The impact of online misinformation on U.S. COVID-19 vaccinations

Francesco Pierri[1,3], Brea Perry[2], Matthew R. DeVerna[3], Kai-Cheng Yang[3], Alessandro Flammini[3], Filippo Menczer[3] and John Bryden[3]*

*Corresponding author: jabryden@iu.edu

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

[2] Department of Sociology, Indiana University, Bloomington, Indiana, United States

[3] Observatory on Social Media, Indiana University, Bloomington, Indiana, United States

Widespread uptake of COVID-19 vaccines is necessary to achieve herd immunity[1–3]. However, surveys have found concerning numbers of U.S. adults hesitant or unwilling to be vaccinated[4,5]. Online misinformation may play an important role in vaccine hesitancy[6–8], but we lack a clear picture of the extent to which it will impact vaccination uptake. Here, we study how vaccination rates and vaccine hesitancy are associated with levels of online misinformation about vaccines shared by 1.6 million Twitter users geolocated at the U.S. state and county levels. We find a negative relationship between misinformation and vaccination uptake rates. Online misinformation is also correlated with vaccine hesitancy rates taken from survey data. Associations between vaccine outcomes and misinformation remain significant when accounting for political as well as demographic and socioeconomic factors. While vaccine hesitancy is strongly associated with Republican vote share, we observe that the effect of online misinformation on hesitancy is strongest across Democratic rather than Republican counties. These results suggest that addressing online misinformation must be a key component of interventions aimed to maximize the effectiveness of vaccination campaigns.

Introduction

The COVID-19 pandemic has killed over 3 million people and infected 140 million worldwide as of April 2021[9]. Vaccination is the lynchpin of the global strategy to fight the SARS-CoV-2 coronavirus[10,11]. However, to be effective, the uptake of vaccines must be widespread. Surveys conducted during February and March 2021 find that around 40-47% of American adults are hesitant to take the COVID-19 vaccine[4,5], suggesting the population may fall short of achieving the vaccination rate required to achieve herd immunity (i.e., 60-70%)[1–3]. Further, uneven distribution of vaccination raises the possibility of localised clusters of non-vaccinated people[12] that will preclude eradication of the virus and may exacerbate racial, ethnic, and socioeconomic health disparities.

Vaccine hesitancy covers a spectrum of intentions, from delaying vaccination to outright refusal to be vaccinated[13]. Several factors have been linked to COVID-19 vaccine hesitancy, with rates in the U.S. highest among three groups: African Americans, women, and conservatives[14]. Other predictors, including education, employment, and income are also associated with hesitancy[15]. Targeted messaging can be used to build confidence and address complacency in target groups[13], but these strategies are undermined by exposure to misinformation.

A number of studies discuss the spread of vaccine misinformation on social media[16] and argue that such campaigns have driven negative opinions about vaccines and even contributed to the resurgence of measles[17,18]. In the COVID-19 pandemic scenario, widely shared misinformation includes false claims that vaccines genetically manipulate the population or contain microchips that interact with 5G networks[7,19]. Exposure to online misinformation has been linked to increased health risks[20] and vaccine hesitancy[8]. Gaps remain in our understanding of how vaccine misinformation is linked to broad-scale patterns of COVID-19 vaccine uptake rates.

The Pfizer-BioNTec COVID-19 vaccine was the first to be given U.S. Food and Drug Administration approval on December 10th 2020[21]. Since then, two other vaccines have been approved in the U.S. Until recently, vaccines have been selectively administered with nationwide priority being given to more vulnerable cohorts such as the more elderly members of the population. As vaccines become available to the entire adult population[22], adoption will be driven by limits in demand rather than in supply. It is therefore important to study the variability in uptake across U.S. states and counties, as reflected in recent surveys[23,24].

In this work we study relationships between vaccine uptake, vaccine hesitancy and online misinformation. We measure vaccine uptake from the daily vaccination rates recorded by the CDC[25] for each U.S. state averaged over the week of March 19 to 25, 2021, when variability across U.S. states became apparent[22]. Vaccine hesitancy is likely to affect uptake rates, so we specify a longer time window to measure that variable, Jan 4th to March 25th, 2021, and likewise for online misinformation. We leverage over 22 M individual responses to surveys administered on Facebook to assess vaccine hesitancy rates[24], and we identify online misinformation by focusing on low-credibility sources shared on Twitter[26–29] by over 1.67M users geolocated within U.S. regions (see Methods for further details on the methodology). For statistical analysis, we use multivariate regression models adjusting for socioeconomic, demographic and political confounding factors. The variables are recorded at group level, which makes drawing inference at the individual level problematic; however, we account for likely issues using interaction variables, logarithmic transforms, heteroskedasticity tests, clustering at multiple levels (county and state), and uncertainty weighting of variables.

Results

Looking across U.S. states, we observe a negative association between vaccination uptake rates and online misinformation (Pearson $R = -0.49$, $p < .001$). Investigating covariates known to be associated with vaccine uptake or hesitancy, we find that an increase in the mean amount of online misinformation is significantly associated with a decrease in daily vaccination rates per million ($b = -3518.00$, $p < .01$, Fig.1**a**, and see Methods and Table S1 in Supplementary Information). Political partisanship (a 10% increase in GOP vote) is also strongly associated with vaccination rate ($b = -640.32$, $p < .01$). These two factors alone explain nearly half the variation in state-level vaccination rates, and are themselves moderately correlated (Fig. S1 and Table S1 in the Supplementary Information), consistent with prior research[30]. Remaining covariates, including religiosity, unemployment rate, and population density, are non-significant and/or collinear with other variables and thus dropped for parsimony.

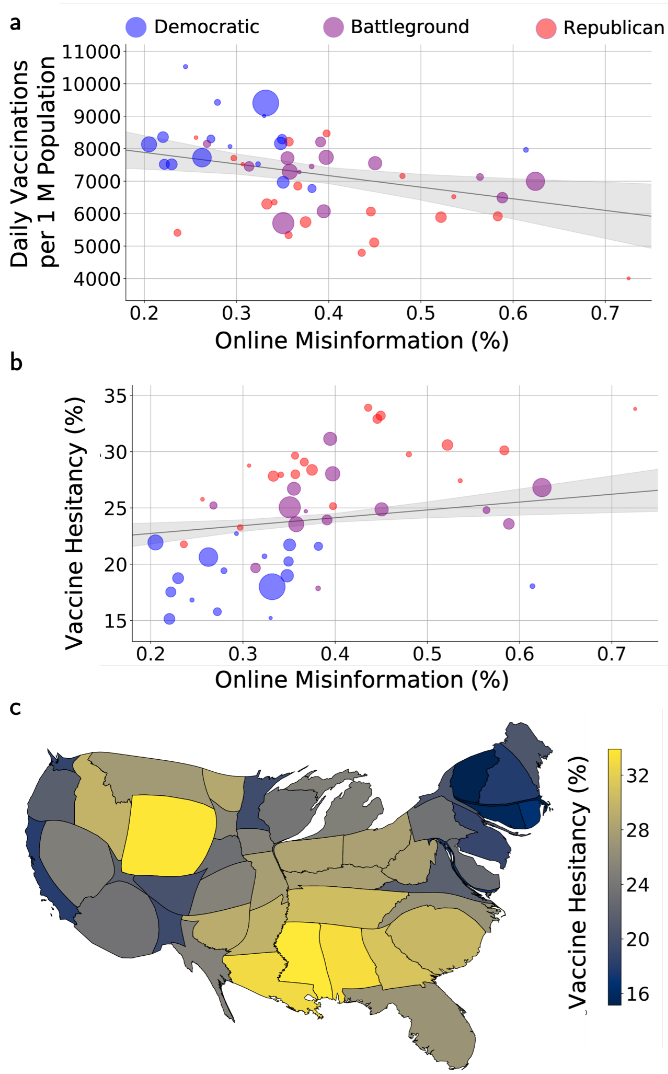

Figure 1. Online misinformation is associated with vaccination uptake and hesitancy at the state level. (a) State-level mean daily vaccinations per million population during the period from March 19 to 25, 2021, against the average proportion of vaccine misinformation tweets shared by geolocated users on Twitter during the period from Jan 4th to March 25th, 2021. **(b)** Levels of state-wide vaccine hesitancy, computed as the fraction of individuals who would not get vaccinated according to Facebook daily surveys administered in the period from January 4th to March 25th, 2021, and misinformation about vaccines shared on Twitter. Each dot represents a U.S. state and is colored according to the share of Republican voters (battleground states have a share between 45% and 55%) and sized according to population. Grey lines show the partial correlation between the two variables after adjusting for socioeconomic, demographic, and political factors in a weighted multiple linear regression model (shaded areas correspond to 95% C.I.). **(c)** Cartogram[31] of the U.S. in which the area of each state is proportional to the average number of misinformation links shared by geolocated users, and the color is mapped to the vaccine hesitancy rate, with lighter colors corresponding to higher hesitancy.

To investigate vaccine hesitancy, we leverage over 22 M individual responses to daily survey data provided by Facebook[24] (see Methods). Reports of vaccine hesitancy are aggregated at the state level (i.e., percent hesitant) and weighted by sample size. We find a strong negative correlation between vaccine uptake and hesitancy across U.S. states (Pearson R = –0.71, $p < .001$, Fig. S1 in Supplementary Information), suggesting that daily vaccination rates largely reflect demand for vaccines rather than supply. Taking into account the same set of potential confounding factors in a weighted regression model, we find a significant positive association between misinformation ($b = 6.88$, $p < .01$) and state-level vaccine hesitancy, and between political partisanship ($b = 2.96$, $p < .001$) and hesitancy (see Fig. 1**b** and Fig. S1 in Supplementary Information). Fig. 1**c** provides an illustration of the correlation between misinformation and hesitancy. For example, the large size and yellow color of Wyoming indicate it is the state with the highest level of misinformation and hesitancy. Among other variables, we find that the percent of Black residents is positively related to reports of hesitancy ($b = 0.12$, $p < .01$), while percent Hispanic or Latinx is negatively associated ($b = –0.07$, $p < .05$). The percent of residents below the poverty line is also positively associated with vaccine hesitancy ($b = 0.53$, $p < .01$).

To test the robustness of these results, we also consider a more granular level of information by examining county data. Similar to previous analyses, we compute online misinformation shared by almost 1.15 M Twitter users geolocated in over 1,300 U.S. counties. We measure vaccine hesitancy rates by leveraging over 17 M daily responses to the Facebook survey for over 700 distinct counties. The total number of observations (i.e. counties) for which we are able to measure both variables is N=548 (see Methods). Political partisanship and misinformation are both significantly correlated with county-level vaccine hesitancy, net covariates (Table S4, Fig. S2 in Supplementary

Information). Using a weighted multiple linear regression model, we find a significant interaction between political partisanship and misinformation. Specifically, as levels of misinformation increase, democratic and republican counties converge to the same level of vaccine hesitancy (Fig. 2).

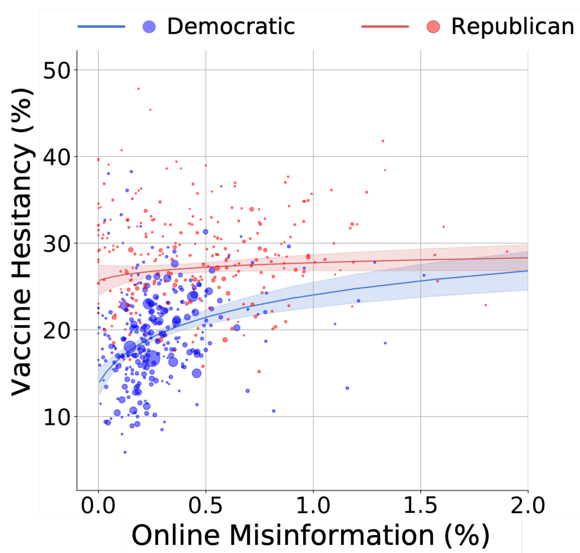

Figure 2. Associations of online misinformation and political partisanship with vaccination hesitancy at the U.S. county level. Each dot represents a U.S. county, with size and color indicating population size and political majority, respectively. The average proportion of misinformation shared on Twitter by geolocated users was fitted on a log scale due to non-normality (i.e., positive skew) at the county level. The two lines show predicted values of vaccine hesitancy as a function of misinformation for majority Democratic and Republican counties, adjusting for county-level confounding factors (see Methods). Shaded area corresponds to 95% C.I.

Discussion

Our results provide evidence for the problem of geographical regions with lower levels of COVID-19 vaccine uptake, which may be driven by online misinformation. Considering variability across regions with low and high levels of misinformation, the best estimates from our data predict a ~20% decrease in vaccine uptake between states, and a ~67% increase in hesitancy rates across democratic counties, across the full range of misinformation prevalence. At these levels of impact on vaccine uptake, the data predict SARS-CoV-2 will remain endemic in many U.S. regions. While our data cannot demonstrate the directionality nor the causal relationship between misinformation and vaccine refusal, we find that both vaccine hesitancy and uptake are associated with misinformation. Vaccine-hesitant individuals are potentially more likely to post more vaccine misinformation, but previous studies have shown a directional effect with exposure to misinformation leading to vaccine hesitancy[8]. Our results thus provide concerning evidence of a damaging effect of online misinformation on the ongoing U.S. COVID-19 vaccination program.

Public opinion is very sensitive to the information ecosystem and sensational posts tend to spread widely and quickly. Our results indicate that there is a geographical component to this spread, with misinformation spreading at a local scale. While social media users are not representative of the general public, existing evidence suggests that vaccine hesitancy flows across online social networks[33], providing a mechanism for the lateral spread of misinformation offline among those connected directly or indirectly to misinformation spreading online. More broadly, our results provide additional insight into the effects of information diffusion on human behavior and the spread of infectious diseases[34].

A limitation of our findings is that they are based on data averaged over geographical regions, which does not provide evidence at an individual level. However, to account for group-level effects we present a number of sensitivity analyses, and note that our findings are consistent over two geographical scales. Our results are also limited to a snapshot in time. Vaccination hesitancy levels will potentially change over time due to novel factors, including changes in COVID-19 infection and death rates, as well as legitimate reports about vaccine safety, among other factors[32].

Associations between online misinformation and detrimental offline effects, like the results presented here, call for better moderation of our information ecosystem. COVID-19 misinformation is shared overtly by known entities on major social media platforms[35]. While people have a constitutional right to free speech, it is important to maintain an environment where individuals have access to good information that benefits public health.

Methods

Our key independent variable is the mean percentage of vaccine-related misinformation shared via Twitter at the U.S. state or county level. We used 55 M tweets from the CoVaxxy dataset[19], which were collected between Jan 4th and March 25th using the Twitter filtered stream API using a comprehensive list of keywords related to vaccines (see S.I.). We leveraged the *carmen* library[36] to geolocate almost 1.67 M users residing in 50 U.S. states, and a subset of approximately 1.15 M users residing in over 1,300 counties. The larger set of users accounts for a total of 11 M shared tweets. Following a consolidated approach in the literature[26–29], we identified misinformation by considering tweets that contained links to news articles from a list of low-credibility websites compiled by a politically neutral third-party (see details in S.I.). We measured the prevalence of misinformation about vaccines in each region by (i) calculating the proportion of vaccine-related misinformation tweets shared by each geo-located account; and (ii) taking the average of this proportion across accounts within a specific region. The Twitter data collection was evaluated and deemed exempt from review by the Indiana University IRB (protocol 1102004860).

Our dependent variables include vaccination uptake rates at the state level and vaccine hesitancy at the state and county levels. Vaccination uptake is measured from the number of daily vaccinations administered in each state during the week 19th-25th March 2021, and measurements are derived from the Centers for Disease Control and Prevention[25]. Vaccine hesitancy rates are based on Facebook Symptom Surveys provided by the Delphi Group[24] at Carnegie Mellon University in the period Jan 4th-March 25th 2021. We computed hesitancy by taking the complementary proportion of individuals "who either have already received a COVID vaccine or would definitely or probably choose to get vaccinated, if a vaccine were offered to them today." See Supplementary Information for further details.

There are no missing vaccine-hesitancy survey data at the state level. Data are missing at the county level because Facebook survey data are available only when the number of respondents is at least 100. We use the same threshold on the minimum number of Twitter accounts geolocated in each county, resulting in a sample size of $N = 548$ counties.

Our multivariate regression models adjust for six potential confounding factors. These include percent of the population below the poverty line, percent aged 65+, percent of residents in each racial and ethnic group (Asian, Black, Native American, and Hispanic; White non-Hispanic is omitted), rural-urban continuum code (RUCC, county level only), number of COVID-19 deaths per thousand, and percent republican vote (in 10 percent units). Other covariates (listed in Supplementary Information table S9) were considered but dropped due to non-significance and/or multicollinearity (i.e., high variance inflation factors).

We also conduct a large number of sensitivity analyses, including different specifications of the misinformation variable (with a restricted set of keywords and different thresholds for the inclusion of Twitter accounts) as well as logged versions of misinformation (to correct positive skew). These results are presented in Supplementary Information (Tables S3-S8).

We conduct multiple regression models predicting vaccination rate and vaccine hesitancy. Both dependent variables are normally distributed, making weighted least squares regression the appropriate model. Data are observed (aggregated) at the state or county level rather than at the individual level. Analytic weights are applied to give more influence to observations calculated over larger samples. The weights are inversely proportional to the variance of an observation such that the variance of the *j*-th observation is assumed to be σ^2/w_j where w_j is the weight. The weights are set equal to the size of the sample from which the average is calculated. We estimate weighted regression with the *aweights* command in Stata 16. In addition, because counties are nested hierarchically in states, we use cluster robust standard errors to correct for lack of independence between county-level observations**.**

Acknowledgments

We are grateful to John Bollenbacher, Bao Tran Truong, David Axelrod, Niklas Loynes, and

Christopher Torres-Lugo for their contributions to the CoVaxxy data collection and for helpful discussion; to George Turner for help with the XSEDE computer platform; and to JangDong Seo, Edwin van Leeuwen, and Francesco Scotti for insightful discussions on statistical analyses. This work is supported in part by Knight Foundation, Craig Newmark Philanthropies, the National Science Foundation (grant ACI-1548562), and EU H2020 (grant 101016233).

Authors' contributions

J.B., F.P., and F.M. designed the study. F.P., M.D., K.Y., and J.B. collected and processed the data. B.P., F.P., and J.B. performed the analyses. J.B., F.P., and B.P. wrote the manuscript with contributions from all authors. J.B., A.F., and F.M. directed the project.

Competing interests

The authors declare no competing interests.

Data and code availability

All measurements of vaccine uptake and vaccine hesitancy rates as well as socioeconomic, political, and demographic variables at the state and county level are publicly available in the online repository associated with this paper[37]. We also provide aggregated measurements of online misinformation shared by geolocated Twitter users. Results at the state and county level can be fully reproduced using the STATA scripts provided in the repository. Due to Twitter's terms of use and service, we can only release IDs of the tweets present in our dataset, which can be reconstructed using the Twitter API. The IDs are accessible in the public dataset[19] associated with the CoVaxxy project (osome.iu.edu/tools/covaxxy) from the Observatory on Social Media at Indiana University.

References

1. Aguas, R. *et al.* Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. *medRxiv* 2020.07.23.20160762 (2020).
2. Gomes, M. G. M. *et al.* Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv* 2020.04.27.20081893 (2020) doi:10.1101/2020.04.27.20081893.
3. Lu, D. *et al.* Data-driven estimate of SARS-CoV-2 herd immunity threshold in populations with individual contact pattern variations. *medRxiv* 2021.03.19.21253974 (2021) doi:10.1101/2021.03.19.21253974.
4. KFF COVID-19 Vaccine Monitor Dashboard. *KFF* https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/ (2021).
5. Funk, C. & Tyson, A. Growing Share of Americans Say They Plan To Get a COVID-19 Vaccine – or Already Have. *Pew Research Center Science & Society* https://www.pewresearch.org/science/2021/03/05/growing-share-of-americans-say-they-plan-to-get-a-covid-19-vaccine-or-already-have/ (2021).
6. DeVerna, M. R. *et al.* CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines. *In Proc. Intl. AAAI Conf. on Web and Social Media (ICWSM), 2021.* Preprint arXiv: 2101.07694. (2021).
7. Hotez, P. *et al.* Correcting COVID-19 vaccine misinformation: Lancet Commission on COVID-19 Vaccines and Therapeutics Task Force Members. *EClinicalMedicine* **33**, (2021).

8. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K. & Larson, H. J. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* 1–12 (2021).

9. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int (2021).

10. Orenstein, W. A. & Ahmed, R. Simply put: Vaccination saves lives. *Proc. Natl. Acad. Sci.* **114**, 4031–4033 (2017).

11. Kim, J. H., Marks, F. & Clemens, J. D. Looking beyond COVID-19 vaccine phase 3 trials. *Nat. Med.* **27**, 205–211 (2021).

12. Salathé, M. & Bonhoeffer, S. The effect of opinion clustering on disease outbreaks. *J. R. Soc. Interface* **5**, 1505–1508 (2008).

13. MacDonald, N. E. Vaccine hesitancy: Definition, scope and determinants. *Vaccine* **33**, 4161–4164 (2015).

14. Callaghan, T. *et al.* Correlates and disparities of intention to vaccinate against COVID-19. *Soc. Sci. Med. 1982* **272**, 113638 (2021).

15. Khubchandani, J. *et al.* COVID-19 Vaccination Hesitancy in the United States: A Rapid National Assessment. *J. Community Health* **46**, 270–277 (2021).

16. Broniatowski, D. A. *et al.* Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *Am. J. Public Health* **108**, 1378–1384 (2018).

17. Burki, T. Vaccine misinformation and social media. *Lancet Digit. Health* **1**, e258–e259 (2019).

18. Wilson, S. L. & Wiysonge, C. Social media and vaccine hesitancy. *BMJ Glob. Health* **5**, e004206 (2020).

19. DeVerna, M. R. *et al.* Data for CoVaxxy: A collection of English-language Twitter posts

about COVID-19 vaccines. https://github.com/osome-iu/CoVaxxy (2021).

20. Gallotti, R., Valle, F., Castaldo, N., Sacco, P. & De Domenico, M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nat. Hum. Behav.* (2020) doi:10.1038/s41562-020-00994-6.

21. U.S. Food & Drug Administration. Pfizer-BioNTech COVID-19 Vaccine. *FDA* (2021).

22. Crist, C. States Begin Opening COVID-19 Vaccines to All Adults. *WebMD* https://www.webmd.com/vaccines/covid-19-vaccine/news/20210324/states-begin-opening-covid-19-vaccines-to-all-adults (2021).

23. Lazer, D. *et al.* The COVID States Project #43: COVID-19 vaccine rates and attitudes among Americans. (2021) doi:10.31219/osf.io/rnw8z.

24. Farrow, D. C., Brooks, L. C., Rumack, A., Tibshirani, R. J. & Rosenfeld, R. Delphi Epidata API. https://github.com/cmu-delphi/delphi-epidata (2015).

25. Centers for Disease Control and Prevention. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker (2020).

26. Lazer, D. M. J. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).

27. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374–378 (2019).

28. Bovet, A. & Makse, H. A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **10**, 7 (2019).

29. Shao, C. *et al.* The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 4787 (2018).

30. Nikolov, D., Flammini, A. & Menczer, F. Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. *Harv. Kennedy Sch. Misinformation Rev.* **1(7)**,

(2021).

31. Gastner, M. T., Seguy, V. & More, P. Fast flow-based algorithm for creating density-equalizing map projections. *Proc. Natl. Acad. Sci.* **115**, E2156–E2164 (2018).

32. Larson, H. J. & Broniatowski, D. A. Volatility of vaccine confidence. *Science* **371**, 1289–1289 (2021).

33. Brunson, E. K. The Impact of Social Networks on Parents' Vaccination Decisions. *Pediatrics* **131**, e1397–e1404 (2013).

34. Funk, S., Salathé, M. & Jansen, V. A. A. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7**, 1247–1256 (2010).

35. Yang, K.-C. *et al.* The COVID-19 Infodemic: Twitter versus Facebook. *ArXiv201209353 Cs* (2021).

36. Dredze, M., Paul, M. J., Bergsma, S. & Tran, H. *Carmen: A Twitter Geolocation System with Applications to Public Health*.

37. Pierri, F. *et al.* Reproducibility code for 'The impact of online misinformation on U.S. COVID-19 vaccinations'. https://github.com/osome-iu/CoVaxxy-Misinfo (2021).

The impact of online misinformation on U.S. COVID-19 vaccinations

Francesco Pierri, Brea Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer and John Bryden

Supplementary Information

Data collection and sources

Twitter data

In our CoVaxxy[1] project, we collected around 55 M English-language posts about vaccines on Twitter by means of the Twitter *POST statuses/filter v1.1 API*, in the period from January 4th, 2021 to March 25th, 2021. Data collection and analysis was done using the Extreme Science and Engineering Discovery Environment (XSEDE)[2].

To define as complete a set as possible of English language keywords related to vaccines, we employed a snowball sampling methodology in December 2020[1] (see reference for full details on the data collection pipeline). The final list contains almost 80 keywords, and it is accessible in the online repository associated with the reference[3]. As a robustness test, we further perform sensitivity analyses using a restricted set of keywords (“vaccine”, “vaccinate”, “vaccination”, “vax”) which covers almost 95% of the total number of geolocated tweets. Results are equivalent to those presented in the main text and are described in the section “Sensitivity Analyses”.

To match Twitter posts with US states and counties, we first identified a collection of Twitter accounts that disclosed a location in their Twitter profile. We then employed the *carmen* Python library[4] to match each location to US states and counties. We were able to match around 1.67 *M* users to 50 US states, and a subset of 1.15 *M* users to over 1,300 US counties; the larger set accounts for a total number of almost 11 *M* shared tweets.

To analyze the spread of low-credibility information, we identified all URLs shared in Twitter posts that originated from a list of low-credibility sources, following a large corpus of literature[5–9]. We employ the Iffy+ Misinfo/Disinfo list of low-credibility sources[10], which is based on information provided by the Media Bias/Fact Check website (MBFC, https://mediabiasfactcheck.com), an independent organization that reviews and rates the reliability of news sources. As defined in the related methodology, political leaning is not a factor for inclusion. The list includes sites labeled by MBFC as having a “Very Low” or “Low” factual-reporting level as well as those classified as “Questionable” or “Conspiracy-Pseudoscience”. The list also includes fake-news websites flagged by BuzzFeed, FactCheck.org, PolitiFact, and Wikipedia, for a total number of 674 low-credibility sources.

Based on this list, we measure the prevalence of low-credibility information about vaccines in each region by (1) calculating the proportion of vaccine-related tweets containing URLs pointing to a low-credibility news website, for each geo-located account; and (2) taking the average of this proportion across all accounts within a specific region. We refer to this average as the state-wide (county-wide) prevalence of misinformation.

At the county level, we omit observations without vaccine hesitancy data (see next section), and we used different thresholds for the minimum number of geolocated accounts, respectively 10, 50, and 100. In the main paper, we present results when using 100 as a threshold. We provide sensitivity analyses using versions including counties with at least 10 and 50 Twitter accounts (see "Sensitivity Analyses" section). The larger threshold is likely to contain less error but also omits more counties.

Election data

We use data provided by the MIT Election Lab to extract state-level returns for the 2020 US presidential election[11]. For counties, we use data provided by Fox News, Politico, and the New York Times. They are publicly available at https://github.com/tonmcg/US_County_Level_Election_Results_08-20.

Vaccine hesitancy data

To compute vaccine hesitancy rates in each state (county), we leverage daily COVID-19 Symptom Surveys produced by the Delphi Group at Carnegie Mellon University[12]. These surveys are voluntarily answered by a random sample of users on Facebook (total reported sample size N = 22,128,855). Within the Vaccination Indicators of the survey, we extract the estimated percentage of respondents (for each state/county) "who either have already received a COVID vaccine or would definitely or probably choose to get vaccinated, if a vaccine were offered to them today." Results are available daily, for all 50 US states and for 764 US counties.

We compute state-wide (county-wide) vaccine hesitancy rates by taking the proportion of negative responses in the period from January 4th to March 25th.

Vaccine uptake data

Vaccination uptake statistics are derived from the Centers for Disease Control and Prevention (CDC) dataset (https://covid.cdc.gov/covid-data-tracker/#vaccinations). Doses monitored for each state include those administered in jurisdictional partner clinics, retail pharmacies, long-term care facilities, Federal Emergency Management Agency partner sites, Health Resources and Services Administration partner sites, and federal facilities. The data have been compiled on a daily basis by *ourworldindata.org*, and we have downloaded them for the period from January 12 to March 25, 2021. The data are available at https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations.

COVID-19 data

We extracted the number of COVID-19 cases and fatalities at the state and county level based on reports made by USAFacts (https://usafacts.org). In particular, we summed the number of daily confirmed COVID-19 cases and fatalities, referring to these as “recent”, in the period from January 4 to March 25, 2021. We then computed the cumulative number of cases and fatalities on March 25th, referring to these as “total”.

Socioeconomic data

To include socioeconomic covariates in our regression model, we use data from the Atlas of Rural and Small-Town America (available at https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/), which includes data at the state and county level from the

American Community Survey (ACS), the Bureau of Labor Statistics, and other sources. We employ data last updated on July 2, 2020, which include county population estimates and annual unemployment/employment data for 2019.

County-level measurements about religion are derived from surveys by the Association of Religion Data Archives (accessible at https://www.thearda.com/Archive/ChCounty.asp).

Additional correlation results

Figures S1 and S2 present additional results about correlations between vaccine demand, vaccine hesitancy, political partisanship, and online misinformation at state and county levels.

Main findings from regression analysis

Table S1 presents results from the weighted (Models 1 and 2) and ordinary (Models 3 and 4) least-squares regression of state-level vaccine hesitancy and vaccination rate, respectively, on covariates. As shown in Model 1, the misinformation variable and the percent of GOP voters explain nearly 80% of the variation in vaccine hesitancy at the state level. These predictors remain significant after the addition of multiple control variables (see Model 2). Misinformation and republican vote percentage explain nearly half of the variation in vaccination rate (see Model 3), and are also significantly associated with vaccination rate at the state level net of controls (see Model 4).

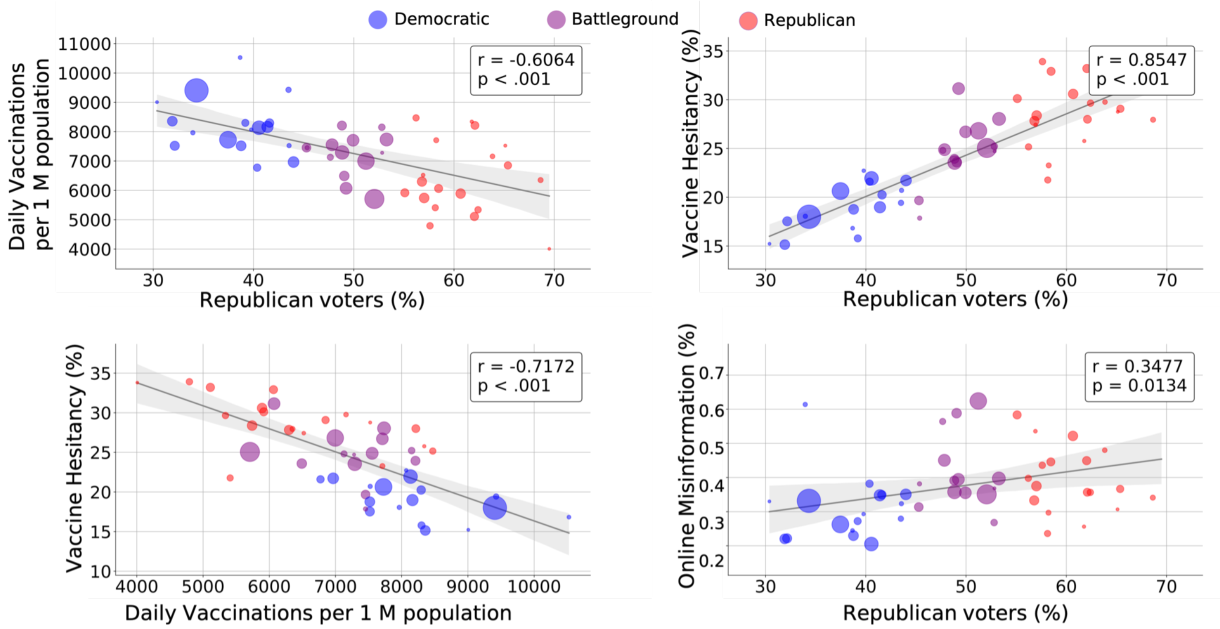

Figure S1. Correlations between vaccine demand, vaccine hesitancy, political partisanship, and online misinformation at the state level. Vaccine demand is computed as the mean number of daily vaccinations per million population in the period 19-25 March 2021. Vaccine hesitancy corresponds to the proportion of individuals who would not get vaccinated according to Facebook daily surveys administered in the period from January 4th to March 25th, 2021. Partisanship is measured as the percentage of Republican voters in the 2020 US Presidential elections. Online misinformation about vaccines shared on Twitter is measured during the period from Jan 4th to March 25th, 2021. Each dot represents a U.S. state, sized according to population and colored according to Republican vote share (battleground states have a share between 45% and 55%).

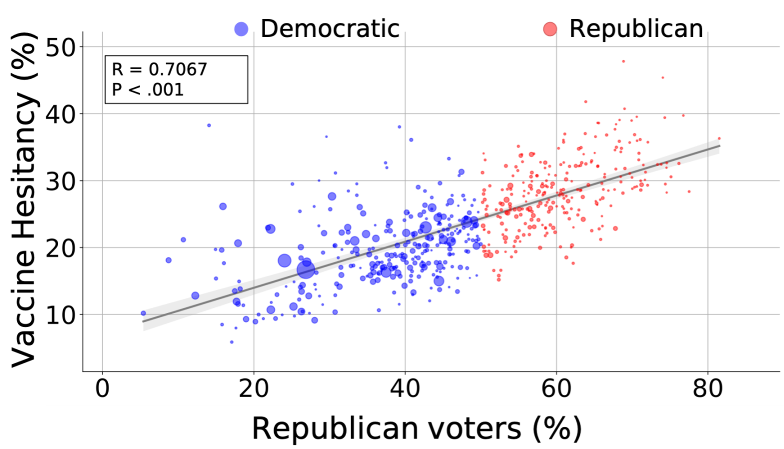

Figure S2. Political partisanship is correlated with vaccine hesitancy at the U.S. county level. Vaccine hesitancy corresponds to the proportion of individuals who would not get vaccinated according to Facebook daily surveys administered in the period from January 4th to March 25th, 2021. Partisanship is measured as the percentage of Republican voters in the 2020 US Presidential elections. Each dot represents a U.S. county, sized according to population and colored according to Republican vote share.

Sensitivity analyses

We conduct a set of sensitivity analyses to ensure that our findings are robust to alternative variable and model specifications. First, we run standard diagnostics for nonlinearity, skewness, multicollinearity, and heteroskedasticity, correcting any problems we discover. Second, because the misinformation measure at the state level is slightly positively skewed, we conduct a model using a natural logarithmic transformation of mean percent misinformation. Results from these models are consistent with the main findings (Table S2). The untransformed variable has a better model fit (lower BIC). Third, because the effect of misinformation may depend on political partisanship, we test for an interaction between misinformation and the percent of GOP voters. There is no evidence of such interaction at the state level. Fourth, we rerun the above models using versions of the mean percentage of vaccine-related misinformation shared by Twitter users by considering a restricted set of keywords to gather tweets (see previous “Twitter Data” section). As shown in Table S3, findings are consistent and robust to this alternate definition of misinformation sharing.

We also conduct a similar set of sensitivity analyses at the county level. First, we test multiple versions of the misinformation variable, which is highly skewed and zero-inflated at the county level. We use the log-transformed version for the main findings due to the best model fit, but obtain significant results with the untransformed variable and very similar findings with a polynomial model that also captures the nonlinear relationship between misinformation and vaccine hesitancy. Second, we test for an interaction between misinformation and percent of GOP voters, finding that being in a majority Republican versus Democratic state moderates the association between misinformation and vaccine hesitancy (Table S4). A scatterplot of republican and democratic-leaning counties confirms the moderation finding (Fig.2 in the main manuscript). Third, we run models adding the number of tweets per county as a control variable to address variation in the volume of Twitter activity across counties. Adding this covariate did not affect results. Fourth, as at the state level, we generate versions of the vaccine misinformation variable using a restricted set of keywords. Again, these results are consistent with our main findings (Table S5). Fifth, we examine the robustness of the threshold of 100 Twitter accounts per county for inclusion in the analysis, setting thresholds of 50 and 10. These results are similar to the main findings (Tables S6 and S7), demonstrating that results are robust to different variable specifications.

To confirm the relationship between misinformation and GOP vote share, we compute a negative binomial regression model predicting mean percent information (untransformed) at the county level using percent GOP vote and a set of control variables. This multivariate analysis confirms the bivariate correlation, indicating a strong relationship between these factors net of potential confounding variables (Table S8).

Table S1. Weighted/ordinary least squares regression of state-level percent vaccine hesitancy and daily vaccination rate per million on misinformation and covariates (N=50 states).

	(1) Vaccine hesitancy	(2) Vaccine hesitancy	(3) Vaccination rate	(4) Vaccination rate
	b (SE)	b (SE)	b (SE)	b (SE)
Mean % low credibility tweets	8.093*	6.877**	-3444.858**	-3518.02**
	(3.04)	(2.43)	(1240.20)	(1277.08)
% GOP vote (10% change)	3.996***	2.960***	-606.567***	-640.319**
	(0.38)	(0.42)	(140.32)	(208.11)
% below poverty line		0.530**		18.173
		(0.15)		(81.84)

% aged 65+		-0.197		171.533
		(0.15)		(100.14)
% Asian		0.011		13.213
		(0.07)		(27.74)
% Black		0.124**		-40.491
		(0.04)		(22.54)
% Hispanic		-0.066*		4.564
		(0.03)		(19.71)
% Indigenous		-0.138		71.890
		(0.12)		(51.00)
COVID deaths/thousand		-0.221		217.490
		(0.42)		(262.06)
Constant	1.858	3.024	11586.785***	9126.137***
	(1.65)	(2.72)	(708.20)	(1537.38)
R^2	0.797***	0.937***	0.457***	0.641***
BIC	225.217	194.454	836.580	843.252

Notes: Vaccine hesitancy is based on state-level means from Facebook survey data. The vaccination rate is vaccines administered per million (CDC data). For models predicting vaccine hesitancy (i.e., state means), analytic weights based on sample size are applied. Unstandardized betas and standard errors are provided.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S2. Weighted/ordinary least squares regression of state-level percent vaccine hesitancy and daily vaccination rate per million on misinformation (logged) and covariates (N=50 states).

	(1)	(2)	(3)	(4)
	Vaccine hesitancy	Vaccine hesitancy	Vaccination rate	Vaccination rate
	b (SE)	b (SE)	b (SE)	b (SE)
Logged mean % low cred tweets	4.136**	3.257**	-1669.206*	-1593.014*
	(1.53)	(1.19)	(636.52)	(660.59)
% GOP vote (10% change)	3.945***	2.962***	-601.418***	-676.915**
	(0.38)	(0.42)	(143.03)	(210.70)
% below poverty line		0.515**		29.711
		(0.15)		(83.31)
% aged 65+		-0.158		158.518
		(0.14)		(101.53)
% Asian		0.009		8.878
		(0.07)		(28.09)
% Black		0.130**		-42.750
		(0.04)		(22.90)
% Hispanic		-0.062*		1.398
		(0.03)		(19.93)
% Indigenous		-0.129		70.503
		(0.12)		(51.98)
COVID deaths/thousand		-0.235		224.368
		(0.42)		(268.26)
Constant	-1.206	0.184	12824.574***	10520.814***
	(2.18)	(2.39)	(980.40)	(1530.22)
R^2	0.798***	0.936***	0.448***	0.627***
BIC	225.049	194.982	837.352	845.150

Notes: Vaccine hesitancy is based on state-level means from Facebook survey data. The vaccination rate is actual vaccines administered per million (CDC data). For models predicting vaccine hesitancy (i.e., state means), analytic weights based on sample size are applied. Unstandardized betas and standard errors are provided. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S3. Weighted/ordinary least squares regression of state-level percent vaccine hesitancy and daily vaccination rate per million on misinformation (restricted keywords) and covariates (N=50 states).

	(1)	(2)	(3)	(4)
	Vaccine hesitancy	Vaccine hesitancy	Vaccination rate	Vaccination rate
	b (SE)	b (SE)	b (SE)	b (SE)
Logged mean % low cred tweets	8.320**	7.108**	-3342.575**	-3517.510**
	(2.97)	(2.37)	(1200.22)	(1236.41)
% GOP vote (10% change)	3.982***	2.944***	-611.854***	-648.565**
	(0.37)	(0.41)	(139.58)	(204.44)
% below poverty line		0.517**		27.129
		(0.15)		(81.32)
% aged 65+		-0.206		170.945
		(0.15)		(99.35)
% Asian		0.003		16.019
		(0.07)		(27.87)
% Black		0.125**		-42.464
		(0.04)		(22.25)
% Hispanic		-0.065*		2.774
		(0.03)		(19.42)
% Indigenous		-0.132		68.678
		(0.12)		(50.75)
COVID deaths/thousand		-0.216		225.119
		(0.42)		(259.70)
Constant	1.841	3.313	11575.126***	9085.430***
	(1.64)	(2.71)	(706.47)	(1530.36)
R^2	0.800***	0.938***	0.457***	0.645***
BIC	224.530	193.465	836.543	842.724

Notes: Vaccine hesitancy is based on state-level means from Facebook survey data. The vaccination rate is actual vaccines administered per million (CDC data). For models predicting vaccine hesitancy (i.e., state means), analytic weights based on sample size are applied. Unstandardized betas and standard errors are provided. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S4. Weighted least squares regression of county-level percent vaccine hesitancy on misinformation (logged) and covariates (N=548 counties, minimum 100 accounts/county).

	(1) b (SE)	(2) b (SE)	(3) b (SE)	(4) b (SE)
Logged mean % low credibility tweets	1.411**	4.304***	1.018***	4.278***
	(0.47)	(0.78)	(0.28)	(0.59)
% GOP vote (10% change)	2.926***		3.663***	
	(0.29)		(0.16)	
Majority GOP state (1=GOP; 0=Dem)		12.147***		11.230***
		(1.53)		(1.23)
GOP state * Logged low credibility		-3.585***		-3.420***
		(0.99)		(0.76)
% below poverty line			0.377***	0.402***
			(0.07)	(0.08)
% aged 65+			-0.055	-0.087
			(0.05)	(0.05)
% Asian			0.028	-0.174**
			(0.03)	(0.05)
% Black			0.203***	0.091***
			(0.02)	(0.03)
% Hispanic			0.002	-0.029
			(0.02)	(0.02)
% Indigenous			0.031	-0.113
			(0.19)	(0.14)
Rural-urban continuum code			0.437	0.594
			(0.26)	(0.34)
COVID deaths/thousand			0.545	0.883**
			(0.28)	(0.28)
Constant	6.979***	13.759***	-3.896***	8.013***
	(1.15)	(1.00)	(0.94)	(1.32)
R^2	0.500***	0.419***	0.804***	0.661***
BIC	3151.490	3240.010	2687.200	2995.562

Notes: Vaccine hesitancy is based on county-level means from Facebook survey data. Misinformation is measured using mean percent of low credibility tweets for counties with at least 100 Twitter accounts. Analytic weights based on Facebook survey sample size are applied, and models use cluster robust standard errors to account for counties being nested in states. Unstandardized betas and standard errors are provided.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S5. Weighted least squares regression of county-level percent vaccine hesitancy on misinformation (logged, restricted keywords) and covariates (N=548 counties, minimum 100 accounts/county).

	(1) b (SE)	(2) b (SE)	(3) b (SE)	(4) b (SE)
Logged mean % low credibility tweets	1.510**	4.382***	1.074***	4.319***
	(0.46)	(0.73)	(0.27)	(0.53)
% GOP vote (10% change)	2.905***		3.641***	
	(0.29)		(0.15)	
Majority GOP state (1=GOP; 0=Dem)		12.010***		11.132***
		(1.49)		(1.16)
GOP state * Logged low credibility		-3.530***		-3.392***
		(0.94)		(0.70)
% below poverty line			0.375***	0.394***
			(0.07)	(0.08)
% aged 65+			-0.058	-0.095
			(0.05)	(0.05)
% Asian			0.028	-0.171**
			(0.03)	(0.05)
% Black			0.202***	0.091***
			(0.02)	(0.03)
% Hispanic			0.002	-0.030
			(0.02)	(0.02)
% Indigenous			0.038	-0.101
			(0.19)	(0.13)
Rural-urban continuum code			0.451	0.648
			(0.26)	(0.33)
COVID deaths/thousand			0.546*	0.916**
			(0.26)	(0.28)
Constant	6.937***	13.673***	-3.849***	7.981***
	(1.14)	(0.95)	(0.93)	(1.29)
R^2	0.501***	0.423***	0.805***	0.665***
BIC	3136.899	3222.391	2673.021	2975.819

Notes: Vaccine hesitancy is based on county-level means from Facebook survey data. Misinformation is measured using mean percent of low credibility tweets for counties with at least 100 Twitter accounts. Analytic weights based on Facebook survey sample size are applied, and models use cluster robust standard errors to account for counties being nested in states. Unstandardized betas and standard errors are provided.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S6. Weighted least squares regression of county-level percent vaccine hesitancy on misinformation (logged) and covariates (N=658 counties, minimum 10 accounts/county).

	(1) b (SE)	(2) b (SE)	(3) b (SE)	(4) b (SE)
Logged mean % low credibility tweets	1.078*	3.252**	0.938***	3.669***
	(0.47)	(1.11)	(0.22)	(0.75)
% GOP vote (10% change)	3.140***		3.754***	
	(0.29)		(0.15)	
Majority GOP state (1=GOP; 0=Dem)		11.307***		10.597***
		(1.35)		(1.19)
GOP state * Logged low credibility		-2.467*		-2.748**
		(1.16)		(0.83)
% below poverty line			0.371***	0.381***
			(0.07)	(0.07)
% aged 65+			-0.058	-0.109
			(0.06)	(0.06)
% Asian			0.023	-0.225***
			(0.02)	(0.05)
% Black			0.205***	0.090***
			(0.02)	(0.03)
% Hispanic			0.003	-0.028
			(0.02)	(0.02)
% Indigenous			-0.003	-0.065
			(0.12)	(0.11)
Rural-urban continuum code			0.591*	0.726*
			(0.22)	(0.32)
COVID deaths/thousand			0.545	1.013***
			(0.28)	(0.28)
Constant	6.565***	14.976***	-4.212***	9.159***
	(1.12)	(1.11)	(0.98)	(1.51)
R^2	0.534***	0.421***	0.812***	0.662***
BIC	3796.413	3945.657	3251.830	3642.051

Notes: Vaccine hesitancy is based on county-level means from Facebook survey data. Misinformation is measured using mean percent of low credibility tweets for counties with at least 10 Twitter accounts. Analytic weights based on Facebook survey sample size are applied, and models use cluster robust standard errors to account for counties being nested in states. Unstandardized betas and standard errors are provided.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S7. Weighted least squares regression of county-level percent vaccine hesitancy on misinformation (logged) and covariates (N=628 counties, minimum 50 accounts/county).

	(1) b (SE)	(2) b (SE)	(3) b (SE)	(4) b (SE)
Logged mean % low credibility tweets	1.347**	4.241***	1.024***	4.229***
	(0.42)	(0.78)	(0.24)	(0.59)
% GOP vote (10% change)	3.039***		3.724***	
	(0.27)		(0.15)	
Majority GOP state (1=GOP; 0=Dem)		12.194***		11.210***
		(1.37)		(1.09)
GOP state * Logged low credibility		-3.350***		-3.239***
		(0.90)		(0.69)
% below poverty line			0.380***	0.410***
			(0.07)	(0.08)
% aged 65+			-0.058	-0.097
			(0.06)	(0.05)
% Asian			0.030	-0.174**
			(0.03)	(0.05)
% Black			0.203***	0.088***
			(0.02)	(0.02)
% Hispanic			0.001	-0.033
			(0.02)	(0.02)
% Indigenous			-0.009	-0.084
			(0.12)	(0.10)
Rural-urban continuum code			0.550*	0.693*
			(0.23)	(0.31)
COVID deaths/thousand			0.535	0.928**
			(0.28)	(0.27)
Constant	6.657***	13.836***	-4.219***	7.994***
	(1.12)	(0.99)	(0.95)	(1.32)
R^2	0.524***	0.439***	0.808***	0.666***
BIC	3619.976	3729.469	3099.757	3455.186

Notes: Vaccine hesitancy is based on county-level means from Facebook survey data. Misinformation is measured using mean percent of low credibility tweets for counties with at least 50 Twitter accounts. Analytic weights based on Facebook survey sample size are applied, and models use cluster robust standard errors to account for counties being nested in states. Unstandardized betas and standard errors are provided.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S8. Negative binomial regression of county-level misinformation on percent GOP vote and covariates (N=548 counties).

	b (SE)
% GOP vote (10% change)	0.261***
	(0.04)
% below poverty line	-0.020*
	(0.01)
% aged 65+	0.043***
	(0.01)
% Asian	0.017
	(0.01)
% Black	0.013***
	(0.00)
% Hispanic	0.006*
	(0.00)
% Indigenous	0.031*
	(0.02)
Rural-urban continuum code	-0.067
	(0.04)
COVID deaths/thousand	-0.092
	(0.06)
Constant	-2.642***
	(0.23)
Wald chi-squared	228.380***
BIC	774.933

Notes: Misinformation is measured using mean percent of low credibility tweets for counties with at least 100 Twitter accounts. Models use cluster robust standard errors to account for counties being nested in states. Negative binomial regression is employed due to zero-inflated Poisson distribution. Unstandardized betas and standard errors are provided.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S9. Description of covariates used during analyses.

(1)	(2)	(3)	(4)
Stata variable	Description	Year	Source
vaxrate	Daily number of people vaccinated per million	2021	Centers for Disease Control and Prevention
lowcred	Mean percentage of low credibility shared (per user)	2021	Twitter API
loglowcred	Natural logarithm of the mean percentage of low credibility shared (per user)	2021	Twitter API
propgop	Proportion of votes for Republican candidate	2020	Fox News, Politico, New York Times
covidmortality	Total COVID 19 deaths	2021	Centers for Disease Control and Prevention
population	Census Population	2010	United States Census
vMedHHInc	Median Household Income	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)
ppoverty	Percentage of people of all ages in poverty	2019	United States Department of Agriculture (County-Level Datasets)
vPercBachelors	Percent of adults with a bachelor's degree or higher	2015-2019	United States Department of Agriculture (County-Level Datasets)
vUnemployment_rate_2019	Unemployment rate	2019	United States Department of Agriculture (County-Level Datasets)
vTOTRATE	Rates of religious adherence per 1,000 population (200+ religions)	2010	Association of Religious Data Archives
vUnder18Pct2010	Percentage of population age 18 years or younger	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)
vAge65AndOlderPct2010	Percentage of population age 65 years or older	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)
vAsianNonHispPct2010	Percentage of population Asians (Non-Hispanic)	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)
vBlackNonHispPct2010	Percentage of population Black (Non-Hispanic)	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)

vHispanicPct2010	Percentage of population Hispanic	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)
vNatAmNonHispPct2010	Percentage of population Native American (Non-Hispanic)	2010	United States Department of Agriculture (Atlas of Rural and Small-Town America)

Supplementary Bibliography

1. DeVerna, M. R. *et al.* CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines. *ArXiv210107694 Cs* (2021).

2. Towns, J. *et al.* XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **16**, 62–74 (2014).

3. DeVerna, M. R. *et al.* Data for CoVaxxy: A collection of English-language Twitter posts about COVID-19 vaccines. https://github.com/osome-iu/CoVaxxy (2021).

4. Dredze, M., Paul, M. J., Bergsma, S. & Tran, H. *Carmen: A Twitter Geolocation System with Applications to Public Health*.

5. Lazer, D. M. J. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).

6. Shao, C. *et al.* The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 4787 (2018).

7. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374–378 (2019).

8. Pennycook, G. & Rand, D. G. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl. Acad. Sci.* **116**, 2521–2526 (2019).

9. Bovet, A. & Makse, H. A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **10**, 7 (2019).

10. Golding, B. Iffy+ Mis/Disinfo Sites. *Iffy.news* https://iffy.news/iffy-plus/ (2020).

11. MIT Election Data and Science Lab. U.S. President 1976–2020. (2021) doi:10.7910/DVN/42MVDX.

12. Farrow, D. C., Brooks, L. C., Rumack, A., Tibshirani, R. J. & Rosenfeld, R. Delphi Epidata API. https://github.com/cmu-delphi/delphi-epidata (2015).