Rochester Institute of Technology

Current Vaccine Sentiments Varying by two Geographic Regions in the United States

By

Michael McTighe

Submitted in partial Fulfillment for the Course Requirements of STAT 611: Statistical software R

Instructor: Huaye Li Ph.D.

**Current Vaccine Sentiments Varying by two Geographic Regions in the United States**

**Overview**

**Novel Corona Virus Pandemic**

During December 2019 clusters of individuals living in the Wuhan province of the People’s Republic of China began developing a pneumonia of unknown origin. Respiratory tract samples taken from the afflicted were analyzed and found to be possessive of a novel zoonotic RNA virus first referred to as 2019-nCoV, the seventh type of corona virus that has the capacity to infect humans (Zhu et al., 2020). Preliminary postmortem biopsies revealed that this novel virus, now referred to as SARS-CoV-2 or COVID 19, could cause deleterious and fatal damage in the alveoli of lungs (Tian et al., 2020). Additional symptoms include elevated body temperature, heavy cough, fatigue, and difficulty capturing one’s breath (Singhal, 2020). Additionally, Singhal (2020) reported that the transmission rate from human to human was higher than previous members of the corona virus family. Given the high transmutability, international infection rates increased rapidly leading the World Health Organization to declare COVID-19 a global pandemic on March 11th 2020.

**Public Health Implications**

At the time of this writing, it is estimated that the reported number of total cases in the United States alone is approximately 45 million with approximately 736,000 total deaths (CDC, 2021). Beyond the medical implications, the global pandemic has caused profound economic hardships across a variety of geographic and geopolitical areas such as income insecurity, food insecurity, labor shortages, and less accessibility to resources and much more (Ali et al., 2020; Biskanaki et al., 2020). Moreover, the threat of physical illness, experienced economic fallout, and social isolation from government-imposed lockdowns have had deleterious psychological consequences includes exacerbated levels of anxiety, depression, stress, negative affect as well as general increases in is levels of life disruption (Kornilaki 2021). All these issues combined have led many to believe that the COVID-19 pandemic has been the most harrowing public health crisis of the past century thereby creating the need for directed research approaches aimed at ameliorating the deleterious effects of the virus.

**Discovery of a vaccine**

Within weeks of the World Health Organization’s declaration of the pandemic, scientists and medical doctors began investigating the efficacy of various treatments for SARS-CoV-2. The later months of 2020 brought about great speculation regarding possible therapies which were endorsed by prominent public figures. Extensive studies exploring the efficacy of drugs like Hydroxychloroquine and Remdesivir dominated health and medicine (Annangi, 2020; Nili et al., 2020), however, questionable research methods and a general lack of consensus allowed for a silver bullet solution to the crisis to remain illusive over a year after the first observed case of COVID-19.

Given the high demand for a safe and effective vaccine, BioNTech and Pfizer developed a coordinated research paradigm designed to provide exactly that. Together, the companies conducted a phase 1 clinical trial on German and American citizens and observed a significant antibody response to the SARS-CoV-2 virus following vaccination using a novel RNA-vaccine (Walsch et al., 2020). The safety and efficacy observed in the study promptly influenced a series of supplemental research which found the vaccine to be about 95% effective following two consecutive inoculations (Polack et al., 2020). In early 2021, the FDA began authorizing emergency use of COVID-19 vaccines which was soon followed by full FDA approval.

As of the current writing is it believed that approximately 58 percent of the United States population have reached full vaccination status (Mayo Clinic, 2021). Epidemiological inquiries have observed that in the fully vaccinated group, there are significantly fewer infections as well as significantly fewer medical complications, hospitalizations, and deaths resulting from breakthrough cases when compared to the non-vaccinated group (Scobie et al., 2021). Though the decision to vaccine should appear intuitive from a public health perspective, socio-political roadblocks leading to a reluctancy to inoculate have led to challenges in reaching the other 42 percent of the public.

**Attitudes toward vaccines**

Given the discovery of a reasonable solution to the medical prevention of the virus, many are saying that vaccine hesitancy is now one of the most prominent issues in overcoming the pandemic (Kreps et al., 2021). Much hesitancy is believed to come from the fact that that research on the COVID-19 vaccine was fast-tracked leaving the long-term implications of inoculation uncertain Kreps et al., (2021). However, a myriad of other variables were determined to predict vaccination status including but not limited to such things as age, sex, income, education level, trust in government, race, endorsement of conspiracy related beliefs, trust in science and medicine, and other demographics (Allington et al., 2021). Additional research has explored attitudes toward covid vaccines, and found they predict vaccination status which leads investigators to suggest that further classification of groups based on attitudes toward vaccines could help guide campaigns aimed at behavioral change (Danabal et al., 2021). Investigations into vaccine attitudes prior to the onset of the 2020 pandemic has found general positive attitudes toward vaccines worldwide, however, certain regions show significant variability in their sentiments (Larson et al., 2016). These findings, taken together, state the importance of further exploration into how vaccine sentiments or attitudes differ by geographical region, so that intervention strategies can allocate resources in a logistically effective manner.

**This Current Study**

The current study seeks to investigate vaccine sentiments and how they differ in valence across two different geopolitical, socioeconomic, and geographic regions. Using a text mining approach, text strings were extracted from the social media platform Twitter using a Twitter developer account connected to R statistical software. 1000 tweets were analyzed from the liberal state California on the pacific coast of the United States and another 1000 were extracted from the southern conservative state of Florida located on the Atlantic coast. Both states are possessive of vast demographical difference. Following a cleaning procedure, and the creation of a set of functions, tweets were coded by sentiment using text libraries developed by Hu and Liu (2004). The primary question of interest in this research is as follows. Do sentiments about vaccines vary significantly by geographic region? The research is strictly exploratory and no directionality about the regions is declared, however, given the variance in demographics, it is predicted that vaccine sentiments will differ significantly based on where they are in the world.

**Methods**

**Data collection and cleaning.**

The ‘TwitteR’ package developed by Gentry, (2016) was installed within R statistical software and a direct method of connection was created using sensitive keys and tokens distributed on Twitter’s developer page. Two objects were created using the “searchTwitter” application. Each consisted of 1000 extracted tweets possessive of the key word “vaccine” posted on “2021-10-15” from two different locations. One object consisted of 1000 tweets posted from the pacific southwest, and the tweets met inclusion criteria if they were posted within a 500-mile radius from the California city of Los Angeles. The other object consisted of 1000 tweets posted in the Atlantic southeast, and the tweets met inclusion criteria if they were posted within 500 miles of the Florida City of Miami. Tweets were further extracted into a vector of text and standardized on formatting.

Both objects then went through similar cleaning processes using the lapply feature and applying functions across the text. First, text was standardized on font style. Next, any hypertext was taken out and replaced with nothing. Thirdly, any hashtags included in the tweet were removed and replaced with nothing. The text was then unlisted and created as a vector. Single tweets were cleaned using lapply to extract punctuation marks, control characters, numerical digits, and special characters. Additional data was collected on retweet status, number of retweets and number of likes the tweet were obtained.

**Sentiment analysis.**

Collections of positive and negative words developed by Hu & Liu (2004) were taken from a working directory and were used to analyze sentiment across the two groups of tweets. A function was created that converted the text of the tweets into vectors of scores. Words within tweets were compared to the collections of positive negative words using true/false statements. A match with a positive word would add a sentiment score of +1. A match with a negative word would add a sentiment score of -1. A total score object was created to quantify the sum of the total sentiment per tweet. A data frame was then created by text and sentiment.

**Analysis**

The two data sets from both locations were combined into a single data set using the rbind application. Sentiment scores for both variables were subjected to mean comparison. Histograms were created to visually inspect the difference between the two regions. An independent samples t-test was used to assess statistical significance between groups.

**Results**

Sentiments toward vaccines varied based on their geographic region. Average sentiment scores were significantly higher in Miami (*M* = 0.01) compared to Los Angeles (*M* = -0.15); *t*(1989) = -3.46, *p* < .001. Visual analysis of histograms suggests positive skewness in the Los Angeles data set with more tweets being generally negative (see Figure 1). The data, taken together, suggests attitudes toward vaccines are generally more positive in the southeast compared to the southwest.

**Figure 1.** *Frequency of tweets by total sentiment score*

Chart, diagram

Description automatically generated

**Discussion**

The general hypothesis related to the primary research question was supported. Different geographic regions within the United States were possessive of different attitudes toward vaccines. Based on the findings, the American southwest had more negative sentiments toward vaccines than the American southeast. Though the results indicate that areas around Florida are more positive toward vaccines than areas around California, both groups possessed average sentiment scores that were closer to attitudinal neutrality than general positivity or negativity.

Widely available political messaging would suggest that a largely conservative state like Florida would have more negative attitudes toward vaccines than a liberal state such as California. However, the research presented by Allington et al (2021) suggests several reasons why Florida would have fewer negative attitudes. First, vaccine hesitancy is associated with youth, and due to increased risk of death, people over the age of 65 view vaccines more positively. According to the U.S census bureau (2019) the percentage of residents over the age of 65 is much higher in Florida (20.9%) than California (14.8%). Secondly, white Americans are reported to have less vaccine hesitancy (Allington et al.., 2021), and the percentage of white residents is also higher in Florida (77.3%) compared to California (71.9%). Thirdly, Allington (2021) states that vaccines hesitancy is associated with lower levels of education. Again, according to the Census Bureau (2019) California has fewer residents possessive of a high school diploma (83.3%) compared to Florida (88.2%). However, current theory should suggest against the results based on income whereby we usually see more hesitancy in low-income groups compared to high income groups (Allington et al., 2021), and we know median household annual income is higher in California ($75,253) compared to Florida (&55,660). Overall, the results are supported by existing research and theory.

The investigation is not without any sets of limitations. Much of the language used to address public health matters such as vaccines are neither positive or negative in valence, and frequently used terms such as COVID are novel and don’t fit within the existing libraries developed Hu and Liu. Beyond that, sampling issues exist. Though the combined samples are large (2000) and are likely to capture even a minute effect size in mean difference, it is possible the tweets were still unrepresentative of the general geographic region. R’s ability to only extract 1000 tweets at a time presents a limitation in accurate attitude identifications. Moreover, the radius used in extracting tweets was 500 miles which is vast. Though it accurately collects data from the general region as intended, it is likely to be divested of important information related to the exact city under study.

Natural Language Processing (NLP) has fundamental limitations built into it, for it often fails to explore contextual information surrounding the positive and negatively coded words. However, the approach itself appears to be a promising method toward capturing attitudes toward vaccines. Preliminary investigations exploring text mining in the context of COVID-19 research found that it provided a method of exploring virus tracking by using various “symptoms” as keywords (Guo et al., 2020), and the authors declare that more studies use this approach toward overcoming the pandemic. Future research using Text mining should continue to look at regional differences in sentiments toward vaccines. By identifying problems regions where vaccine attitudes are generally negative, public health officials can better direct educational and interventional campaigns aimed at ameliorating vaccine hesitancy.

**References**

Ali, A., Ahmed, M., Hassan, M. (2020). Socioeconomic impact of COVID-19 pandemic: Evidence from rural mountain community in Pakistan. *Journal of Public Affairs 2020,* 1 – 9. <https://doi.org/10.1002/pa.2355>

Allington, D., McAndrew, S., Moxham-Hall, V., Duffy, B. (2021). Coronavirus conspiracy suspicions, general vaccine attitudes, trust, and coronavirus information source as predictors of vaccine hesitancy among UK residents during the COVID-19 Pandemic. *Psychological Medicine. DOI: 10.1017/S9933291721991434*

Annangi, S. (2020) Chloroquine and hydroxychloroquine for COVID-19: A word of caution. *Respirology.* 25, 683 – 684. DOI:10.1111/resp.13845

Biskanaki, F., Rallis, E., Andreous, E., Sfyri, E., Teripi, N., Kefala, V. (2020) Social-economic impact of COVID-19 pandemic on aesthetic centers in Greece. *Journal of Cosmetic Dermatology*. 19, 2165 – 2168. https;//doi.org/10.1111/jocd.13517

Center for Disease Control and Prevention (2021)

Chagla, Z. (2021). The BNT162n2 (BioNTech/Pfizer vaccine had 95% efficacy against COVID-19 > 7 days after the 2nd dose. *Annals of Internal Medicine.* 172(*2*).

Danabal, K.G.M., Magesh, S.S., Saravanan, S., Gopichandran, V. (2021). Attitude toward COVID 19 vaccines and vaccine hesitancy in urban and rural communities in Tamil Nadu, India – a community based survey. *BMC Health Services Research.* 21, 1 – 10. <https://doi.org/10,1186/s12913-021-07037-4>

Gentry, J. (2016, August 29). Package ‘twitteR’. Retrieved 2021, from <https://cran.r-project.org/web/packages/twitteR/twitteR.pdf>

Guo, J. W., Radlof, C. L. Wawrzynski, S. W., Cloyes, K. G. (2020). Mining twitter to explore the emergence of COVID-19 symptoms. *Public Health Nursing*. 37, 934 - 940. DOI:10.1111/phn.12809

Hu, M., Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining. *Aug 22 – 25.* Seattle, Washington, U.S.A

Kreps, S., Dasgupta, N., Brownstein, J. S., Hswen, Y., Kriner, D. L. (2021). Public attitudes toward COVID – 19 Vaccination: The role of vaccine attributes, incentives, and misinformation. *Nature Partner Journal.* 6(*73*), 6 – 7. <https://doi.org/10.1038/s41541-021-00335-2>

Kornilaki, K. N. (2021). The psychological effect of COVID-19 quarantine on Greek young adults: Risk factors and the protective role of daily routine and altruism. *International Journal of Psychology,* 2021, 1 – 10*. DOI: 10.1001/ijop.12767*

Larson, H. J., Figueiredo, A., Xiahong, Z., Schulz, W. S., Verger, P., Johnston, I. G., Cook, A. R., Jones, N. S. (2016). The state of vaccine confidence 2016: Global insights through a 67-country survey. *EBioMedicine.* 12, 295 – 301. <http://dx.doi.org/10.1016/j.ebiom.2016.08.042>

Mayo Clinic (2019)

Nili, A., Farbod, A., Neishabouri, A., Mozafarihashjin, M., Tavakolpour, S., Mahmoudi, H. (2020). RemdesivirL a beacon of hope from Ebola virus disease to COVID-19. *Reviews in Medical Virology*. 30, 1 – 13. <https://doi.org/10.1002/rmv.2133>

Scobia, H. M., Et. Al., (2021). Monitoring incidence of COVID-19 cases, hospitalizations, and deaths by vaccination status – 13 U.S jurisdictions, April 4 – July 17, 2021. *Morbidity and Mortality Weekly Report*. 70(*37*), 1284 – 1290.

Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics.* 87(*4)*, 281 – 286. <https://doi.org/10.1007/s12098-020-03263-6>

Tian, S., Xiong, Y., Lui, H., Nio, L., Guo, J., Liao, M., Xiao, S. Y. (2020). Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. *Modern Pathology.* 33, 1007 – 1014. <https://doi.org/10.1038/241379-020-0536-x>

U.S. Census Beurau (2019)

Walsh, E. E., Frenck, R. W., Falsey, A. R., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Mulligan, M. J., Bailey, R., Swanson, K, A., Li, P., Koury, K., Kalina, W., Cooper, D., Fontes-Garfias, C., Shi, P-Y., Tureci, O., Tompkins, K. R., Lyke, K., Raabe, V., Dormitzer, P. R., Jansen, K. U., Sahin, U., Gruber, W. C. (2020). Safety and immunogenicity of two RNA-Based COVID-19 vaccine candidates. *New England Journal of Medicine.* 383, 2439 - 2450

Zhu, N., Zhang, D., Wang, W., Xingwang, L., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Phil, G., Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019). *New England Journal of Medicine.* 382(*8*), 727 – 733. DOI: 10.1056/NEJMoa2001017