3/17/2020	Moderna initiates the first human trials of a COVID-19 vaccine at Kaiser Permanente in US
4/13/2020	The president of the US (Donald Trump) pulls out funding to the WHO
5/2/2020	WHO declares COVID-19 pandemic as a global health crisis
10/2/2020	The president of the US tested positive for COVID-19 infection
11/9/2020	Pfizer BioNTech COVID-19 vaccines announced to be 90% effective in new trials
12/11/2020	Food and Drug Administration (FDA) declares an Emergency Use Authorization (EUA) for the Pfizer BioNTech COVID-19 vaccines, the first publicly available COVID-19 vaccine
12/14/2020	The first person receives the Pfizer BioNTech COVID-19 vaccine in New York
12/15/2020	Moderna COVID-19 vaccines announced to be 94% safe and effective in new trials
12/30/2020	A new variant of COVID-19 virus reported in the US
2/27/2021	FDA declares an Emergency Use Authorization (EUA) for the Johnson & Johnson COVID-19 one-shot vaccines

3/8/2021	Center for Disease Control (CDC) issues a statement that people that are fully vaccinated can begin meeting indoors without a mask
4/2/2021	CDC issues a statement that people that are fully vaccinated can begin traveling domestically without proof of COVID-19 test
4/13/2021	CDC announces the blood clot complications from Johnson & Johnson COVID-19 one-shot vaccines and recommends pausing it
5/7/2021	Pfizer BioNTech COVID-19 submits request for FDA approval of its COVID-19 vaccines
6/1/2021	The new COVID-19 variant, Delta, begins a new wave of COVID-19 cases in the US
8/6/2021	CDC study reveals that fully vaccinated individuals were less than half as likely to be re-infected by COVID-19 than non-vaccinated individuals
8/23/2021	FDA fully approves Pfizer BioNTech COVID-19 vaccines

Data Cleaning of Tweet Text Data

We cleaned the full text of each tweet through multiple steps of data cleaning processes and stored it in a new column 'edited' for processing later. This data cleaning process started by splitting the tweet text into tokens or single words and then converted to a lower-case format. The following steps include dropping 'stopwords' or high-frequency words in the English language that are not important for text processing. [81] We also dropped Custom' stopwords', precisely terms used in the initial search query seen in Table 1 in the tweets, as they do not provide value in the data processing steps. We further cleaned tweets by removing any emojis, hyperlinks, non-characters, punctuations, and symbols in the text. All the remaining text we converted to their dictionary or root form using the stemmer and lemmatization package of the Natural Language Toolkit (NLTK) 3.7.

Data Processing of Location, Sentiment, and Emotion Data

The geotagged tweets extracted from Twitter API v2 have a polygon bounding box that contains a four-sided geographical area (e.g., west longitude, south latitude, east longitude, north latitude). [82] We generated the centroid of the latitudes and longitudes in each of the bounding boxes and used the centroid to represent the final location of the tweet. We then reverse geocoded the centroid of the geotagged tweets [83] or processed to retrieve more information about the place, specifically to get the tweet's city, county, and zip code using the GeoPy 2.2.0 package. The U.S. Department of Agriculture (USDA) Rural-Urban Continuum Codes, [77] which contains the geographic designations (metropolitan and nonmetropolitan). for U.S zip codes, was then used for classifying the tweet locations. The USDA updated this dataset every decade, and the latest version, 2013, was used. The Rural-Urban continuum codes classified counties by degree of urbanization and proximity to metropolitan areas. [84]

For sentiment analysis of tweets, we used the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon, a well-known rule-based lexicon for calculating sentiment scores. [78] It produces a compound score, which was calculated by summing the sentiment score of each word and yields an output score for each tweet between -1 (extreme negative sentiment score) to 1 (extreme positive sentiment score). [78] We then created a categorical column, sentiment, which outputs negative (tweet compound score <= -0.05), neutral (tweet compound score between >-0.05 and <0.05), and positive (tweet compound score >= 0.05). For emotion analysis of tweets, we used the National Research Council Canada Lexicon (NRCLex) [79] a crowdsourced dictionary containing 14,128 rows of unique English words and scores for each of the ten columns with matching negative sentiment, positive sentiment, and eight emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). Each word contains a score for each of the columns (i.e., no, weak, moderate, and strong) and a corresponding score of 0 to 4. The approach was to match each word of the tweets to one or more emotions and their scores from the lexicon. We then produce a compound score by summing the emotion score of each word to yield an output score for each tweet between 0 to 4 for each of the emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust).

Data Analyses

For phase 1, we started on March 11, 2020, when the World Health Organization declared COVID-19 a pandemic. Then, we used the mid-point on December 11, 2020, when the Food and Drug Administration announced an Emergency Use Authorization (EUA) for the first COVID-19 vaccine, Pfizer. Finally, we ended on September 12, 2021, the same amount of time between the first two events, specifically, -9 months after the EUA of the first COVID-19 vaccines.

For phase 2, we analyzed the data in two separate periods (i.e., March 11, 2020, to December 11, 2020, and December 12, 2020, to September 12, 2021). The first time-period, dating from March 11, 2020, to December 11, 2020, represents when the public knew that COVID-19 was a pandemic and was anticipating the development of vaccines. The second time-period, dating from December 12, 2020, to September 12, 2021, represents when the first COVID-19 vaccines were authorized and consumed by the public.

For the unit of analysis, we calculated the mean sentiment and emotion scores, exploring how the public sentiment and emotions about COVID-19 vaccines change over time for phase 1 and changes between the two time periods for phase 2. We also calculated the sum of sentiment and emotion scores, exploring how the sum of the public sentiment and emotions about COVID-19 vaccines changes over time for phase 1 and between the two time periods for phase 2. The primary metric explored was the sum and mean sentiment and emotion scores of these two comparisons aggregated into sentiment categories and the eight emotion categories. For visualization purposes, the absolute value of the mean sentiment scores less than 0 was used to have a visual of the line chart superimposed on top of one another.

We also explored these changes by aggregating the tweets by sentiment categories, emotion categories, and geographical designation in phases 1 and 2. We used the Welch's t-test specifically in phase 2 to determine if there are statistically significant differences between the two time periods (i.e., March 11, 2020, to December 11, 2020, and December 12, 202,0 to September 12, 2021) and between the two geographical designations of metropolitan and nonmetropolitan. We considered the unequal sample sizes of different periods. We used Welch's t-test, or unequal variances t-test, to test the significance of the difference in the means between different groups.

Appendix B – Results

Table 5. Significance testing of the difference in the means before and after December 11, 2020.

Test	Mean Difference	Standard Deviation Difference	Confidence Interval Difference	t-value	degrees of freedom	p-value
Sentiments_Before-After	-0.0125	-0.0015	[0.01 0.02]	3.9113	43692.1791	0.0001
Anger_Before-After	-0.0305	-0.0435	[0.02 0.04]	7.0846	41562.9196	0.0000
Anticipation_Before-After	-0.1094	-0.0763	[0.1 0.12]	17.3589	41152.9165	0.0000
Disgust_Before-After	-0.0161	-0.0248	[0.01 0.02]	4.2137	42339.7792	0.0000
Fear_Before-After	0.0014	0.0143	[-0.01 0.01]	-0.2831	44468.7849	0.7771
Joy_Before-After	-0.0203	-0.0110	[0.01 0.03]	4.5426	43234.0482	0.0000
Sadness_Before-After	0.0033	-0.0008	[-0.01 0.01]	-0.6959	43760.5957	0.4865
Surprise_Before-After	-0.1402	-0.1423	[0.13 0.15]	30.6449	37634.3728	0.0000
Trust_Before-After	-0.0874	-0.0517	[0.07 0.1]	13.3994	42027.8639	0.0000

Table 6. Significance testing of the difference in the means between metropolitan and nonmetropolitan.

Test	Mean Difference	Standard Deviation Difference	Confidence Interval Difference	t-value	degrees of freedom	p-value
Sentiments_Metro_NonMetro	-0.0046	0.0014	[-0. 0.01]	1.4896	49281.5969	0.1363
Anger_Metro_NonMetro	-0.0004	0.0024	[-0.01 0.01]	0.1052	49240.9551	0.9162
Anticipation_Metro_NonMetro	-0.0016	0.0016	[-0.01 0.01]	0.2771	49331.3502	0.7817
Disgust_Metro_NonMetro	0.0030	0.0105	[-0.01 0]	-0.8330	48642.7894	0.4048
Fear_Metro_NonMetro	0.0043	-0.0079	[-0.01 0.01]	-0.8841	49838.3958	0.3767
Joy_Metro_NonMetro	-0.0024	-0.0046	[-0.01 0.01]	0.5659	49694.6503	0.5714
Sadness_Metro_NonMetro	0.0066	0.0028	[-0.02 0]	-1.4350	49238.6745	0.1513
Surprise_Metro_NonMetro	-0.0073	-0.0134	[-0. 0.01]	1.8857	50347.6968	0.0593
Trust_Metro_NonMetro	-0.0033	-0.0014	[-0.01 0.02]	0.5438	49462.4309	0.5866

Figure 7. Sentiment score mean before and after December 11, 2020.

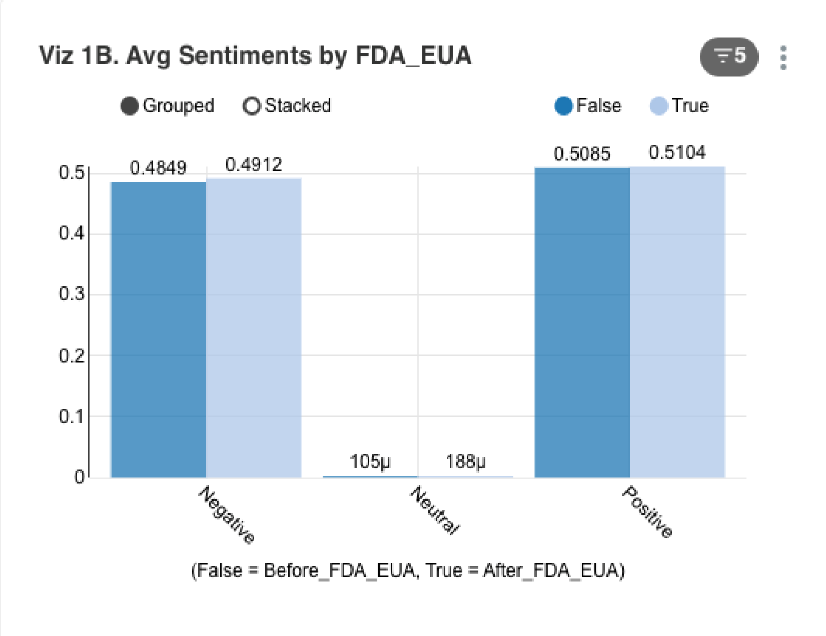


Table 1B. Avg Sentiments by FDA_EUA

	metric	(AVG(sentiment_intensity))	
	FDA_VaxEUA	False	True
sentiment			
Negative		-0.48494	-0.491203
Neutral		0.000104958	0.000187602
Positive		0.508532	0.510429

Note. Bar chart and Table of the mean sentiment scores (-1 to 1) of COVID-19 Vaccine Tweets, grouped by Sentiment Categories (Negative, Neutral, Positive) comparing the first time-period (March 11, 2020-Dec 11, 2020) vs. the second time-period (December 12, 2020-Sep 12, 2021).

Figure 8. Sentiment scores sum before and after December 11, 2020.

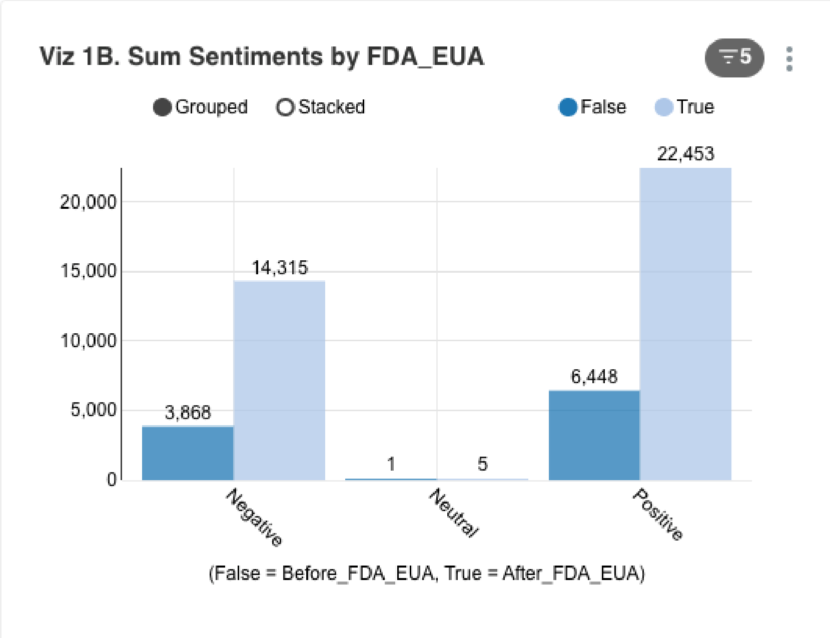
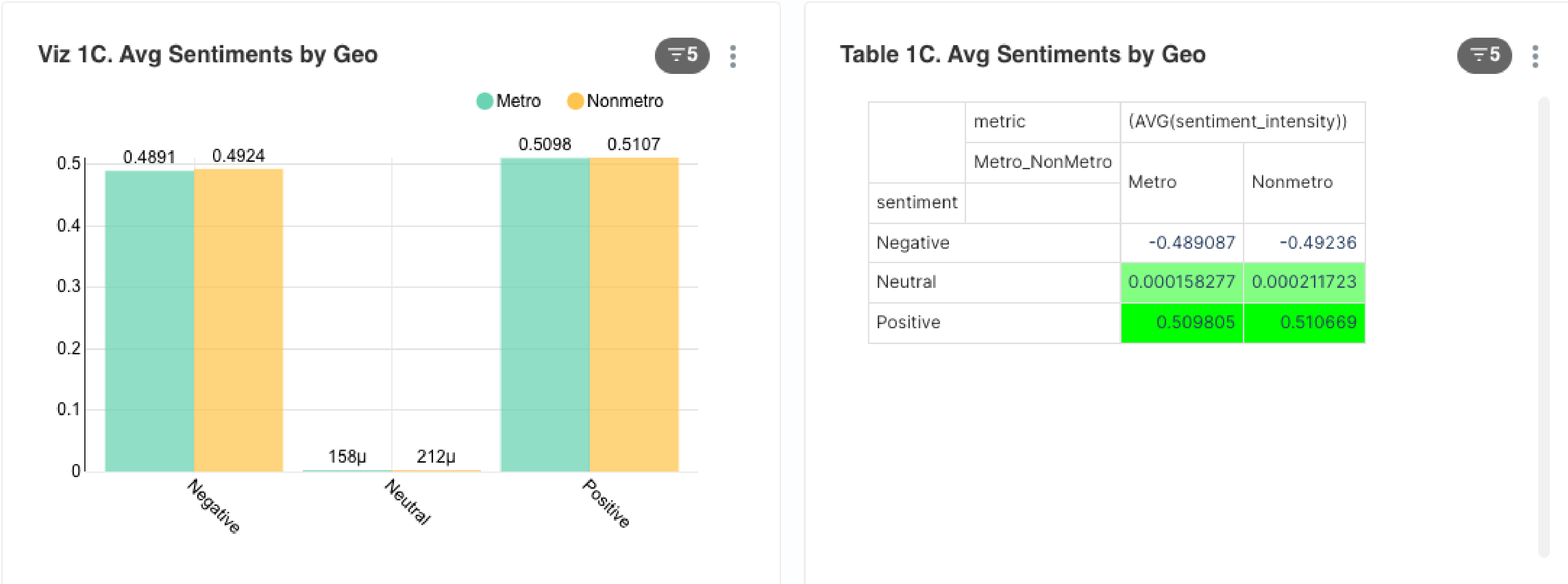


Table 1B. Sum Sentiments by FDA_EUA

	metric	(SUM(sentiment_intensity))	
	FDA_VaxEUA	False	True
sentiment			
Negative		-3867.88	-14314.6
Neutral		0.7197	5.066
Positive		6448.19	22453.3

Note. Bar chart and Table of the sum of sentiment scores (-1 to 1) of COVID-19 Vaccine Tweets, grouped by Sentiment Categories (Negative, Neutral, Positive) comparing the first time-period (March 11, 2020-Dec 11, 2020) vs. the second time-period (December 12, 2020-Sep 12, 2021).

Figure 9. Sentiment score mean for metropolitan and nonmetropolitan.



	metric	(AVG(sentiment_intensity))	
	Metro_NonMetro	Metro	Nonmetro
sentiment			
Negative		-0.489087	-0.49236
Neutral		0.000158277	0.000211723
Positive		0.509805	0.510669

Note. Bar chart and Table of the mean sentiment scores (-1 to 1) of COVID-19 Vaccine Tweets, comparing Geo Categories (Metro, Nonmetropolitan).

Figure 10. Sentiment scores sum for metropolitan and nonmetropolitan.

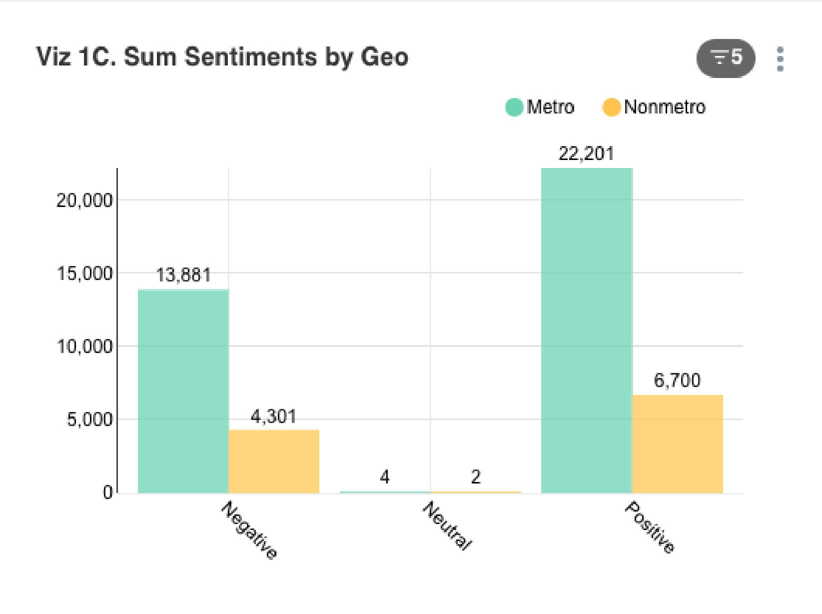


Table 1C. Sum Sentiments by Geo

	metric	(SUM(sentiment_intensity))	
	Metro_NonMetro	Metro	Nonmetro
sentiment			
Negative		-13881.3	-4301.25
Neutral		4.097	1.6887
Positive		22201	6700.48

Note. Bar chart and Table of the sum of sentiment scores (-1 to 1) of COVID-19 Vaccine Tweets, comparing Geo Categories (Metro, Nonmetropolitan).

Figure 11. Emotion score mean before and after December 11, 2020.

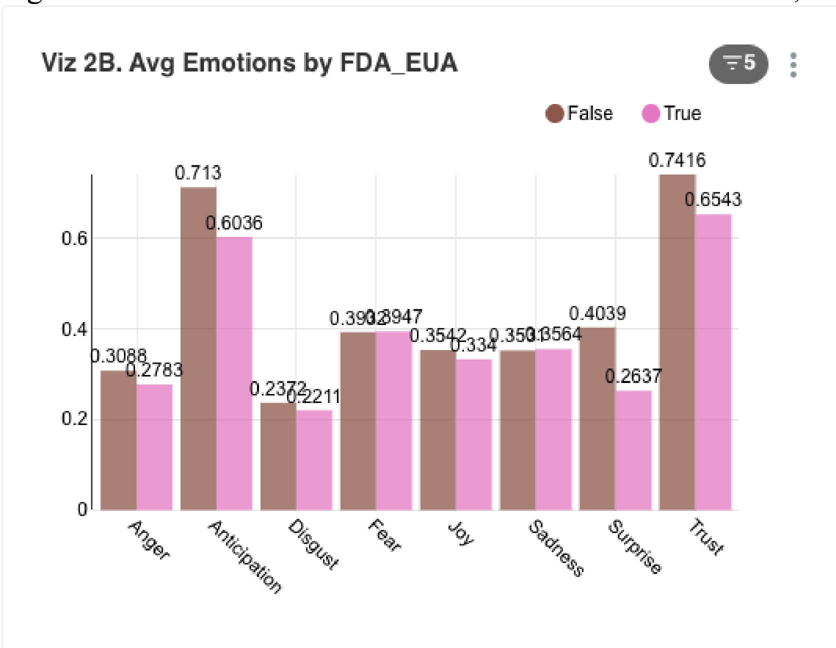


Table 2B. Avg Emotions by FDA_EUA

	metric	AVG(Emotion_Score)	
	FDA_VaxEUA	False	True
Emotions			
Anger		0.3088	0.2783
Anticipation		0.713	0.6036
Disgust		0.2372	0.2211
Fear		0.3932	0.3947
Joy		0.3542	0.334
Sadness		0.3531	0.3564
Surprise		0.4039	0.2637
Trust		0.7416	0.6543

Note. Bar chart and Table of the mean emotion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) comparing the first time-period (March 11, 2020-Dec 11, 2020) vs. the second time-period (December 12, 2020-Sep 12, 2021).

Figure 12. Emotion scores sum before and after December 11, 2020.

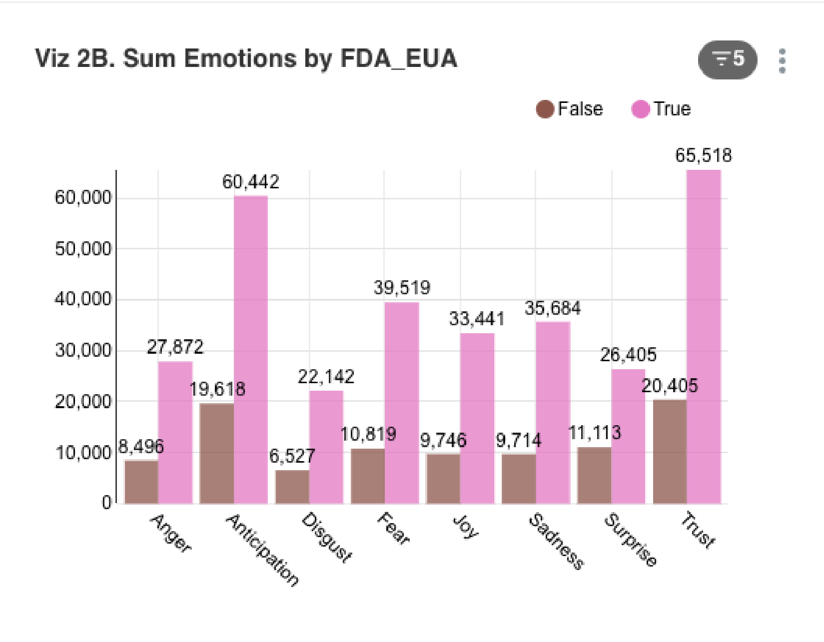


Table 2B. Sum Emotions by FDA_EUA

	metric	SUM(Emotion_Score)	
	FDA_VaxEUA	False	True
Emotions			
Anger		8,496	27,872
Anticipation		19,618	60,442
Disgust		6,527	22,142
Fear		10,819	39,519
Joy		9,746	33,441
Sadness		9,714	35,684
Surprise		11,113	26,405
Trust		20,405	65,518

Note. Bar chart and Table of the sum of motion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) comparing the first time-period (March 11, 2020-Dec 11, 2020) vs. the second time-period (December 12, 2020-Sep 12, 2021).

Figure 13. Emotion score mean for metropolitan and nonmetropolitan.

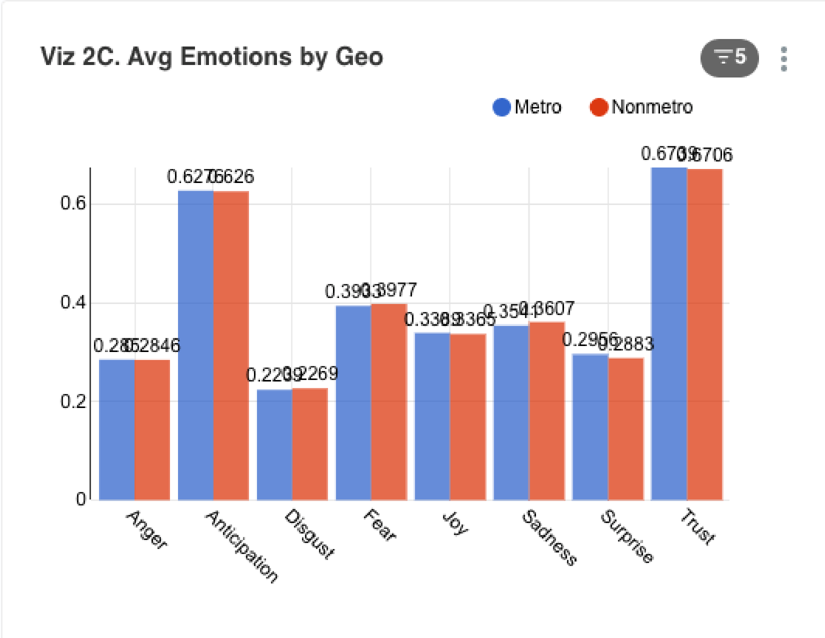


Table 2C. Avg Emotions by Geo

	metric	AVG(Emotion_Score)	
	Metro_NonMetro	Metro	Nonmetro
Emotions			
Anger		0.285	0.2846
Anticipation		0.6276	0.626
Disgust		0.2239	0.2269
Fear		0.3933	0.3977
Joy		0.3389	0.3365
Sadness		0.3541	0.3607
Surprise		0.2956	0.2883
Trust		0.6739	0.6706

Note. Bar chart and Table of the mean emotion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) comparing Metropolitan vs. Nonmetropolitan areas.

Figure 14. Emotion scores sum for metropolitan and nonmetropolitan.

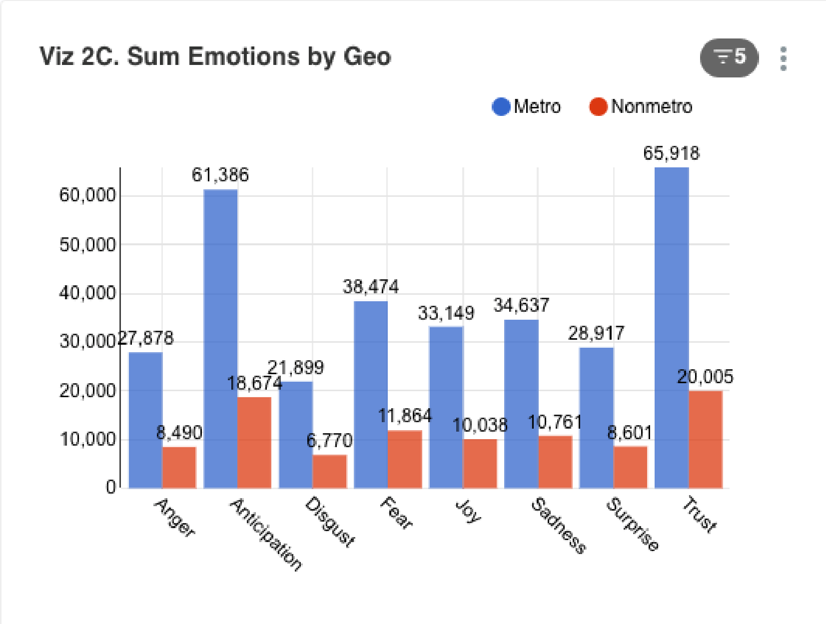


Table 2C. Sum Emotions by Geo

	metric	SUM(Emotion_Score)	
	Metro_NonMetro	Metro	Nonmetro
Emotions			
Anger		27,878	8,490
Anticipation		61,386	18,674
Disgust		21,899	6,770
Fear		38,474	11,864
Joy		33,149	10,038
Sadness		34,637	10,761
Surprise		28,917	8,601
Trust		65,918	20,005

Note. Bar chart and Table of the sum of motion scores (0 to 4) of COVID-19 Vaccine Tweets, grouped by Emotion Categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust) comparing Metropolitan vs. Nonmetropolitan areas.

CHAPTER 4. Implications and Conclusion

Implications and Future Work

This dissertation provides a better understanding of how communities in the United States perceive COVID-19 vaccines and the factors that affect these perceptions. Our findings include implications to combat COVID-19 vaccine hesitancy amongst rural and non-metropolitan communities, including leveraging surveys and social media to understand and share evidence-based health information.

This work contributes to the body of knowledge in public health by guiding healthcare professionals and government officials to take the appropriate actions to contain and reduce cases and mortality of the COVID-19 pandemic. Lack of accurate information about COVID-19 vaccinations combined with negative emotions may be connected to how people act on myths. It was vital to inform action against the spread of misinformation and disinformation and hesitancy about COVID-19 vaccinations. [33] Perceptions of the messages on vaccination and COVID-19 appear associated with how the public's health behaviors are shaped. [33]

We know that first, we can study perceptions about COVID-19 at the local levels by conducting interviews or analyzing sizeable social media data and generating novel insights. Second, we know that COVID-19 confidence was mainly based on the source and content of the news. Our findings, for example, show that people take to social media when major public and private institutions announce COVID-19 vaccine developments. Third, geographic location has not been a factor in the accessibility and perceptions of COVID-19 vaccines. In this study, the geographical location of residents did not appear to be associated with people receiving, sharing, and perceiving information about COVID-19 vaccines, potentially due to easy digital access to information and communication.

What was inadequately known was the effects of using the distinction of rural versus nonmetropolitan in studying public perception of COVID-19 vaccines. The study on social media data relied on the more granular definition of non-metropolitan designation, which looks at the degree of urbanization and proximity to metropolitan areas. This may not be a direct comparison to the survey data, which relied on the general definition of rural designation. Studying survey and social media data methodologies with more recent data after the release of COVID-19 vaccinations and boosters and comparable definitions of geographical designations can provide more insights.

Second, we still do not know the emerging themes within the survey and Twitter data that are associated with attitudes toward COVID-19 vaccination. Both studies relied on major news on specific days to make general assumptions on associations with particular events. Adding more advanced analysis methods, such as probabilistic clustering or topic modeling, can help identify intrinsic themes within the survey and social media data, rather than relying on general event assumptions.

Third, we remain unknown is the overlap in the characteristics of individuals who share their perceptions through surveys versus social media. For example, demographic characteristics were taken for the survey study, but these data were not collected for the social media study. Leveraging other data sources such as electronic health record data and social media data can provide a holistic view of the characteristics of individuals and factors that affect their perceptions.

Conclusion

The focus of this research was important as the public increasingly relies on social media for health information. [40,43,85] Understanding the behavioral patterns and relationship of a pandemic to people’s sentiments and emotions about vaccination was important for improving public health. [86] To the best of our knowledge, this dissertation work was the first to use both survey and social media data at the geographical precision of rural versus urban and nonmetropolitan versus metropolitan designations, in addition to applying novel natural language processing methodologies to perform sentiment and emotion analysis to further our understanding on the United States’ public perceptions about COVID-19 vaccines.

REFERENCES

1. Ten health issues WHO will tackle this year. Accessed April 10, 2019. https://www.who.int/emergencies/ten-threats-to-global-health-in-2019

2. Immunization coverage. Accessed January 29, 2019. https://www.who.int/news-room/fact-sheets/detail/immunization-coverage

3. Coronavirus (COVID-19) statistics data - Google Search Help. Accessed July 1, 2022. https://support.google.com/websearch/answer/9814707?hl=en&visit_id=637922968897923652-373838948&p=cvd19_statistics&rd=1

4. Steele MK, Couture A, Reed C, et al. Estimated Number of COVID-19 Infections, Hospitalizations, and Deaths Prevented Among Vaccinated Persons in the US, December 2020 to September 2021. *JAMA Network Open*. 2022;5(7):e2220385. doi:10.1001/jamanetworkopen.2022.20385

5. MacDonald NE. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*. 2015;33(34):4161-4164.

6. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. *The Lancet*. 2011;378(9790):526-535.

7. Dempsey AF, Schaffer S, Singer D, Butchart A, Davis M, Freed GL. Alternative vaccination schedule preferences among parents of young children. *Pediatrics*. Published online 2011:peds-2011.

8. Gust DA, Darling N, Kennedy A, Schwartz B. Parents with doubts about vaccines: which vaccines and reasons why. *Pediatrics*. 2008;122(4):718-725.

9. Brodeur A, Grigoryeva I, Kattan L. *Stay-at-Home Orders, Social Distancing and Trust*. Social Science Research Network; 2020. Accessed June 25, 2020. https://papers.ssrn.com/abstract=3602410

10. Engle S, Stromme J, Zhou A. *Staying at Home: Mobility Effects of COVID-19*. Social Science Research Network; 2020. doi:10.2139/ssrn.3565703

11. Influenza Pandemics | History of Vaccines. Accessed January 29, 2019. /content/articles/influenza-pandemics

12. Flu broke records for deaths, illnesses in 2017-2018, new CDC numbers show. Washington Post. Accessed January 12, 2019. https://www.washingtonpost.com/national/health-science/last-years-flu-broke-records-for-deaths-and-illnesses-new-cdc-numbers-show/2018/09/26/97cb43fc-c0ed-11e8-90c9-23f963eea204_story.html

13. Vaccines: Vac-Gen/What Would Happen If We Stopped Vaccinations. Published June 29, 2018. Accessed January 29, 2019. https://www.cdc.gov/vaccines/vac-gen/whatifstop.htm

14. Vanderslott S, Roser M. Vaccination. *OWID*. 2019;1(1):e1455.

15. Kirzinger A, Sparks G, Brodie M. KFF COVID-19 Vaccine Monitor- Rural America | KFF. KFF COVID-19 Vaccine Monitor- Rural America. Published April 9, 2021. Accessed August 15, 2021. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-rural-america/

16. Macias Gil R, Marcelin JR, Zuniga-Blanco B, Marquez C, Mathew T, Piggott DA. COVID-19 pandemic: disparate health impact on the Hispanic/Latinx population in the United States. *The Journal of infectious diseases*. 2020;222(10):1592-1595.

17. Kickbusch I, Leung G. Response to the emerging novel coronavirus outbreak. *BMJ*. 2020;368. doi:10.1136/bmj.m406

18. Biasio L. Vaccine hesitancy and health literacy. *Human Vaccines & Immunotherapeutics*. 2017;13(3):701-702. doi:10.1080/21645515.2016.1243633

19. Stets JE. Emotions and Sentiments. In: Delamater J, ed. *Handbook of Social Psychology*. Handbooks of Sociology and Social Research. Springer US; 2006:309-335. doi:10.1007/0-387-36921-X_13

20. Razai MS, Oakeshott P, Esmail A, Wiysonge CS, Viswanath K, Mills MC. COVID-19 vaccine hesitancy: the five Cs to tackle behavioural and sociodemographic factors. *J R Soc Med*. 2021;114(6):295-298. doi:10.1177/01410768211018951

21. Ransohoff DF, Ransohoff RM. Sensationalism in the media: when scientists and journalists may be complicit collaborators. *Effective clinical practice : ECP*. 2001;4(4):185-188.

22. Shen S (Cindy), Dubey V. Addressing vaccine hesitancy: Clinical guidance for primary care physicians working with parents. *Canadian Family Physician*. 2019;65(3):175-181.

23. Salmon DA, Dudley MZ, Glanz JM, Omer SB. Vaccine hesitancy: Causes, consequences, and a call to action. *Vaccine*. 2015;33:D66-D71. doi:10.1016/j.vaccine.2015.09.035

24. Dorell C, Yankey D, Kennedy A, Stokley S. Factors That Influence Parental Vaccination Decisions for Adolescents, 13 to 17 Years Old: National Immunization Survey–Teen, 2010. *Clin Pediatr (Phila)*. 2013;52(2):162-170. doi:10.1177/0009922812468208

25. Saxén H, Virtanen M. Randomized, placebo-controlled double blind study on the efficacy of influenza immunization on absenteeism of health care workers. *Pediatr Infect Dis J*. 1999;18(9):779-783.

26. Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Sources and Perceived Credibility of Vaccine-Safety Information for Parents. *Pediatrics*. 2011;127(Supplement 1):S107-S112. doi:10.1542/peds.2010-1722P

27. Saelee R. Disparities in COVID-19 Vaccination Coverage Between Urban and Rural Counties — United States, December 14, 2020–January 31, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71. doi:10.15585/mmwr.mm7109a2

28. Larson HJ, de Figueiredo A, Xiahong Z, et al. The State of Vaccine Confidence 2016: Global Insights Through a 67-Country Survey. *EBioMedicine*. 2016;12:295-301. doi:10.1016/j.ebiom.2016.08.042

29. Coronado A. Spatial Associations and Network Dynamics Between the Vaccine Exemption Dicsussion in Twitter and the Corresponding Geographic Space. Published online 2017.

30. Peretti-Watel P, Larson HJ, Ward JK, Schulz WS, Verger P. Vaccine Hesitancy: Clarifying a Theoretical Framework for an Ambiguous Notion. *PLoS Curr*. 2015;7:ecurrents.outbreaks.6844c80ff9f5b273f34c91f71b7fc289. doi:10.1371/currents.outbreaks.6844c80ff9f5b273f34c91f71b7fc289

31. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm–An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*. 2012;30(25):3778-3789.

32. Jang SM, Mckeever BW, Mckeever R, Kim JK. From social media to mainstream news: The information flow of the vaccine-autism controversy in the US, Canada, and the UK. *Health communication*. 2019;34(1):110-117.

33. Tang L, Bie B, Park SE, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *American Journal of Infection Control*. 2018;46(9):962-972. doi:10.1016/j.ajic.2018.02.010

34. Kim C, Lee JK. Social media type matters: Investigating the relationship between motivation and online social network heterogeneity. *Journal of Broadcasting & Electronic Media*. 2016;60(4):676-693.

35. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. *JMIR Public Health and Surveillance*. 2016;2(1):e1. doi:10.2196/publichealth.5059

36. Filter realtime Tweets. Twitter Developer. Published 2020. https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/faq

37. Burgoon M, Denning VP, Roberts L. Language expectancy theory. *The persuasion handbook: Developments in theory and practice*. Published online 2002:117-136.

38. Li L, Zhang Q, Wang X, et al. Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo. *IEEE Transactions on Computational Social Systems*. 2020;7(2):556-562. doi:10.1109/TCSS.2020.2980007

39. Himelboim I, Xiao X, Lee DKL, Wang MY, Borah P. A Social Networks Approach to Understanding Vaccine Conversations on Twitter: Network Clusters, Sentiment, and

Certainty in HPV Social Networks. *Health Communication*. 2019;0(0):1-9. doi:10.1080/10410236.2019.1573446

40. Chen T, Dredze M. Vaccine Images on Twitter: Analysis of What Images are Shared. *Journal of Medical Internet Research*. 2018;20(4):e130. doi:10.2196/jmir.8221

41. Berger J, Milkman K. Emotion and Virality: What Makes Online Content Go Viral? in: Marketing Intelligence Review Volume 5 Issue 1 (2013). Published July 16, 2014. https://content.sciendo.com/view/journals/gfkmir/5/1/article-p18.xml

42. Xu Z, Guo H. Using Text Mining to Compare Online Pro- and Anti-Vaccine Headlines: Word Usage, Sentiments, and Online Popularity. *Communication Studies*. 2018;69(1):103-122. doi:10.1080/10510974.2017.1414068

43. Broniatowski DA, Jamison AM, Qi S, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health*. 2018;108(10):1378-1384.

44. How will country-based mitigation measures influence the course of the COVID-19 epidemic? - The Lancet. Accessed June 25, 2020. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30567-5/fulltext?utm_campaign=tlcoronavirus20&utm_content=120403755&utm_medium=social&utm_source=twitter&hss_channel=tw-27013292

45. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Stud Health Technol Inform*. 2015;216:761-765.

46. Bureau UC. What is Rural America? The United States Census Bureau. Accessed August 15, 2021. https://www.census.gov/library/stories/2017/08/rural-america.html

47. Alcendor DJ. Racial disparities-associated COVID-19 mortality among minority populations in the US. *Journal of clinical medicine*. 2020;9(8):2442.

48. Rozenfeld Y, Beam J, Maier H, et al. A model of disparities: risk factors associated with COVID-19 infection. *International journal for equity in health*. 2020;19(1):1-10.

49. Sommers BD. Targeting in Medicaid: The costs and enrollment effects of Medicaid's citizenship documentation requirement. *Journal of Public Economics*. 2010;94(1-2):174-182.

50. Quick Safety Issue 13: Overcoming the challenges of providing care to LEP patients | The Joint Commission. Accessed August 29, 2021. https://www.jointcommission.org/resources/news-and-multimedia/newsletters/newsletters/quick-safety/quick-safety--issue-13-overcoming-the-challenges-of-providing-care-to-lep-patients/

51. Castro-Sánchez E, Chang PW, Vila-Candel R, Escobedo AA, Holmes AH. Health literacy and infectious diseases: why does it matter? *International Journal of Infectious Diseases*. 2016;43:103-110.

52. Hawryluk M. Rural Americans in Pharmacy Deserts Hurting for Covid Vaccines. Kaiser Health News. Published March 3, 2021. Accessed August 15, 2021. https://khn.org/news/article/rural-america-pharmacy-deserts-hurting-for-covid-vaccine-access/

53. Statistics U. Labor force characteristics by race and ethnicity, 2013. Published online 2014.

54. US Bureau of Labor Statistics. Workers who could work at home, did work at home, and were paid for work at home, by selected characteristics, averages for the period 2017–2018. Published online 2018.

55. Fontenot K, Semega J, Kollar M. Income and poverty in the United States: 2017. *Washington DC: US Government Printing Office*. Published online 2018.

56. Kupferschmidt K, Wadman M. Delta variant triggers new phase in the pandemic. *Science*. 2021;372(6549):1375-1376. doi:10.1126/science.372.6549.1375

57. Jacobson HE, Hund L, Mas FS. Predictors of English Health Literacy among US Hispanic Immigrants: The importance of language, bilingualism and sociolinguistic environment. *Literacy & numeracy studies: an international journal in the education and training of adults*. 2016;24(1):43.

58. Largent EA, Persad G, Mello MM, et al. Incorporating Health Equity Into COVID-19 Reopening Plans: Policy Experimentation in California. *Am J Public Health*. 2021;111(8):1481-1488. doi:10.2105/AJPH.2021.306263

59. COVID-19 Race and Ethnicity Data. Accessed January 16, 2022. https://www.cdph.ca.gov/Programs/CID/DCDC/s/COVID-19/Race-Ethnicity.aspx

60. Brumfiel G. Vaccine refusal may put herd immunity at risk, researchers warn. Published online April 7, 2021.

61. Montanaro D. Poll: Americans Don't Trust What They're Hearing From Trump On Coronavirus. *NPR*. https://www.npr.org/2020/03/17/816680033/poll-americans-dont-trust-what-they-re-hearing-from-trump-on-coronavirus. Published March 17, 2020. Accessed August 28, 2021.

62. Czeisler MÉ, Marynak K, Clarke KEN, et al. Delay or Avoidance of Medical Care Because of COVID-19-Related Concerns - United States, June 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(36):1250-1257. doi:10.15585/mmwr.mm6936a4

63. Hong YR, Lawrence J, Jr DW, Iii AM. Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and

National Hospital Survey Data. *JMIR Public Health and Surveillance*. 2020;6(2):e18961. doi:10.2196/18961

64. U.S. Census Bureau QuickFacts: California. Accessed January 16, 2022. https://www.census.gov/quickfacts/CA

65. U.S. Census Bureau QuickFacts: United States. Accessed January 16, 2022. https://www.census.gov/quickfacts/fact/table/US/RHI725219

66. Tolbert J, 2020. Where are the COVID-19 Hotspots? Tracking State Outbreaks. KFF. Published June 30, 2020. Accessed May 30, 2022. https://www.kff.org/coronavirus-covid-19/issue-brief/where-are-the-covid-19-hotspots-tracking-state-outbreaks/

67. CDC. CDC Museum COVID-19 Timeline. Centers for Disease Control and Prevention. Published January 5, 2022. Accessed March 27, 2022. https://www.cdc.gov/museum/timeline/covid19.html

68. CDC. Frequently Asked Questions about COVID-19 Vaccination. Centers for Disease Control and Prevention. Published August 19, 2021. Accessed August 28, 2021. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/faq.html

69. Robinson JS and D. *2 Sentiment Analysis with Tidy Data | Text Mining with R*. Accessed August 5, 2020. https://www.tidytextmining.com/

70. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *Journal of Medical Internet Research*. 2021;23(6):e24435. doi:10.2196/24435

71. Hu T, Wang S, Luo W, et al. Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research*. 2021;23(9):e30854. doi:10.2196/30854

72. Monselise M, Chang CH, Ferreira G, Yang R, Yang CC. Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis. *Journal of Medical Internet Research*. 2021;23(10):e30765. doi:10.2196/30765

73. Saleh SN, McDonald SA, Basit MA, et al. Public Perception of COVID-19 Vaccines through Analysis of Twitter Content and Users. Published online April 22, 2021:2021.04.19.21255701. doi:10.1101/2021.04.19.21255701

74. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine*. 2021;39(39):5499-5505. doi:10.1016/j.vaccine.2021.08.058

75. Su Y, Venkat A, Yadav Y, Puglisi LB, Fodeh SJ. Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities. *Computers in Biology and Medicine*. 2021;132:104336. doi:10.1016/j.compbiomed.2021.104336

76. Twitter API for Academic Research | Products. Accessed March 27, 2022. https://developer.twitter.com/en/products/twitter-api/academic-research

77. USDA ERS - Rural-Urban Continuum Codes. Accessed March 27, 2022. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx

78. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: ; 2015.

79. Mohammad saif. Explore the NRC Word-Emotion Association Lexicon. NRC Word-Emotion Association Lexicon. Published 2009. Accessed February 3, 2020. https://public.tableau.com/views/NRC-Emotion-Lexicon-viz1/NRCEmotionLexicon-viz1?:embed=y&:showVizHome=no&:host_url=https%3A%2F%2Fpublic.tableau.com%2F&:tabs=no&:toolbar=yes&:animate_transition=yes&:display_static_image=no&:display_spinner=no&:display_overlay=yes&:display_count=yes&:showVizHome=no&:showTabs=y&:loadOrderID=0

80. Developer Agreement and Policy – Twitter Developers. Accessed January 17, 2020. https://developer.twitter.com/en/developer-terms/agreement-and-policy

81. 2. Accessing Text Corpora and Lexical Resources. Accessed March 27, 2022. https://www.nltk.org/book/ch02.html

82. Filtering Tweets by location. Accessed March 27, 2022. https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location

83. GET geo/reverse_geocode. Accessed March 27, 2022. https://developer.twitter.com/en/docs/twitter-api/v1/geo/places-near-location/api-reference/get-geo-reverse_geocode

84. USDA ERS - Documentation. Accessed May 30, 2022. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/

85. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *J Med Internet Res*. 2015;17(6):e144. doi:10.2196/jmir.4343

86. Hassan Zadeh A, Zolbanin HM, Sharda R, Delen D. Social Media for Nowcasting Flu Activity: Spatio-Temporal Big Data Analysis. *Inf Syst Front*. Published online January 5, 2019. doi:10.1007/s10796-018-9893-0