

FORECASTING DISTRUST: A Nationwide Analysis of the Climate of Vaccine Hesitancy

Alexander Reid Ross

Senior Research Fellow,
Network Contagion Research Institute

Peter Norvig

Director of Research, Google Inc.

Aaron Bernstein

Senior Director, Insights and Advocacy, Walmart

Jacob Shapiro

Professor of Politics and International Affairs,
Princeton University

Alex Goldenberg

Senior Intelligence Analyst,
Network Contagion Research Institute;
Research Fellow, Miller Center for
Community Protection and Resilience,
Rutgers University

Malav Modi

Research Fellow,
Network Contagion Labs at Rutgers University

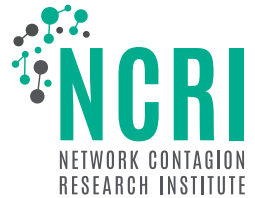
Joel Finkelstein

Chief Science Officer and Director,
Network Contagion Research Institute;
Senior Research Fellow, Miller Center for
Community Protection and Resilience,
Rutgers University

PRESENTED BY



POWERED BY



INTRODUCTION

Widespread institutional distrust along with fears of Covid vaccine side effects create continued hesitancy to take the Covid vaccine. This reluctance hampers efforts to fully vaccinate the U.S. population and manage the Covid-19 pandemic. Inspired by distinct, regional trends in disinformation and misinformation, this continued reluctance to vaccinate, combined with dangerous new Covid variants, is causing a new surge in caseloads and deaths.

Large-scale, data driven modeling and analysis is essential to combating Covid vaccine reluctance. Data driven spatial models of Covid vaccine reluctance have shown efficacy in predicting where and why it persists.¹ Data-rich performance models of specific Covid vaccine distributors can help us understand which distributors are performing best for specific reluctant populations in a region-specific fashion. Such analyses are crucial to creating an informed and effective vaccination strategy.

We contribute a unique perspective on this challenge by running the same analysis on two different measures of vaccination: CDC data which captures all vaccination activity and one large-scale vaccine distributor's activities. By comparing these two data streams in the same model we can see where particular organizations are having success and thus learning how they can contribute to addressing vaccine reluctance.

Vaccine reluctance is challenging to model however. To begin with, a wide variety of distinct population data and trends (including demographic information, political mobilizations, and voting patterns) are likely to predict Covid vaccine reluctance and these data would need to be carefully curated. Online signifiers such as Google searches and Tweets are becoming increasingly critical to understanding covid infection rates,² and these are non-trivial to collect and analyze.

Another challenge in predicting Covid vaccine reluctance is measuring the impact of conspiracy theories and the groups that subscribe to them. These groups also often disseminate anti-vaccine misinformation and disinformation on social media. Groups such as QAnon can take hold quickly and foster distrust within vulnerable communities.

It is intuitive that participation in such groups should predict Covid vaccine hesitancy. However, absent better modes of measurement, the structure of online conspiracy groups and the fact that their communications are pervasive and often coded makes it difficult to assess whether and where group participation predicts Covid vaccine reluctance in any given region. Tracking and monitoring misinformation and disinformation requires expansive data collection from forums both popular and specialized, and necessitates immersive comprehension of subtle cues.

Finally, efforts to develop data-rich models on the performance of specific vaccine distributors are faced with additional challenges. Although the CDC releases up-to-date maps of national vaccination rates on the county level, data on the performance of corporate distributors is proprietary. Using only CDC data leaves researchers unable to determine which distributors are succeeding at reaching specific and reluctant populations. Without proprietary data, we are in the dark about what makes vaccine distributors successful.

In order to overcome these challenges, the Network Contagion Research Institute created a large-scale and data-driven climate model of vaccine reluctance. This model combines data from the CDC, search data from the Google public health team, scaled analyses of online conspiracy trends, aggregated event and demographic databases, and the Walmart's vaccine distribution data.

Using CDC data, we are able to predict the underlying factors associated with vaccine reluctance in a region-specific and county-specific fashion. Using vaccine distribution data, we can determine what factors may predict its success with vaccine administration on a county level. Finally, we can compare these models to reveal the relative strengths and advantages that Walmart may have in reaching reluctant populations.

Online Conspiracy Activity, Anti-vaccine Protests, Google Searches for Vaccine Side Effects and Political and Racial Leanings Predict Where Lower Levels of Vaccination are Likely

Using the Centers for Disease Control and Prevention's COVID Data Tracker, we first obtained data on vaccination by county in the United States. These data were then used to generate a map showing the level of vaccination by county (figure 3).

We then performed a linear regression to determine what of the following factors predict lower vaccine administration by county:

- Percentage of Trump voters from Townhall.com³
- Percentage of African Americans from 2019 American Community Survey (US Census Bureau)⁴
- Percentage of uninsured from Small Area Health Insurance Estimates (US Census Bureau)⁵
- Percentage of population reporting poor health from Behavioral Risk Factor Surveillance System (CDC)^{6A}
- Data from the Centers for Disease Control on distribution of Flu Vaccine^{6B}
- Data on Google searches for vaccine side effects published by the Google public health team⁷
- Data on Google searches for vaccine availability and/or location published by the Google public health team⁸
- Rate of Anti-Vaccine protests and mobilizations from ACLED⁹
- Data on religious community composition from U.S. Religious Census¹⁰
- All Tweets (excluding retweets) for a 30 day period starting in July 14th 2020, employing the QAnon conspiracy meme: “WWG1WGA” from NCRI social media collections.

We used a dot-and-whisker plot with scaled, standardized coefficients along 90% confidence bands to illustrate the relevance of different factors. Because large retailers are roughly distributed to equalize catchment areas by market size, we focus on shots delivered per capita.

A Spatial Model of Covid Vaccine Reluctance

The analysis (figure 1) indicated that Trump-voting counties and counties with high population levels of African Americans, as described elsewhere, were key demographic predictors for persistently low vaccination levels. These findings are consistent with other studies of vaccine hesitancy among African Americans.^{11,12,13} Online behaviors also predicted regional vaccination levels. (This has also been substantiated in academic studies.^{14,15}) Google searches for vaccine availability or vaccination locations predicted.

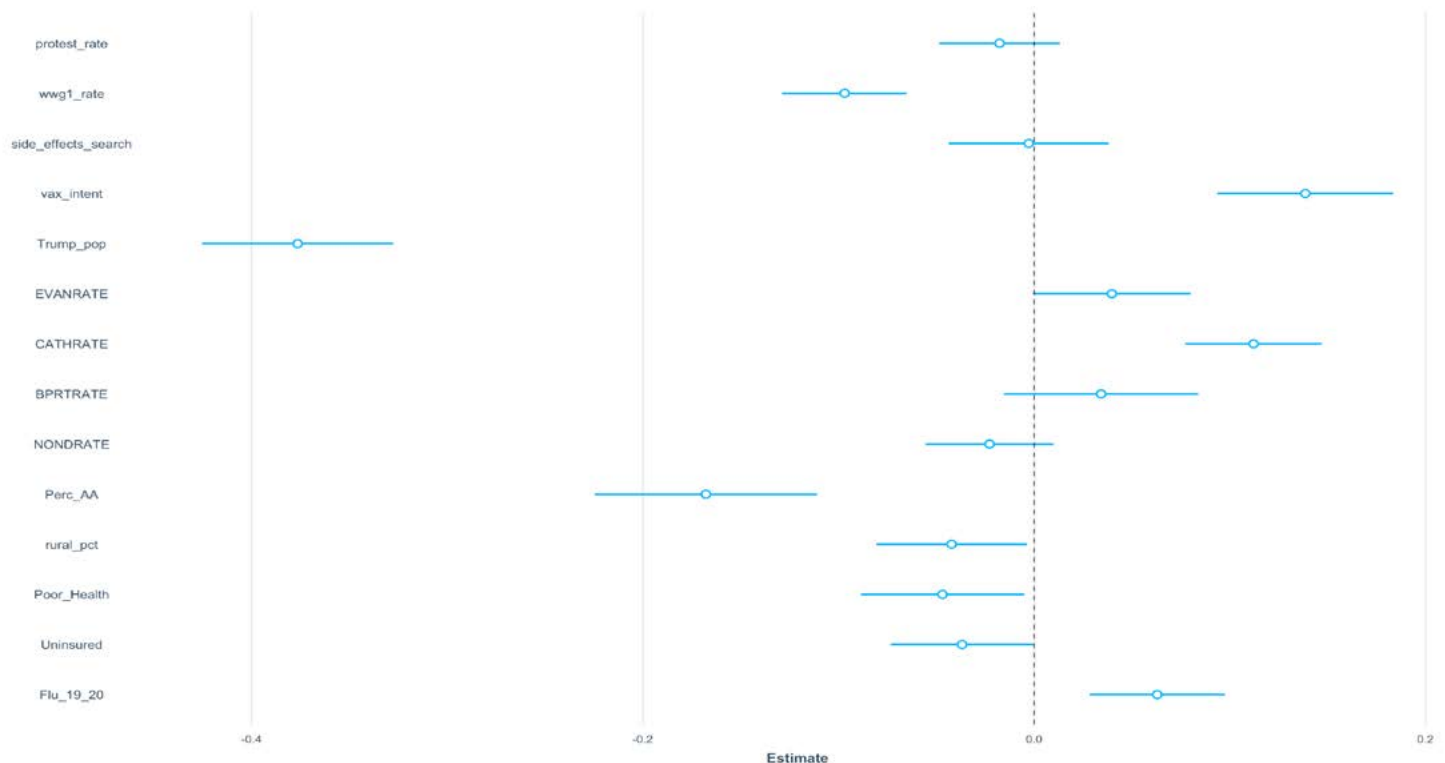


Figure 1. A dot and whisker plot of coefficients in the regression predicting CDC vaccination rates. Note: while Evangelicals appear positively correlated in this model, their bivariate correlation is -0.292

Engagement in political anti-vaccine groups and gatherings, both online and offline, was associated with lower Covid vaccination rates on the county level. We matched geo-spatial signifiers from Twitter with keywords, and though the sample of small (n=43) we found that use of the “WWG1WGA” meme (where we go one we go all, code for subscribing to the QAnon conspiracy narrative) was a statistical indicator for regions with lower Covid vaccine levels. As one might expect, in addition to engagement with online conspiracy theories, engagement in real-world anti-vaccine protests and mobilizations predicted (non significantly) lower vaccination levels. Finally, our research suggested that Covid-19 vaccination levels could be regionally predicted by flu vaccine, and reluctance for these vaccines is correlated.



Figure 2. (L) A topic network from a recent report out of the Network Contagion Labs¹⁷ using over 3 million comments on Twitter from QAnon related conspiracy reveals themes of global oppression, deceit by powerful elite and apocalyptic myths as some of the most commonly related content. (R) An image from a BBC report depicting a Q enthusiast protesting Covid vaccines.

Previous work from the Network Contagion Labs at Rutgers found that at the beginning of 2021, the QAnon conspiracy narrative substantially changed to focus on the Covid vaccine and other global public health efforts. Subsequent research indicated that QAnon-related conspiracy theories predict real world mobilizations and protests against Covid vaccines and public health restrictions related to Covid (figure 2).¹⁸ However, to our knowledge, this current analysis is the first evidence that the use of QAnon memes on social media predicts Covid vaccine reluctance.

Using the variables above, a geographically weighted regression analysis that analyzes each county’s data in relationship with neighboring counties offers a post-hoc forecast of the persistence of low vaccine levels at the county level, and proved strongly predictive of Covid vaccine reluctance (R2= 0.71, Figure 3). In addition, in the online version of this report we include an [interactive, topographical map](#) of many of the variables analyzed here, layered beneath the forecast values. This allows users to explore county-specific underlying factors that might predict low Covid vaccination levels.

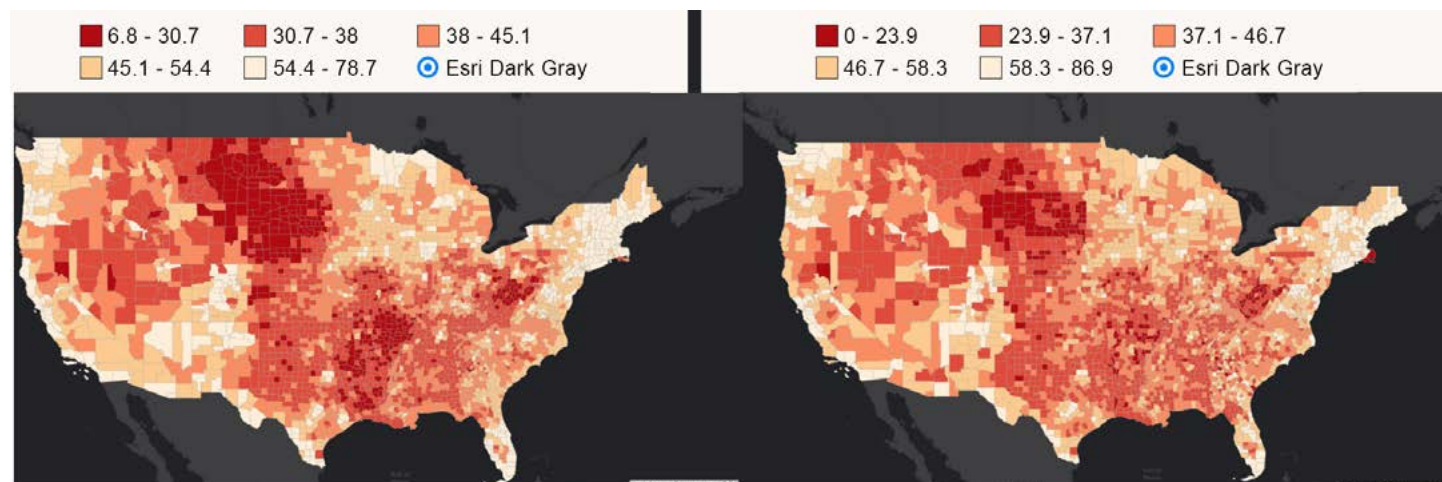


Fig 3. A geographically weighted regression model (left) of vaccination levels produces predictions that closely match actual vaccination levels (right).

A Spatial Model of Walmart as a Covid Vaccine Distribution Network

We next sought to examine and model Walmart’s performance as a Covid vaccine distribution network to better understand how underlying components of the broader distribution effort fare when compared to the overall average. Using data provided by Walmart on its county-level Covid vaccine distribution, we ran a linear model (R2= 0.17) to determine what factors predict Walmart’s performance in distributing the Covid vaccine. Our analysis (figure 4) concludes that Walmart is a strong performer in vaccine distribution. Its performance in key regions suggests it may serve as a critical player for reaching reluctant populations.

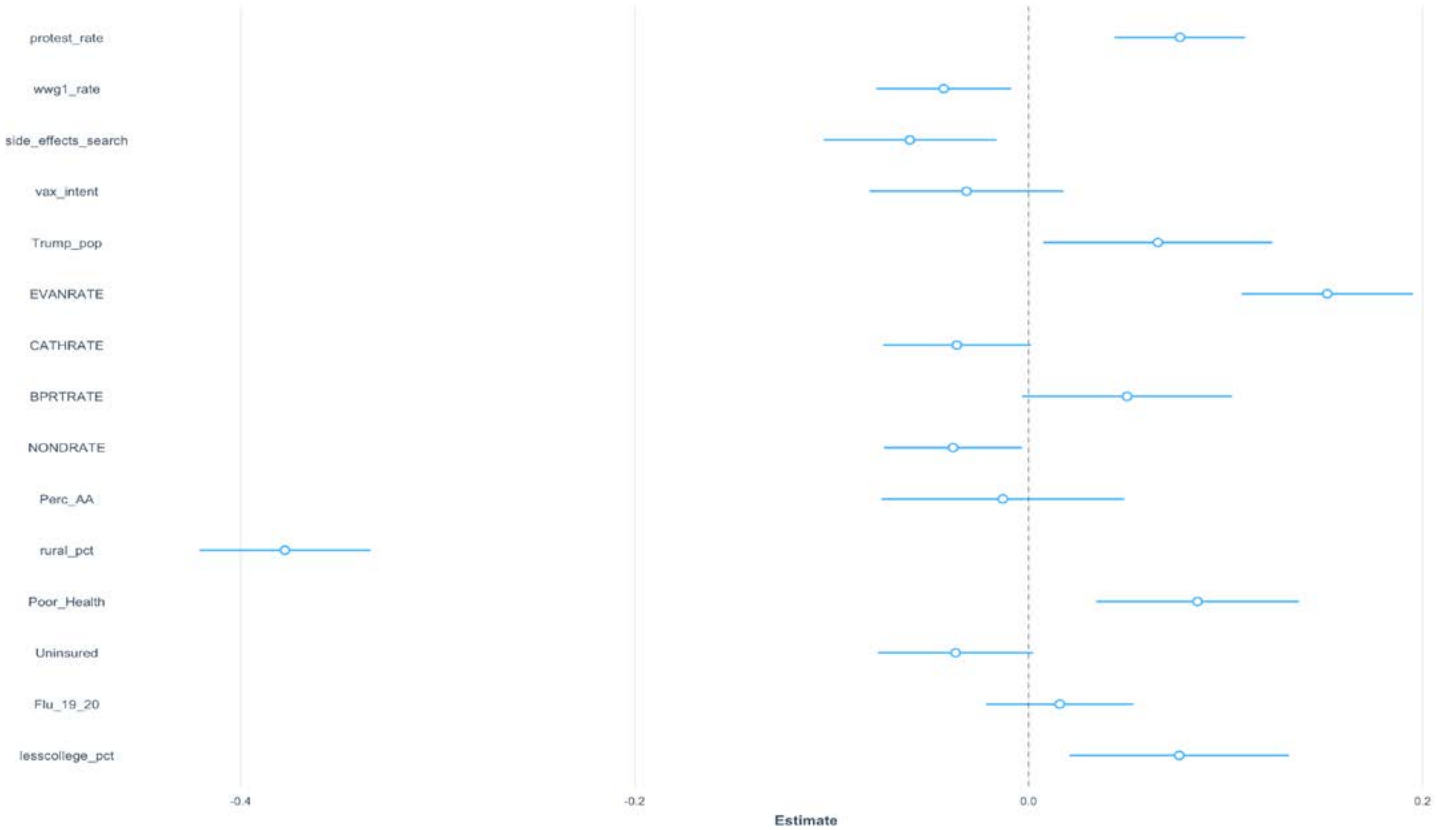


Figure 4. A dot and whisker plot of coefficients in a spatial regression for vaccination levels for Walmart.

We compared Walmart’s performance to the aggregate, and after characterizing sources of error in the model,¹⁹ we found specific areas of strength for Walmart as a Covid vaccine distribution network. Where distribution is lagging overall, Walmart outperforms the aggregate. We find that counties with high proportions of generally reluctant populations, such as Trump-voters and African Americans, do not appear to be significant challenges for Walmart. In fact, counties with a high proportion of Trump supporters and counties with a high proportion of evangelicals see stronger Walmart performance (figure 5). In addition, Walmart is successful at reaching important at-risk populations such as those reporting poor health. We found areas that were not as strong, including rural communities.

Our models expose key areas of need and may expose possible areas of future calamity, such as South Dakota, which early GWR (figure 3) predictions emphasized and which is currently leading the nation in COVID cases.²⁰

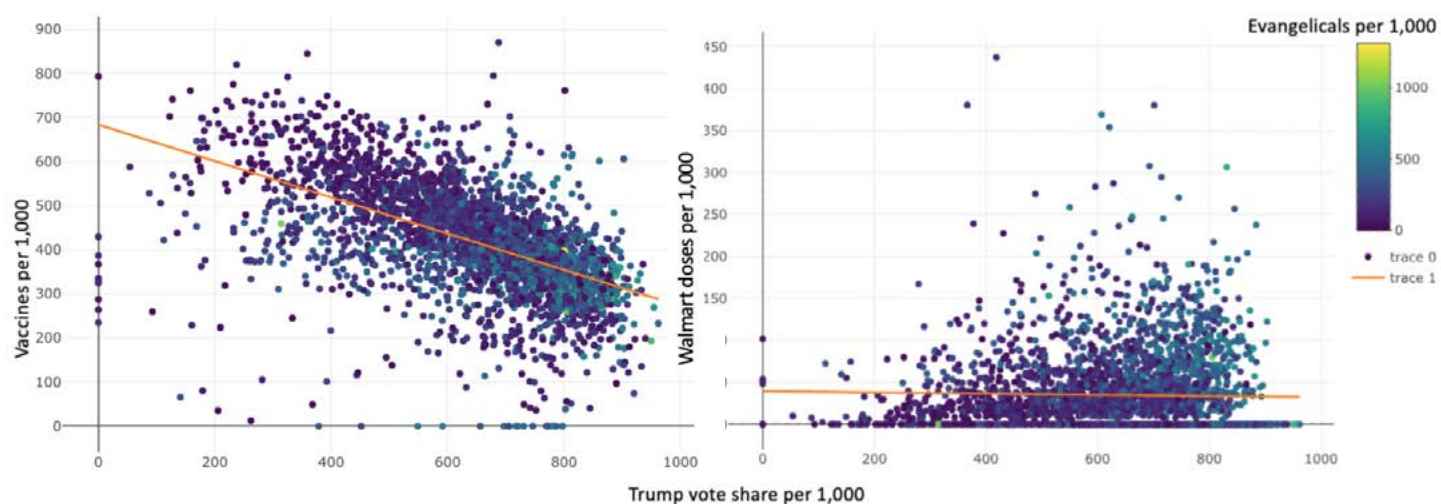


Figure 5. Scatter plots of doses of vaccine (Y-Axis) vs Trump voter share (X-Axis) shows that while general vaccine distribution (L) decreases with Trump voter share by county, Walmart’s distribution (R) does not. Evangelical counties are shown in yellow scale.

CONCLUSIONS

Making private Covid vaccine distribution data available in a safe, considered, and public facing way, helps the research community rapidly gather important insights into problems that now impact the entire globe. When private distributors make such data available, it permits researchers to evaluate their performance. With access to enough data, we can determine which key distributors have the critical reach needed to address highly hesitant regions and populations.

Evaluating the performance of key distributors is not possible using only limited, aggregate data on Covid vaccine distribution from the Centers for Disease Control and Prevention. Yet, determining the relative makeup and reach of specific distributors is a crucial criteria for consideration for public health officials, law enforcement, policy makers and the broader medical and health-research community.

In addition to modeling the performance of a specific Covid vaccine distribution network, this report supported a large-scale spatial analysis of Covid vaccine reluctance and its underlying predictors more broadly. Among the notable findings in these analyses, the influence of online behaviors including Google searches for any kind of vaccine and concerns about their side effects proved to be indicators of vaccine reluctance regionally. This finding illustrates the importance of data sharing in crisis. Our models suggest Google’s data proves a valuable tool for predicting real world Covid vaccine rates, and without Walmart’s proprietary data, we would not have been able to do many of these analyses at all.

In addition to online search behavior, this report showed a key role for political misinformation, disinformation, and mobilization in predicting Covid vaccine reluctance. The spatial distribution of tweets for the conspiracy code word “WWG1WGA,” QAnon’s key slogan, predicted a lower Covid vaccination rate on the county level. The spatial distribution of anti-vaccine rallies and mobilizations, similarly, predicted reluctance. Taken together, these results suggest that online community signifiers for political belonging and real world political unrest both serve as important spatial indicators of Covid vaccine reluctance.

The Covid-19 pandemic is entering a dangerous new phase described by unpredictable variants and rapid spread. At the same time, our ability to successfully vaccinate against the disease is predicted by a simultaneous and rapidly mutating infodemic of political and institutional distrust, and misinformation and disinformation, much of which is disseminated through conspiracy theories.

In the face of these threats to public health, a new paradigm of mutual data sharing, transparency, and civic participation is emerging and is critical to helping contain contagious threats.

Private corporations possess essential data for solving global challenges. Mainstream brands and institutions that entrust their data to the public good can save lives, and in turn, they can gain the public’s trust in a virtuous cycle; helping to turn the tide against both a deadly pandemic and a dangerous infodemic.

FOOTNOTES

- ¹ Biswas, Monirujjaman. "Exploring district-level geographical variance of complete immunization coverage in India." *European Journal of Public Health* 30.Supplement_5 (2020): ckaa166-956
- ² Inks Amir Masoud, and Rina Ghose. "Geospatial analysis of misinformation in COVID-19 related tweets." *Applied Geography* 133 (2021): 102473.
- ³ https://github.com/tonmcg/US_County_Level_Election_Results_08-20
- ⁴ <https://www.census.gov/programs-surveys/acs>
- ⁵ <https://www.census.gov/programs-surveys/sahie.html>
- ^{6A} <https://www.cdc.gov/brfss/index.html>
- ^{6B} <https://www.cdc.gov/flu/fluview/interactive-general-population.htm>
- ⁷ <https://google-research.github.io/vaccination-search-insights/?placelId=ChIJCzYy5IS16lQRQrfeQ5K50xw>
- ⁸ Ibid
- ⁹ <https://acleddata.com/special-projects/us-crisis-monitor/>
- ¹⁰ <http://www.usreligioncensus.org/>
- ¹¹ Khubchandani, Jagdish, and Yilda Macias. "COVID-19 Vaccination Hesitancy in Hispanics and African-Americans: A Review and Recommendations for Practice." *Brain, behavior, & immunity-health* (2021): 100277.
- ¹² Razai, Mohammad S., Tasnime Osama, Douglas GJ McKechnie, and Azeem Majeed. "Covid-19 vaccine hesitancy among ethnic minority groups." (2021).
- ¹³ Willis, Don E., Jennifer A. Andersen, Keneshia Bryant Moore, James P. Selig, Christopher R. Long, Holly C. Felix, Geoffrey M. Curran, and Pearl A. McElfish. "COVID-19 vaccine hesitancy: Race/ethnicity, trust, and fear." *Clinical and Translational Science* (2021).
- ¹⁴ Chadwick, Andrew, Johannes Kaiser, Cristian Vaccari, Daniel Freeman, Sinéad Lambe, Bao S. Loe, Samantha Vanderslott et al. "Online social endorsement and Covid-19 vaccine hesitancy in the United Kingdom." *Social Media+ Society* 7, no. 2 (2021): 20563051211008817.
- ¹⁵ Getman, Rebekah, Mohammad Helmi, Hal Roberts, Alfa Yansane, David Cutler, and Brittany Seymour. "Vaccine hesitancy and online information: the influence of digital networks." *Health Education & Behavior* 45, no. 4 (2018): 599-606.
- ¹⁶ <https://www.bbc.com/news/blogs-trending-53997203>
- ¹⁷ Stahl, Lesley. "QAnon's Corrosive Impact on the U.S." *CBS New, 60 Minutes*. 02-22-2021. <https://www.cbsnews.com/news/qanon-conspiracy-united-states-60-minutes-2021-02-21/>. Accessed 07-21-2021.
- ¹⁸ Reid Ross, Alexander, Malav Modi, Pamela Paresky, Lee Jussim, Alex Goldenberg, Paul Goldenberg, Danit Finkelstein, John Farmer, Kelli Holden, Denver Riggelman, Jacob Shapiro, Joel Finkelstein. Network Contagion Research Institute report. March 11, 2021. "A contagion of institutional distrust: Viral Disinformation of the COVID Vaccine and the Road to Reconciliation". Available at: <https://networkcontagion.us/>
- ¹⁹ Many of the outliers in our GWR model had fairly high vaccination rates along with high percentages of Trump voters, owing to smaller, more rural populations with fewer Evangelicals but more uninsured people and reports of poor or fair health, as well as higher Google searches for vaccine side effects. A number of these counties were located in New Mexico and Virginia. Where our GWR over-predicted vaccination rates, counties tended to be less rural, with fewer uninsured, lower reports of poor health, more searches indicating vaccination intent, higher rates of Catholics and non-denominationalists. Interestingly, in both under- and over-predicted outliers, Walmart served as an important hub, with Walmart alone vaccinating an average of 3% of the population of the top 27 over-predicted communities (the highest percentage being 17% at Prince Edward County, Virginia). This indicates that Walmart has access to and can serve communities that have unexpectedly low vaccination rates.
- ²⁰ Todd, Anna, "South Dakota leads nation in new COVID-19 case increases per 100,000 residents," *Argus Leader*, Aug 19, 2021, <https://www.argusleader.com/story/news/2021/08/19/new-york-times-says-covid-case-rate-south-dakota-most-in-united-states/8204546002/>, accessed 8/22/2021

For a complete list of variables and exposition of findings, including collection and annotation, See [methods](#). See updated [methods](#).