Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Vaccine 39 (2021) 4034–4038

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Short communication

Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States

Sharath Chandra Guntuku [a,b,f,*], Alison M. Buttenheim [c,d,f], Garrick Sherman [b], Raina M. Merchant [a,e,f]

[a] Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA, United States
[b] Department of Computer and Information Science, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, United States
[c] Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
[d] Center for Health Incentives and Behavioral Economics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[e] Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[f] Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia PA, United States

ARTICLE INFO

Article history:
Received 8 April 2021
Received in revised form 4 June 2021
Accepted 5 June 2021
Available online 9 June 2021

Keywords:
COVID-19 vaccination
Twitter
Natural language processing
Machine learning

ABSTRACT

The speed at which social media is propagating COVID-19 misinformation and its potential reach and impact is growing, yet little work has focused on the potential applications of these data for informing public health communication about COVID-19 vaccines. We used Twitter to access a random sample of over 78 million vaccine-related tweets posted between December 1, 2020 and February 28, 2021 to describe the geographical and temporal variation in COVID-19 vaccine discourse. Urban suburbs posted about equitable distribution in communities, college towns talked about in-clinic vaccinations near universities, evangelical hubs posted about operation warp speed and thanking God, exurbs posted about the 2020 election, Hispanic centers posted about concerns around food and water, and counties in the ACP African American South posted about issues of trust, hesitancy, and history. The graying America ACP community posted about the federal government's failures; rural middle American counties posted about news press conferences. Topics related to allergic and adverse reactions, misinformation around Bill Gates and China, and issues of trust among Black Americans in the healthcare system were more prevalent in December, topics related to questions about mask wearing, reaching herd immunity and natural infection, and concerns about nursing home residents and workers increased in January, and themes around access to black communities, waiting for appointments, keeping family safe by vaccinating and fighting online misinformation campaigns were more prevalent in February. Twitter discourse around COVID-19 vaccines in the United States varied significantly across different communities and changed over time; these insights could inform targeted messaging and mitigation strategies.

1. Introduction

Authorization and rollout of the first vaccines against the severe acute respiratory syndrome coronavirus (SARS-CoV-2) that causes COVID-19 commenced in December 2020 in the United States (US). Three vaccines have received emergency use authorization to mitigate COVID-19 in the US, and large-scale clinical trials are in progress for two other vaccines. While the number of individuals expressing hesitancy about the vaccine has been decreasing [1], concern about COVID-19 vaccines remains high. In a culturally diverse society such as the United States, salience in messaging requires a deep understanding of cultural nuance across groups, and the development of targeted messaging based on those nuances. The use of social media data is expanding exponentially due to its low barrier to entry [2] and has the potential to be harnessed to deliver precision public health communication that addresses dynamically changing misinformation and heterogenous belief systems across communities. During COVID-19, Twitter has been utilized to measure changes in mental health [3], to identify misinformation [4], study psychosocial effects [5], and to uncover emerging symptoms [6]. In this paper, we study online discourse about the COVID-19 vaccine using Twitter to gain insight into variation across communities and over time.

* Corresponding author at: 3330 Walnut St., Philadelphia, PA 19104, United States.
E-mail address: sharathg@upenn.edu (S.C. Guntuku).

https://doi.org/10.1016/j.vaccine.2021.06.014
0264-410X/

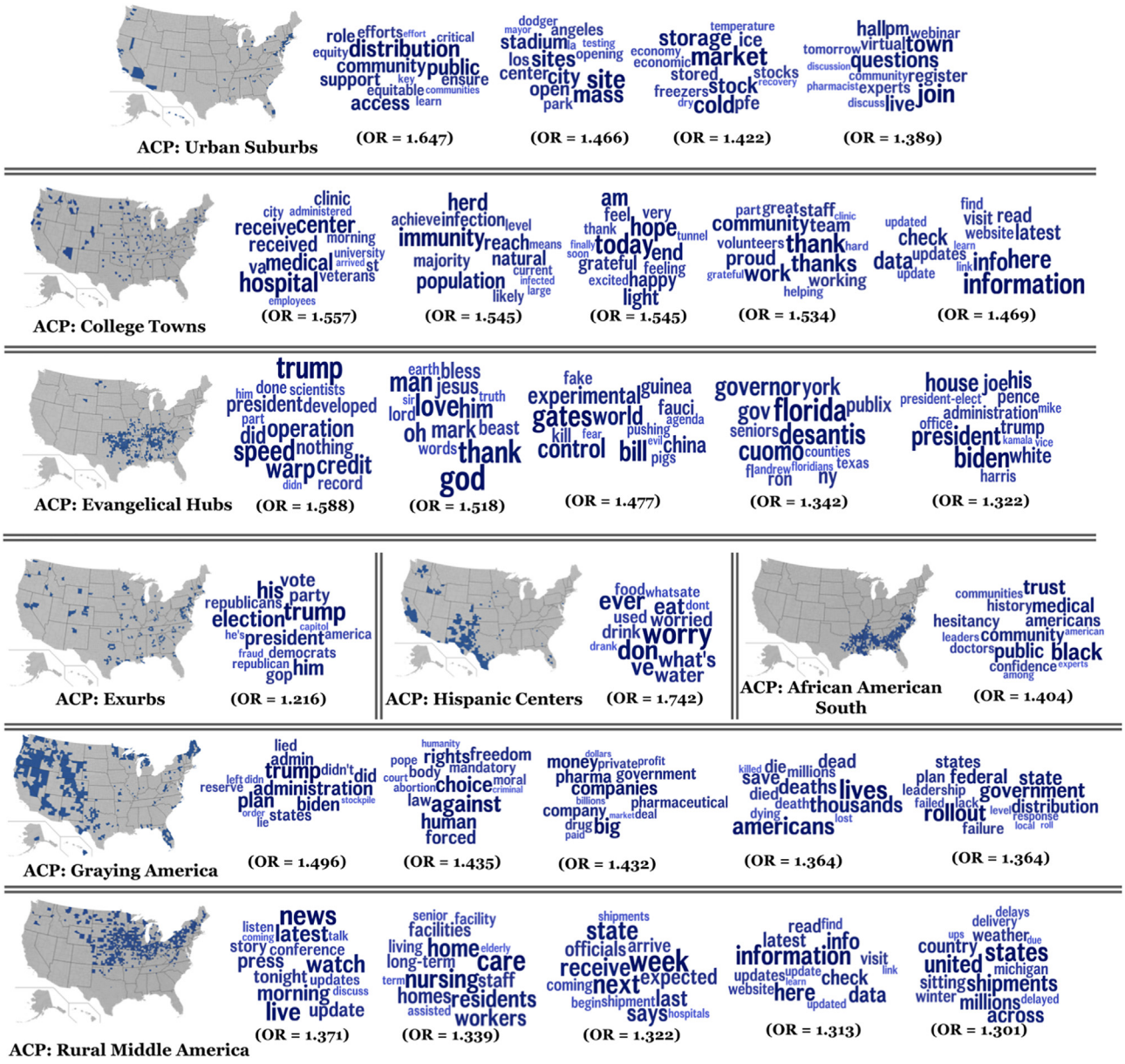

Fig. 1. COVID-19 vaccine topics associated with eight ACP communities showing significant differences in vaccine topics, along with their corresponding odds ratios (OR). Only top five significant topics per ACP sorted by OR after Benajamini-Hochberg p-correction ($p < 0.05$) are shown. Higher *odds ratio (OR)* indicates a stronger association of topic with the ACP community compared to other ACP communities. All topics along with 95% CIs are shown in Supplementary Table S1.

2. Methods

Using publicly available data, we identified over 78.1 million vaccine-related messages posted on Twitter between December 1, 2020 to February 28, 2021. We geolocated tweets posted in the US to different counties using location information available for each tweet from the Twitter API [7]. We then identified words including emoticons and created a set of one hundred open vocabulary data-driven word clusters (topics) with Latent Dirichlet Allocation on original tweets [8]. We then extracted the weekly prevalence of topics across tweets aggregated to US counties. Heterogeneity in communities is not necessarily spatial - i.e., a metropolitan area and a rural location a few miles away can be more distinct than two metro areas several hundred miles apart. Consequently, we obtained 15 community types identified by the American Communities Project (ACP), which is a non-spatial proximity-based county-level clustering using 36 demographic, cultural, and socio-economic indicators, including income, race, education, ethnicity, religious affiliation, etc [9]. Each county is assigned a membership to one of the 15 communities by the ACP. The data used to define the type in the ACP came from two sources: the U.S. Census American Community Survey[10] and the Religious Congregations and Membership Study [11]. We also obtained vaccination rates per county from the CDC [12] between December 13, 2020 and June 3, 2021 and calculated weekly averages across all counties in each ACP community. Review of this study was waived by the University of Pennsylvania's institutional review board because it is based on publicly available data.

We took a data-driven approach to allow for a more transparent view of the topics that differentiate geographic and temporal trends. Topics were used as input in a logistic regression model with dummy variables for each of the ACP communities as the outcomes for geographical analyses and for each week as the outcomes for temporal analyses. We considered counties where

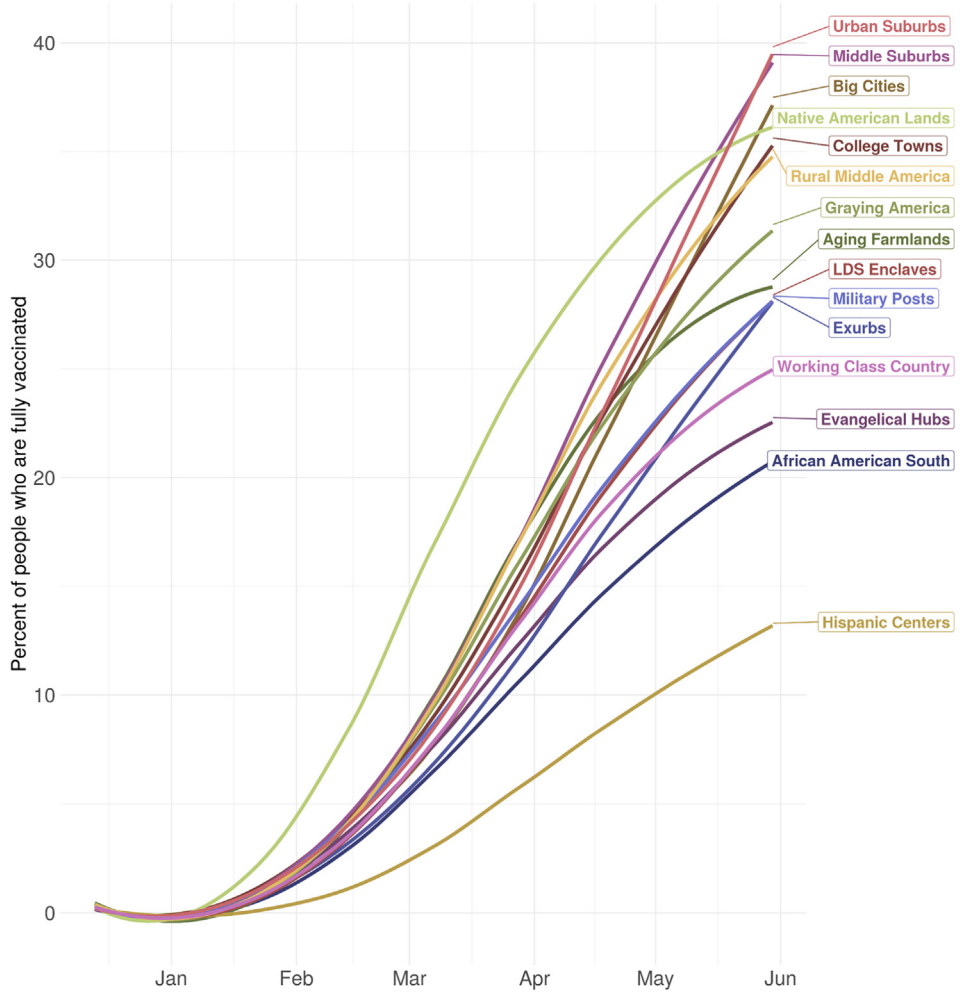

Fig. 2. Percentage of people who received two doses of the COVID-19 vaccine in each ACP community. Data was obtained from CDC between December 13, 2020 and June 3, 2021 for counties and aggregated to weeks across ACP communities.

tweets from those counties total at least 500 words per week for analyses and *p*-value of < 0.05, after adjusting for multiple comparisons using Benjamini-Hochberg's multitest correction, as a heuristic for identifying potentially meaningful associations [13]. We report effect sizes of each topic in terms of odds ratio (OR) along with 95% confidence intervals to quantify the differences. Higher *OR* indicates a stronger association of a topic with each ACP community for the geographical and each week for the temporal analyses.

3. Results

Of 9.6 million vaccine-related tweets posted from 2958 counties in the United States from December 1, 2020 to February 28, 2021, there were 4 million original tweets (non-retweet) from 2957 counties. 1853 counties had at least 500 words per week.

Eight ACP communities showed significant differences in vaccine topics (Fig. 1). Urban suburbs posted about equitable distribution in communities (OR = 1.65 (1.61, 1.69), p < 0.05), mass vaccine sites (OR = 1.47 (1.43, 1.5), p < 0.05), cold storage (OR = 1.42 (1.38, 1.46), p < 0.05), and live public town hall webinars with experts (OR = 1.39 (1.35, 1.43), p < 0.05). College towns talked about in-clinic vaccinations near universities (OR = 1.56 (1.52, 1.6), p < 0.05), likelihood of reaching herd immunity (OR = 1.55 (1.51, 1.59), p < 0.05), feeling hopeful (OR = 1.55 (1.51, 1.59), p < 0.05), expressing gratitude to community volunteers (OR = 1.54 (1.5, 1.58), p < 0.05), and data and information tracking (OR = 1.47 (1.43, 1.51), p < 0.05). Evangelical hubs posted about operation warp speed (OR = 1.59 (1.55, 1.63), p < 0.05), thanking God (OR = 1.52 (1.48, 1.56), p < 0.05), conspiracy theories around Bill Gates and China (OR = 1.48 (1.44, 1.52), p < 0.05), local (OR = 1.34 (1.31, 1.38), p < 0.05), and federal administration (OR = 1.32 (1.29, 1.36), p < 0.05). Exurbs posted about the 2020 election (OR = 1.22 (1.18, 1.25), p < 0.05), Hispanic centers posted about concerns around food and water (OR = 1.74 (1.7, 1.79), p < 0.05), and counties in the ACP African American South posted about issues of trust, hesitancy, and history (OR = 1.4 (1.37, 1.44), p < 0.05). US counties with mostly retirees, termed Graying America in the ACP schema, posted about the federal government's failures (OR = 1.49 (1.46, 1.53), p < 0.05), personal choice and freedom (OR = 1.43 (1.4, 1.47), p < 0.05), big pharma (OR = 1.43 (1.4, 1.47), p < 0.05), and deaths (OR = 1.36 (1.33, 1.4), p < 0.05). Rural counties from Maine to the Great Lakes to Washington, termed Rural Middle America by the ACP, posted about news press conferences (OR = 1.37 (1.34, 1.41), p < 0.05), nursing homes and long term senior resident facilities (OR = 1.34 (1.3, 1.38), p < 0.05), states receiving vaccines shipments (OR = 1.32 (1.29, 1.36), p < 0.05), data and information tracking (OR = 1.31 (1.28, 1.35), p < 0.05), and delays in shipments (OR = 1.3 (1.27, 1.34), p < 0.05). While Fig. 1 shows only top 5 topics per ACP community sorted by OR, the list of all significant topics along with 95% CIs is shown in Supplementary Table S1.

Fig. 3. Weekly variation of data-driven COVID-19 Twitter topics from December 1, 2020 to February 28, 2021. Significant topics are colored according to their association with each week after Benajamini-Hochberg correction ($p < 0.05$). Each row represents a topic, each column represents a week, and each cell represents an odds ratio between both.

Fig. 2 shows the vaccination rates documented by CDC grouped by ACP communities from December 13, 2020, until June 3, 2021. Urban Suburbs, Middle Suburbs, and Big Cities lead with over 35% of the population being fully vaccinated, while Evangelical Hubs, African American South have lower than 25%, and Hispanic Centers have lower than 15% vaccination rates as of June 2021.

Fig. 3 shows variation of Twitter vaccine topics over different weeks from December 2020 to February 2021. Topics related to allergic and adverse reactions, misinformation around Bill Gates and China, and issues of trust among Black Americans in the healthcare system were higher in December; topics related to questions about mask wearing, reaching herd immunity and natural infection, and concerns about nursing home residents and workers increased in January. Themes around access to black communities, waiting for appointments, keeping family safe by vaccinating and fighting online misinformation campaigns were more prevalent in February.

4. Discussion

Discourse around COVID-19 vaccines in the United States varies significantly across different geographic communities and is changing over time. Hesitancy and acceptance of vaccines and, in particular the COVID-19 vaccine, has varied by access, sociodemographic, and cultural factors [14]. Much public health messaging for the COVID-19 vaccine is being developed based on behavior change models that incorporate health beliefs, social norms, self-efficacy [15–17]. However, our results suggest that messaging campaigns should also incorporate dynamic news cycles as well as cultural markers in messaging that often signify in-group affiliation. The speed at which social media is propagating COVID-19 related misinformation, and its potential reach and impact necessitate nimble, real-time, and adaptive approaches for messaging. Going beyond the data, language used on social media in different communities could be indicative of current and future vaccination rates. Limitations of this study include that Twitter is not representative of the general population in the US and the tweets analyzed in this stream are a random sample provided by the Twitter API. Social media provides an opportunity to understand the rapidly evolving public information spaces across diverse populations and communities that can inform targeted messaging and mitigation strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

S.C.G acknowledges support from Google Cloud.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2021.06.014. Dataset of LDA topics generated in this paper is available at https://github.com/wwbp/covid_vaccine_lda_topics.

References

[1] KFF COVID-19 Vaccine Monitor: December 2020. Published December 15, 2020. Accessed March 18, 2021. https://www.kff.org/coronavirus-covid-19/report/kff-covid-19-vaccine-monitor-december-2020/
[2] Koeze E, Popper N. The Virus Changed the Way We Internet. The New York Times. https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html. Published April 8, 2020. Accessed March 18, 2021.
[3] Guntuku SC, Sherman G, Stokes DC, et al. Tracking Mental Health and Symptom Mentions on Twitter During COVID-19. J Gen Intern Med. 2020;35(9):2798–800.
[4] Kouzy R, Jaoude JA, Kraitem A, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*. Published online 2020. https://doi.org/10.7759/cureus.7255.
[5] Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media. J Med Internet Res. 2020;22(11):e22600.
[6] Santosh R, Andrew Schwartz H, Eichstaedt JC, Ungar LH, Guntuku SC. Detecting Emerging Symptoms of COVID-19 using Context-based Twitter Embeddings. *arXiv [csCL]*. Published online November 8, 2020.
[7] Schwartz HA, Eichstaedt JC, Kern ML, et al. Characterizing Geographic Variation in Well-Being Using Tweets. ICWSM 2013:583–91.
[8] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J machine Learning res 2003;3:993–1022.
[9] Chinni D, Gimpel J. Our Patchwork Nation: The Surprising Truth About the "Real" America. Penguin; 2011.
[10] Julian T, Kominski R. Education and synthetic work-life earnings estimates. American community survey reports. ACS-14. *US Census Bureau*. Published online September 2011. Accessed May 27, 2021.
[11] Grammich CA. 2010 U.S. Religion Census: Religious Congregations & Membership Study : An Enumeration by Nation, State, and County Based on Data Reported for 236 Religious Groups. Association of Statisticians of American Religious Bodies; 2012.
[12] CDC. Reporting county-level COVID-19 vaccination data. Published May 14, 2021. Accessed June 4, 2021. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/distributing/reporting-counties.html
[13] Schwartz HA, Andrew Schwartz H, Giorgi S, et al. DLATK: Differential Language Analysis ToolKit. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Published online 2017. doi:10.18653/v1/d17-2010
[14] Kreps S, Prasad S, Brownstein JS, et al. Factors Associated With US Adults' Likelihood of Accepting COVID-19 Vaccination. JAMA Network Open. 2020;3(10):. https://doi.org/10.1001/jamanetworkopen.2020.25594e2025594.
[15] Zampetakis LA, Melas C. The health belief model predicts vaccination intentions against COVID-19: A survey experiment approach (aphw.12262). Appl Psychol Health Well Being. 2021. https://doi.org/10.1111/aphw.12262.
[16] Wong LP, Alias H, Wong P-F, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother. 2020;16(9):2204–14.
[17] Sherman SM, Smith LE, Sim J, et al. COVID-19 vaccination intention in the UK: Results from the "COVID-19 Vaccination Acceptability Study" (CoVAccS), a nationally representative cross-sectional survey. doi:10.1101/2020.08.13.20174045