MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231111

Facebook Post Credibility as a Predictor of Vaccine Hesitancy in the US

Ana ALEKSANDRIC[a,b], Anisha DANGAL[b], Shirin NILIZADEH[a] and Gabriela MUSTATA WILSON[b,1]
[a] *Department of Computer Science and Engineering, The University of Texas at Arlington, United States*
[b] *Multi-Interprofessional Center for Health Informatics, The University of Texas at Arlington, United States*
ORCiD ID: Gabriela Mustata Wilson https://orcid.org/0000-0003-3090-9242

Abstract. Vaccine hesitancy represents a barrier to public health efforts aiming to mitigate the pandemic by performing global interventions. One of the reasons behind vaccine hesitancy is mistrust towards the health system that partially originated due to the misinformation shared over the internet. This study examined the association between the credibility of the sources regarding the COVID-19 vaccine posted on social media and the vaccination rate at the state level in the United States. Study findings suggest that sharing more Facebook posts with links to low-credibility sources about vaccination is associated with a lower number of new vaccinations at the state level in the US. This indicates an urgent need for social media-leveraged interventions through which public health officials can share reliable information to educate populations about vaccine benefits and reduce vaccine hesitancy.

Keywords. Social media, misinformation, vaccine hesitancy, covid-19

1. Introduction

The World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020 [1]. However, doubts about the reliability and effectiveness of the quickly developed vaccines have caused a significant number of US citizens to indicate a reluctance to become vaccinated against COVID-19 in addition to a section of the population that opposes immunizations [2].

Misinformation plays a significant role in vaccine reluctance. Moreover, many people use social media to find information and news [3], making these platforms suitable environments for the rapid spread of false information, yielding a high probability that individuals will come across misleading information regarding vaccination [4]. Also, vaccine hesitancy in Jordan and Kuwait was shown to be significantly impacted by social media rumors [5]. Furthermore, individuals tend to believe in conspiracy and religious misinformation, which makes them avoid proposed behavioral guidelines [6]. Therefore, social media disinformation is powerful in reducing vaccination coverage over time [7] while raising public health concerns regarding the impact of anti-vaccination content on platforms that increase vaccine hesitancy [8]. Thus, this study aims to provide more insights into the effect of

[1] Corresponding author details: Gabriela Mustata Wilson, email: gabriela.wilson@uta.edu.

misinformation spread on social media regarding the credibility of the sources that are shared over the network while also measuring the impact of social determinants of health (health literacy and social vulnerability index) and US state race and ethnicity composition. To accomplish that, data were collected from Facebook, one of the most popular social media platforms, and each post URL was labeled with its credibility. Finally, a regression model was used to test the formulated hypothesis.

2. Methods

2.1. Data Collection

This study examines how the presence of low-credibility sources on social media is associated with the vaccination rate at the state level in the US. Therefore, CrowdTangle [9] was used to collect Facebook posts, which is a social media listening platform allowing academics to collect posts from the most influential Facebook pages and public groups. Therefore, for each state in the US, we queried CrowdTangle using the set of keywords previously developed in another study to collect vaccine-related posts on Twitter [10]. In addition, only posts written in English and those shared in 2021 were selected. The reason behind analyzing only posts in English is the variety of sources that other language speakers might use to obtain information (e.g., the Spanish-speaking population might use different sources than the English-speaking population). Therefore, the final dataset consisted of 858,622 English Facebook vaccine-related posts that contained links other than www.facebook.com.

The next step required labeling the data URLs shared within posts with their credibility. A *low* label was assigned to all posts that contained a link to the website classified as mixed or low credibility. In contrast, a *high* label was given to all posts that included a website that was classified as high-credibility according to source credibility lists collected by OSoMe [10] utilizing the Iffy+ Misinfo/Disinfo list created by Iffy.news [11]. Finally, all posts that contained a link to the source not present in any of these lists were labeled as *unknown*. The total number of unique URL sources in the Facebook dataset is 641,602. The number of unique high-credibility sources in the dataset is 58,466 (9.1%), while the number of unique low-credibility sources is 5,947 (0.9%). Posts containing *unknown* sources were discarded from the analysis as they could impact the results due to no indication of their credibility. Consequently, the dataset size was reduced to 95,578 Facebook posts containing 64,413 unique URLs.

2.2. Hypothesis Formulation

As shown in previous studies, misinformation spread on social media might negatively affect vaccination acceptance. Therefore, this study compares the impact of low-credibility posts with the impact of high-credibility sources on new weekly vaccinations in the US states. Thus, the study tests the following hypothesis:

States in the US with more Facebook posts *with low-credibility sources and fewer posts with high-credibility sources tend to have lower vaccination rates.*

2.3. Variables

The vaccination increase (new vaccinations) per week is used as a dependent variable in the statistical model, as the research aims to pinpoint the correlation

between new weekly vaccinations per state with respect to the credibility of social media sources posted in the same state. Vaccination rate is obtained from *Our World in Data* [12], and the measure used is a weekly increase in *total_vaccinations_per_hundred,* which is the total number of doses administrated per 100 people in the total population of the state.

Health Literacy (HL) is the ability to locate, comprehend, and use the information required to make necessary health-related decisions [13], which can have an effect on the choice of sources where individuals look for health-related information. HL was collected from the University of North Carolina at Chapel Hill [14].

Social Vulnerability Index (SVI) is a measure of negative consequences for a society whenever outward tension appears to have an impact [15]. Therefore, SVI has emerged to be a control variable in this study as it can play a significant role in the US vaccination rate. The data was collected from CDC/ATSDR [15].

Race and Ethnicity Composition is the proportion of the population with regard to the distinct racial groups residing within the US (percentage of Hispanic, Asian, or Black population per state). Thus, these measures are used as control variables in statistical models as they might have an impact on vaccine hesitancy at the state level. Data was downloaded from the United States Census Bureau [16].

The independent binary variable used in the regression analysis is **credibility**, which contains two categories: *low* and *high.* The distribution of low-credibility and high-credibility sources per state is presented in **Figure 1**.

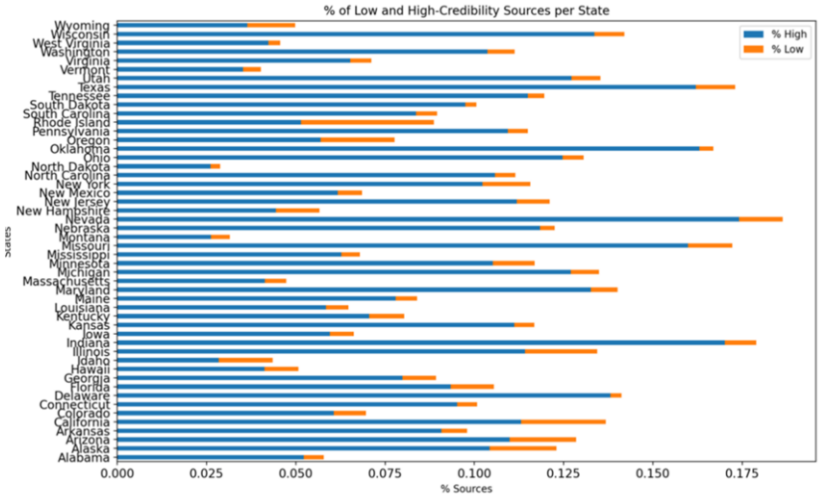

Figure 1. The distribution of low-credibility and high-credibility sources in 2021 per state**.**

As shown in **Figure 1**, certain states contain a larger number of low-credibility sources compared to other states raising concerns about what other factors (e.g., SVI & HL) might have led to this phenomenon and the effects that this might have on the vaccination rate in these states.

3. Results

In the final dataset, each post contained the following attributes: credibility, state and week of the year when the post was shared, new vaccinations in that state in that week, HL, SVI, and race and ethnicity composition of that state. Therefore, a linear regression model was employed with the new vaccinations as a numeric dependent variable. In addition, the binary independent variable used in the model is credibility being *low* or *high*, while all control variables were included in the model. Moreover, standard errors were clustered by state and week to avoid any data dependence occurring due to the data repetition that could potentially impact the results. The model revealed a statistically significant relationship between new weekly vaccinations and source credibility. In more detail, a significant negative relationship implies that the presence of low credibility sources at the state and weekly level is associated with a lower number of new weekly vaccinations compared to the presence of high credibility sources, which are associated with a higher number of new weekly vaccinations ($p < 0.001$). The results of the regression model are presented in **Table 1**.

Table 1. The association between source credibility and new weekly vaccinations.

Variable	Estimate	Std. Error	P- value
Credibility - Low	-0.41	0.09	P < 0.001
Health Literacy	0.03	0.06	0.6
SVI	-1.56	1.09	0.15
Hispanic Population (%)	1.9	1.24	0.13
Black Population (%)	0.07	1.35	0.96
Asian Population (%)	-0.002	2.61	0.99

4. Discussion

The negative correlation between low-credibility sources and new vaccinations might indicate that social media users in certain areas do not distinguish between low and high-credibility sources, or they might not be aware of where to search for reliable health-related information. Surprisingly, HL, SVI, and race and ethnicity composition did not yield any significance in the model. The reason behind it might be that the effect of credibility is much higher compared to mentioned control variables, or it is due to the generally low number of states. This is the first study that establishes a connection between post credibility on Facebook and new weekly vaccinations at the state level in the US. The study findings emphasize the importance of social media listening, which should be utilized in future emergencies to develop interventions in real-time and rapidly overcome the crisis.

Intuitively, this study contains certain limitations. For example, posts collected are only in English and they only belong to Facebook pages and public groups, so they do not provide any insight into the credibility of the sources that are shared by the general Facebook population. Moreover, detecting misinformation relies on URL credibility, discarding a big portion of the data that does not contain any URLs.

5. Conclusions

Even though vaccination represents one of the most effective ways to combat the pandemic, vaccine hesitancy is a big problem for public health officials. This study examined the association between the credibility of the sources posted on Facebook and the vaccination rate at the state level in the US. Despite the obvious limitations of this study, statistical analysis yielded relevant results that show that low-credibility sources are associated with a lower vaccination rate compared to high-credibility sources. This analysis provides a better understanding of the serious impact that misinformation shared on social media has nationwide. It might be an initial step in creating successful interventions targeting specific populations which are excessively exposed to and influenced by misleading information. Combating online misinformation can help educate society about vaccination importance and reduce vaccine hesitancy in the US.

References

[1] Ghebreyesus TA. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19-25 May 2020.
[2] Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. Public Health. 2021 May;194:245-51, doi: 10.1016/j.puhe.2021.02.025.
[3] Elisa Shearer, Amy Mitchell. News Use Across Social Media Platforms in 2020 2021. https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/ (accessed November 5, 2022).
[4] Lee SK, Sun J, Jang S, Connelly S. Misinformation of COVID-19 vaccines and vaccine hesitancy. Sci Rep. 2022 Aug;12(1):13681, doi: 10.1038/s41598-022-17430-6.
[5] Sallam M, Dababseh D, Eid H, Al-Mahzoum K, Al-Haidar A, Taim D, Yaseen A, Ababneh NA, Bakri FG, Mahafzah A. High Rates of COVID-19 Vaccine Hesitancy and Its Association with Conspiracy Beliefs: A Study in Jordan and Kuwait among Other Arab Countries. Vaccines (Basel). 2021 Jan;9(1):42, doi: 10.3390/vaccines9010042.
[6] Barua Z. COVID-19 Misinformation on Social Media and Public's Health Behavior: Understanding the Moderating Role of Situational Motivation and Credibility Evaluations. Hu Arenas. 2022 May;1-24, doi: 10.1007/s42087-022-00291-w.
[7] Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health. 2020 Oct;5(10):e004206, doi: 10.1136/bmjgh-2020-004206.
[8] Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother. 2020 Nov;16(11):2586-93, doi: 10.1080/21645515.2020.1780846.
[9] CrowdTangle n.d. https://www.crowdtangle.com/ (accessed November 5, 2022).
[10] DeVerna MR, Pierri F, Truong BT, Bollenbacher J, Axelrod D, Loynes N, Torres-Lugo C, Yang KC, Menczer F, Bryden J. CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines. In Proceedings of the International AAAI Conference on Web and Social Media 2021 May 22 (Vol. 15, pp. 992-999).
[11] Iffy+. Iffy+ Mis/Disinfo Sites n.d. https://iffy.news/iffy-plus/ (accessed November 5, 2022).
[12] Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, Giattino C, Rodés-Guirao L. A global database of COVID-19 vaccinations. Nat Hum Behav. 2021 Jul;5(7):947-53, doi: 10.1038/s41562-021-01122-8.
[13] CDC. What Is Health Literacy? 2020. https://www.cdc.gov/healthliteracy/learn/index.html (accessed November 5, 2022).
[14] National Health Literacy Mapping to Inform Health Care Policy. Health Literacy Data Map 2014. http://healthliteracymap.unc.edu/# (accessed November 5, 2022).
[15] CDC. CDC/ATSDR Social Vulnerability Index n.d. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html (accessed November 5, 2022).
[16] United States Census Bureau. State Population by Characteristics: 2010-2019 n.d. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html (accessed November 5, 2022).