Social Media and COVID-19 Vaccine Inoculation Rates

By SHIJIA HUANG*

Does the discussion of the COVID-19 vaccines on social media influence the inoculation rates? This paper addresses this question by conducting sentiment analysis on the Twitter posts around the COVID-19 vaccines in the United States and modeling the inoculation rates. The results show that the tweets, which reflected people's general positive attitudes towards the vaccine, are negatively correlated with the inoculation rates. Furthermore, I analyzed the tweets by State and considered the COVID-19 case and death count to find the State-fixed effects of tweets sentiment and the effect of the pandemic situation on inoculation rates. My results suggest that social media sentiments may discourage people from receiving vaccinations even when the attitude is positive.

The COVID-19 is an ongoing global pandemic declared by the WHO as an international public health emergency that has influenced more than 200 countries and territories (Alabdulla et al., 2021). It has immeasurable negative impacts on public health, economies, trade, and society. For now, the primary strategy to end this pandemic is establishing immunization to populations above a certain percentage (Praveen et al., 2021). As a result, governments worldwide are working on increasing the inoculation rates for COVID-19 through various means, one of them being the vaccine propaganda on social media. There are upsurges of discussion of the COVID-19 vaccines on social media about their principle of action, protection efficacy, side effects, and comparison of different brands. People share and receive each other's opinions on the COVID-19 vaccines on social media, thus influencing their behavior and overall inoculation rates. Therefore, the propaganda of COVID-19 vaccines on social

* Huang: Vancouver School of Economics, University of British Columbia (email: shijia2@student.ubc.ca). Thank for the supervision of Professor Jonathan Graves, Vancouver School of Economics, University of British Columbia.

media might be an effective strategy to encourage people to receive vaccinations.

However, sometimes it would be much more difficult for people to make vaccination decisions after being exposed to more detailed information about the COVID-19 vaccines. Moreover, people may be more cautious about following a popular trend than usual during the decision-making. Therefore, the governments' efforts on vaccination propaganda on social media may be futile or disproportionate with the increase in inoculation rates and thus could lead to misuse of resources. For now, it is unclear whether the discussion of the COVID-19 vaccines on social media can impact people's attitudes towards the vaccine and their behaviors, thus the overall inoculation rates.

Based on the previous information, my research question is: Does the discussion of the COVID-19 vaccines on social media influence the inoculation rates? In particular, after being educated about the features of the vaccines or being exposed to the vaccination boom frequently on social media, would people's primary and secondary demand for the vaccines increase based on these exposures, thus influencing their behavior? Or does it play no significant role or even negatively impact the inoculation rates? By answering the research question, governments can know whether the propaganda of COVID-19 vaccines on social media is an effective strategy to encourage people to receive vaccinations, thus can help to allocate resources wisely to speed up the establishment of herd immunity and avoid misuses of time and money.

This paper has three goals. First, I aim to find people's general attitudes towards the COVID-19 vaccines on social media by conducting sentiment analysis on social media posts. It requires an in-depth analysis of the text of the posts and quantifying the attitudes to find correlations. Second, I aim to investigate whether the effects of social media sentiments on inoculation rates vary by State and whether pandemic situations such as case and death count also influence people's decision on receiving vaccinations. Finally, I aim to investigate the mechanisms for the effect of social media sentiments on vaccination behaviors to suggest solutions to encourage people to receive vaccinations.

The results of my research provided evidence that the discussion of the COVID-19 vaccines on social media is negatively correlated with the inoculation rates. The results also suggest that the effect of States the user located were significantly different, and the number of daily new deaths can negatively impact the inoculation rates. In contrast, the number of daily new cases can positively impact the inoculation rates. The results contribute to improving the policy-making about vaccine propaganda and help establish herd immunity.

The rest of the paper is organized as follows: Section I introduced some background information about the COVID-19 vaccine in the US and the previous studies about the COVID-19 vaccine on social media. Section II introduced the data used and the preliminary findings from the exploratory data analysis. Section III introduced the model specifications and the results. Section IV provides the interpretation of the results, explores alternative specifications, and discusses the research's limitations. Section V concluded the findings and provided suggestions for future research.

I. Background and Literature

A. COVID-19 Vaccine in the US

United States is one of the first countries that start mass COVID-19 vaccinations for its residents. As of February 2021, more than 30 million people had received at least one dose of the COVID-19 vaccine in the US (Nguyen et al., 2021). In addition, there have been plenty of COVID-19 vaccine supplies in the US since the start of mass vaccination (Nguyen et al., 2021). Therefore, the vaccination decision is highly voluntary for people in the US with limited restrictions to receive vaccines. However, a national survey indicates that many people were quite hesitant to receive the COVID-19 vaccines in the US (Nguyen et al., 2021). Therefore, I focused my research on the United States to minimize other factors that could also influence the inoculation rates.

B. COVID-19 Vaccine on Social Media

Despite the numerous negative impacts of the pandemic on the economy and society, one aspect of its influence is that social media plays a more important role in connecting the world. With social distancing, social media brings us together during the pandemic that we rely on to connect to the world. Social media were also the primary source for people receiving COVID-19 related information and updates. Therefore, the discussions of COVID-19 vaccines on social media may influence people's vaccination decisions.

Generally, scholars are concerned about the misinformation and conspiracy on social media that could result in confusion and mistrust, thus could bolster people's hesitancy on receiving the COVID-19 vaccines (Baines et al., 2021; Federico & Esmaeli-Azad, 2021; Lockyer et al., 2021). The thematic analysis of discussions on other social media platforms suggests that several reasons for the refusal of vaccines include its side effect, skeptics about vaccine efficacy, fear of allergic reactions, and population control through COVID-19 vaccines (Baines et al., 2021; Praveen et al., 2021). Another study conducted by Alabdulla et al., 2021 suggests that the socio-demographic determinants could also play an important role in influencing individuals' vaccine hesitancy.

II. Data

To answer the research question, I utilized data from several sources. The primary data (Dataset A) is the tweets posts under the hashtag #CovidVaccine directly extracted from the Twitter platform. I focused my research on Twitter since it is the main social media platform where people share their opinions about the COVID-19 vaccines and receive COVID-19 updates during the pandemic (Praveen et al., 2021). The data contains numerous discussions about the COVID-19 vaccines with the users' location, posting date, and tweets texts. I focused my research only on the United States and filtered out the tweets posts that the user-defined location is in the US. I also mapped the users' location to the associated US States, see Figure 1 for the distribution of the users who posted these tweets by State.

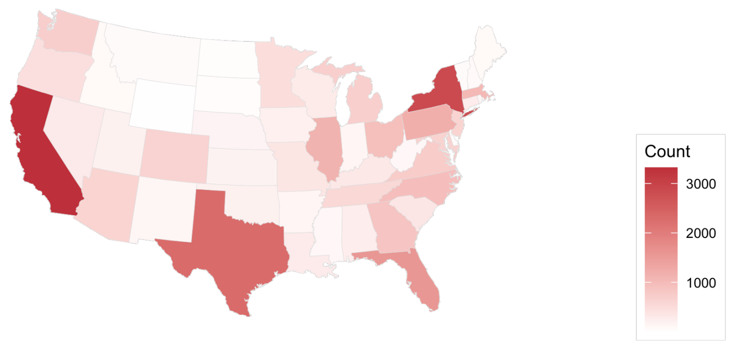


FIGURE 1. TWITTER USERS DISTRIBUTION

Note: The distribution of Twitter users who posted these tweets by State in the United States.

The range of the tweets posting date is between August 2020 and October 2021, which includes both before and after the start of mass vaccination periods in the US. The primary variable of interest is the text of the tweets, which showed the users' attitudes and thoughts about the COVID-19 vaccines. I processed the text using Natural Language Processing techniques to quantify the attitudes and calculated a numeric sentiment score for each tweet.

Other than the tweets data, I also utilized the COVID data (Dataset B) and vaccination data (Dataset C) in the US as supplementary. They provided the daily number of vaccines administrated after the start of comprehensive vaccination in the US, the number of new daily cases and deaths, and the number of total cases and deaths each day. Then I combined the three data sets by location and date and used this data frame to analyze the relationships between tweets sentiment and inoculation rates. The variables in the data frame were explained in Table 1 and the summary statistics in Table 2.

TABLE 1—TABLE OF VARIABLE EXPLANATION

Variable	Description	Source
Date	UTC date when the tweet was created	Dataset A
State	US State where the user-defined location belongs to	Dataset A
User name	The name of the tweet user	Dataset A
User location	The user-defined location for this account's profile	Dataset A
Tweet text	The actual text of the tweet	Dataset A
Sentiment score	Sentiment score of the tweet text	Dataset A
Hashtags	All the other hashtags posted in the tweet along with #CovidVaccine	Dataset A
Total cases	Total number of cases at the tweet posting date and the user's state	Dataset B
New case	Number of new cases at the tweet posting date and the user's state	Dataset B
Total death	Total number of deaths at the tweet posting date and the user's state	Dataset B
New death	Number of new deaths at the tweet posting date and the user's state	Dataset B
Vaccines administered per 100k population	Total number of doses administered per 100,000 census population	Dataset C

Note: Dataset A: Covid Vaccine Tweets (Kaggle, 2020); Dataset B: United States COVID-19 Cases and Deaths by State over Time (CDC, 2020); Dataset C: COVID-19 Vaccinations in the United States, Jurisdiction (CDC, 2020).

III. Empirical Methodology and Results

A. Sentiment Analysis

To analyze the effect of social media discussions around the COVID-19 vaccines on inoculation rates, first we need to investigate people's attitudes towards the COVID-19 vaccine on social media platforms. I conducted sentiment analysis on the text of the tweets and calculated a numeric sentiment score between -10 to 10 for each tweet by using the R package Syuzhet. The positive score indicates the user's positive attitude reflected by the tweet post, and the sentiment score's absolute value indicates the emotion level. I assumed that all tweet posts were posted by actual people and generally expressed the user's honest opinions on COVID-19 vaccines.

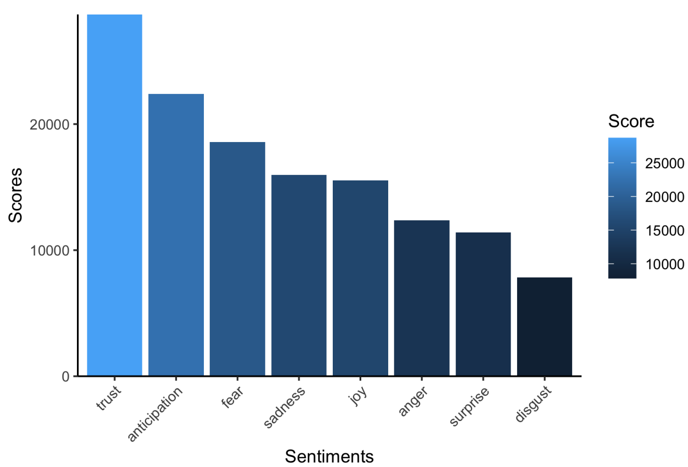
Figure 2 shows the results of the sentiment analysis by producing a bar chart for the overall score of eight different emotions towards the COVID-19 vaccines by calculating the presence of these emotions in the text of the tweets. According to the plot, we can see the top 3 emotions towards the COVID-19 vaccine on Twitter are trust, anticipation, and fear. It indi-

TABLE 2—SUMMARY STATISTICS

Variable	Datatype	Mean	Median	Min	Max	SD
Date	Date	2021-03-22	2021-02-20	2020-08-09	2021-10-22	86.37
State	Character					
User name	Character					
User location	Character					
Tweet text	Character					
Sentiment score	Number	0.319	0.25	-7.3	6.5	1.034
Hashtags	Character					
Total cases	Number	1193019	809897	1350	4826113	1112133
New case	Number	5076	2694	-35	60949	7369.78
Total death	Number	20533	15959	41	70884	17455.49
New death	Number	82	37	-175	721	126.01
Vaccines administered per 100k population	Number	40443	18177	0	152148	42656.78

Note: The negative number in the new and total cases and death caused by the correction from previous day's data

cates that generally, people have a positive attitude towards the vaccine but also have some concerns. Figure 3 shows the most frequent word in the discussion of the vaccine presented as a word cloud. The most frequent word is "get" which can also identify the positive trend of Twitter users towards the vaccine on social media.



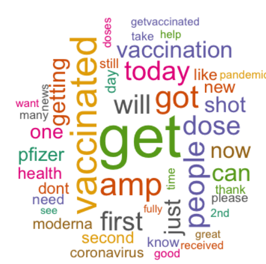


FIGURE 2. SENTIMENTS TOWARDS THE COVID-19 VACCINES

FIGURE 3. TWEETS TEXT

Note: The overall score of eight different emotions towards the COVID-19 vaccines and the most frequent word in the discussion of the COVID-19 vaccine on Twitter

The advantage of sentiment analysis is that it can present attitudes as numeric sentiment scores for each tweet and find its correlation with inoculation rates by fitting a model. However, the scores themselves cannot indicate the main factors that shaped the user attitudes towards the vaccine, thus needing further text investigation.

Figure 4 plotted the average attitudes of users towards the COVID-19 vaccines by States before and after the start of mass vaccination in the US. We can see that before mass vaccination, users in Florida, Delaware, North Dakota, Montana, and Kentucky are more skeptical about COVID-19 vaccines. While after the start of mass vaccination, users generally be more positive towards the vaccine. Another finding is that users in the center US usually be more positive towards the vaccine than users in other regions. Thus the influence of vaccine tweets may vary between different States.

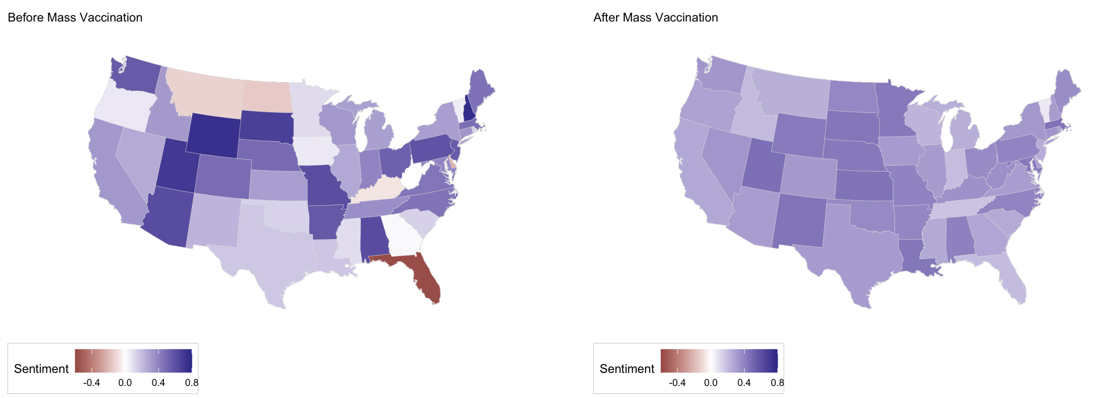
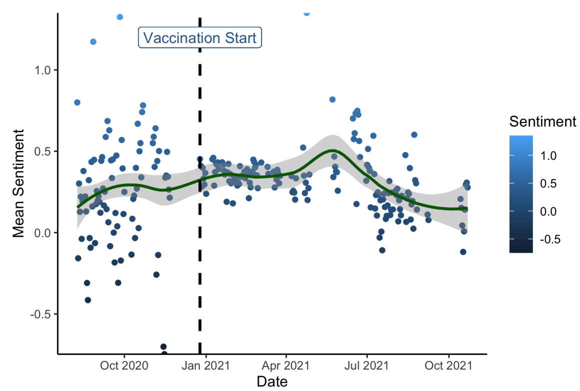


FIGURE 4. SENTIMENT OF USERS TOWARDS THE COVID-19 VACCINES BEFORE AND AFTER MASS VACCINATION

Note: The State-level sentiment were the average of all tweets sentiment score in each State during the period of before and after the start of mass vaccination in the US.

By plotting the data by date for the US as a whole, we can find the attitudes changing trend and investigate its relationship with the trend of other factors. Figure 5 shows the time series trend of attitudes towards the vaccine, and the vertical line indicates the time for the start of mass vaccination. Figure 7 shows the daily number of vaccines administrated per 100K populations. Figure 6 and Figure 8 show the number of new cases and death each day.



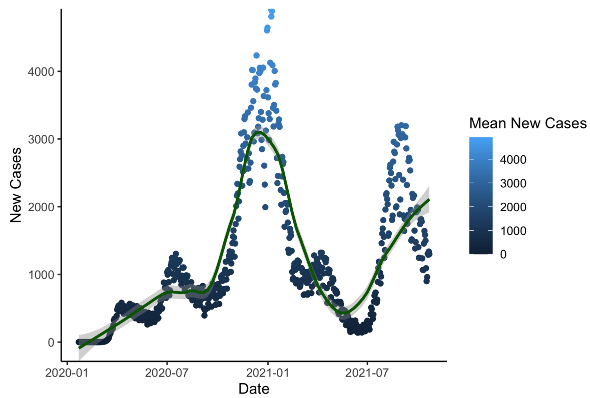
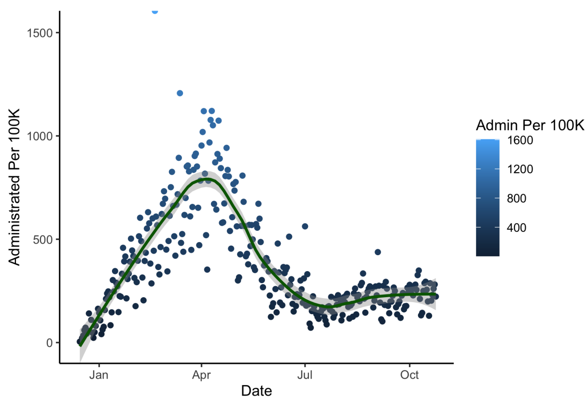


FIGURE 5. SENTIMENT TREND

FIGURE 6. COVID-19 CASES TREND



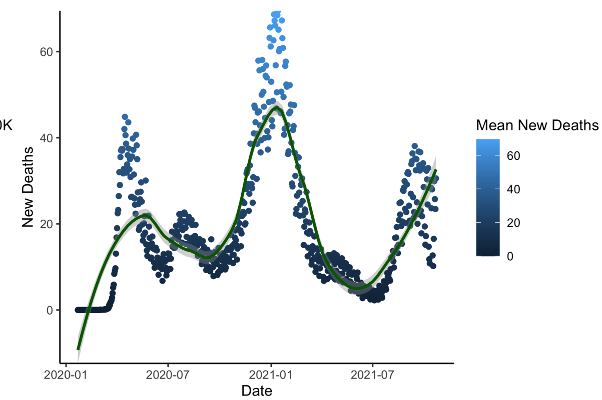


FIGURE 7. COVID-19 INOCULATION RATES TREND

FIGURE 8. COVID-19 DEATH TREND

Note: Figure 5: time series trend of attitudes towards the vaccine, and the vertical line indicates the time for the start of mass vaccination. Figure 7: daily number of vaccines administrated per 100K populations. Figure 6 and Figure 8: number of new cases and death count each day.

After comparing the plots, I found a positive trend towards the COVID-19 vaccine before mass vaccination. The positive trend continued to grow but dropped suddenly after June 2021, after the start of mass vaccination. Interestingly, there are also spikes of new cases and death after June 2021, which suggest that the number of new cases and death may negatively influence people's attitudes towards the vaccine. In addition, there is an increase in the daily number of vaccines administrated per 100K population after June/July 2021, which may be related to the drop in users' attitudes on Twitter.

B. Model

To study the impact of tweets sentiment on inoculation rates across different States and dates, I estimated a panel regression model as the baseline specification. The dependent variable $VacRate_{it}$ is the inoculation rates, measured by the number of vaccines administrated per 100,000 census population in the corresponding State i and date t. In the baseline model, $State_i$ are the unobserved factors that are invariant with date t across the State $i = 1,..,n$. I aim to estimate the coefficient β_1 to measure the impact of tweets sentiment $Sentiment_{it}$. Note that since the inoculation rates are zero before the start with mass vaccination in the US, I fitted the following model using only the data collected after the start of mass vaccination and grouped by date and State with computed average sentiment score $Sentiment_{it}$ for State i and date t to minimize the bias of extreme personal attitudes. The term u_{it} is the error term that varies by State i and date t.

$$VacRate_{it} = \beta_0 + \beta_1 Sentiment_{it} + \beta_2 State_i + u_{it} \tag{1}$$

Column (1) of Table 3 presents the results of estimating the effect of tweets sentiment on inoculation rates for the baseline model. I focused on the State-level impact for only four different States: California, Florida, Texas, and New York. The reason is that most of the tweets were posted by users in California, Texas, and New York. And Florida has a major attitude change after the start of mass vaccination. To simplify the analysis, comparing the effect of tweets sentiment on inoculation rates between the four States can give a general overview of the difference in States' effects. The results indicate that the sentiment of the tweets is negatively correlated with the inoculation rates with a 5% significance level. Moreover, the effects caused by factors in California and New York were significant at a 5% level. Since the coefficients β_2 for these two States were positive, this model suggests that California and New York users were less likely to be negatively influenced by the social media sentiments that discourage them from receiving vaccinations.

To estimate the effect of a change in $Sentiment_i$ on $VacRate_i$ holding the $State_i$ constant, I fitted a fixed effect model for estimation of the relation between tweets sentiment and inoculation rates by letting $StateFE_i = \beta_0 + \beta_2 State_i$ as the fixed effect of State i over time. Therefore $StateFE_i$ are the State specific intercepts that capture heterogeneity across different States including demographic factors.

$$VacRate_{it} = \beta_1 Sentiment_{it} + StateFE + u_{it} \tag{2}$$

Column (2) of Table 3 presents the results of estimating the effect of tweets sentiment on inoculation rates for the fixed effect model. The coefficient estimation for β_1 is the same as the baseline model in that the sentiment of the tweets is negatively correlated with the inoculation rates with a 5% significance level. Now the coefficients β_2 for these two States were higher than before with a 1% significance level. Therefore, the effect of different States that the user located would be quite different on the inoculation rates.

IV. Discussion

A. Effects of Tweets Sentiment

Based on the previous analysis, my results suggest that the discussion of the COVID-19 vaccines on social media may discourage people from receiving vaccinations even when the attitude is positive. This might result from the mistrust of people on social media content, thus causing their hesitancy on receiving the vaccinations (Baines et al., 2021). Therefore, the propaganda of the COVID-19 vaccines on social media may not be an effective strategy to increase the vaccination rates. However, due to the limitation of the sentiment analysis, the exact reason that social media negatively impacts the inoculation rates needs to be found

TABLE 3—MODEL RESULTS

	Dependent variable:			
	Vaccines administered per 100k population			
	(1)	(2)	(3)	(4)
Sentiment	$-1,724.787^{**}$	$-1,724.787^{**}$	$-1,290.897^{*}$	$-1,409.545^{*}$
	(812.054)	(812.054)	(771.113)	(760.451)
New death			-356.208^{***}	-435.412^{***}
			(22.074)	(28.183)
New case			1.299^{***}	-1.650^{***}
			(0.325)	(0.369)
New death : New case				0.010^{***}
				(0.001)
State: CA	$11,627.730^{**}$	$66,870.350^{***}$	$118,902.000^{***}$	$124,937.500^{***}$
	(5,817.456)	(3,925.877)	(3,268.803)	(3,469.505)
State: FL	7,338.340	$62,580.970^{***}$	$94,358.230^{***}$	$114,820.300^{***}$
	(5,611.346)	(3,606.172)	(3,805.377)	(3,890.586)
State: NY	$13,962.540^{**}$	$69,205.160^{***}$	$78,729.570^{***}$	$89,331.380^{***}$
	(5,910.924)	(4,064.443)	(3,386.655)	(3,024.234)
State: TX	1,491.749	$56,734.380^{***}$	$91,421.290^{***}$	$110,557.400^{***}$
	(5,388.032)	(3,258.853)	(3,236.377)	(3,472.703)
Constant	$55,242.630^{***}$			
	(4,305.995)			
Observations	6,192	6,192	6,192	6,192
R^2	0.045	0.635	0.686	0.697
Adjusted R^2	0.037	0.632	0.684	0.695
Residual Std. Error	44,119.220	44,119.220	40,922.900	40,196.030
	(df = 6139)	(df = 6139)	(df = 6137)	(df = 6136)
F Statistic	5.583^{***}	201.838^{***}	244.222^{***}	252.631^{***}
	(df = 52; 6139)	(df = 53; 6139)	(df = 55; 6137)	(df = 56; 6136)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

by a deep investigation of the Twitter posts texts and the collection of more data.

B. Effects of New Cases and New Death

To consider the effect of pandemic situation as well, I fitted the fixed effect model with extra variables $Case_{it}$ and $Death_{it}$. The variable $Case_{it}$ is the number of new cases confirmed for State i and date t, and the variable $Death_{it}$ is the number of new death confirmed for State i and date t.

$$VacRate_{it} = \beta_1 Sentiment_{it} + \beta_2 Case_{it} + \beta_3 Death_{it} + StateFE + u_{it} \tag{3}$$

Column (3) of Table 3 presents the results of estimating the effect of tweets sentiment and the effect of COVID-19 cases and death on inoculation rates for the model. The results indicate that the sentiment of the tweets is still negatively correlated with the inoculation rates with a 10% significance level. The results show that the number of daily new deaths negatively impacts the inoculation rates at a 1% significance level. This can result from the death count making people question the effectiveness of the COVID-19 vaccine to prevent bad outcomes and death, thus harming their trust in the vaccine. On the other hand, the number of daily new cases positively impacts the inoculation rates at a 1% significance level. This can be interpreted with the fact that when there is a growing new case count, people start to be concerned with the pandemic and start to ask for vaccines' help.

Since the number of new deaths and cases has quite different effects on the inoculation rates, I want to study their combined effect by adding the interaction term $Case_{it} \times Death_{it}$ to the fixed effect model. It represents the interaction effect between new cases and death for State i and date t on the inoculation rates.

$$VacRate_{it} = \beta_1 Sentiment_{it} + \beta_2 Case_{it} + \beta_3 Death_{it} + \beta_4 Case_{it} \times Death_{it} + StateFE + u_{it} \tag{4}$$

Column (4) of Table 3 presents the results of estimating the effect of tweets sentiment and the effect of COVID-19 cases and death on inoculation rates for model 4. The results indicate that the sentiment of the tweets is still negatively correlated with the inoculation rates with a 10% significance level. The coefficient for the interaction term β_4 is 0.01 with a 1% significance level. We can see the interaction effect between new cases and death was positive but relatively small to influence the inoculation rates.

C. Limitations

A few limitations in the research can provide suggestions for future improvement. First, the distribution of the tweets extracted was not equal across States due to the available data. This might result in bias when the data were dominated by Twitter users in a few significant States. Second, the model was fitted using only the data after the start of mass vaccination to have a positive inoculation rate as the dependent variable. The results only reflected the effect of social media on inoculation rates during a specific period when the pandemic situation improved with vaccines' help. A better representation of people's decision on receiving the vaccines needs to be created other than the inoculation rates to find the effect before the start of mass vaccination.

V. Conclusions

Does the discussion of the COVID-19 vaccines on social media influence the inoculation rates? In conclusion, I found that social media discussion about the COVID-19 vaccines is negatively correlated with the inoculation rates. This might result from mistrusting social media content, thus questioning the information about the vaccines. The paper also explores the differences caused by the user's location and the effect of the pandemic situation. The results suggest that the effect of States the user located were significantly different, and the number of daily new deaths may negatively impact the inoculation rates. In contrast, the number of daily new cases can positively impact the inoculation rates.

For future research opportunities, I suggest that other than using the inoculation rates, which can be influenced by many other factors such as vaccine supply and labor efficiency, there could be another variable of interest to represent people's vaccination decisions, such as the vaccine booking information or survey results. This would allow the study on the effect of social media even before starting the mass vaccination.

REFERENCES

Alabdulla, M., Reagu, S. M., Al-Khal, A., Elzain, M., & Jones, R. M. (2021). COVID-19 vaccine hesitancy and attitudes in Qatar: A national cross-sectional survey of a migrant-majority population. *Influenza and Other Respiratory Viruses, 15(3)*, 361–370. https://doi.org/10.1111/irv.12847

Baines, A., Ittefaq, M., & Abwao, M. (2021). Scamdemic, #Plandemic, or #Scaredemic: What Parler Social Media Platform Tells Us about COVID-19 Vaccine. *Vaccines (Basel), 9(5)*, 421. https://doi.org/10.3390/vaccines9050421

Centers for Disease Control and Prevention. (2020). COVID-19 Vaccinations in the United States, Jurisdiction (Version 2020-10-24). [Data file]. Retrieved from https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc

Centers for Disease Control and Prevention. (2020). United States COVID-19 Cases and Deaths by State over Time (Version 2020-10-24). [Data file]. Retrieved from https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36

Federico, F., & Esmaeli-Azad, B. (2021). Social Media Misinformation and Conspiracy Theories pose a Great Threat to Successful COVID-19 Vaccination Among Adults ages 18-23. *The FASEB Journal, 35(S1)*. https://doi.org/10.1096/fasebj.2021.35.S1.04634

Freed, G. L. (2021). Actionable lessons for the US COVID vaccine program. Israel *Journal of Health Policy Research, 10(1)*, 14. https://doi.org/10.1186/s13584-021-00452-2

Kaggle. (2020). Covid Vaccine Tweets (Version 36). [Data file]. Retrieved from https://kaggle.com/kaushiksuresh147/covidvaccine-tweets

Lockyer, B., Islam, S., Rahman, A., Dickerson, J., Pickett, K., Sheldon, T., Wright, J., McEachan, R., Sheard, L., & Group, the B. I. for H. R. C.-19 S. A. (2021). Understanding COVID-19 misinformation and vaccine hesitancy in context: Findings from a qualitative study involving citizens in Bradford, UK. *Health Expectations, 24(4)*, 1158–1167. https://doi.org/10.1111/hex.13240

Nguyen, K. H., Srivastav, A., Razzaghi, H., Williams, W., Lindley, M. C., Jorgensen, C., Abad, N., & Singleton, J. A. (2021). COVID-19 vaccination intent, perceptions, and reasons for not vaccinating among groups prioritized for early vaccination—United States, September and December 2020. *American Journal of Transplantation, 21(4)*, 1650-1656. http://dx.doi.org/10.1111/ajt.16560

Praveen, S., Ittamalla, R., & Deepak, G. (2021). Analyzing the attitude of Indian citizens towards COVID-19 vaccine – A text analytics study. *Diabetes & Metabolic Syndrome Clinical Research & Reviews, 15(2)*, 595–599. https://doi.org/10.1016/j.dsx.2021.02.031