



OPEN Mask images on Twitter increase during COVID-19 mandates, especially in Republican counties

Xiaofeng Lin¹, Georgia Kernell^{2,3✉}, Tim Groeling², Jungseock Joo², Jun Luo² & Zachary C. Steinert-Threlkeld^{3,4}

Wearing masks reduces the spread of COVID-19, but compliance with mask mandates varies across individuals, time, and space. Accurate and continuous measures of mask wearing, as well as other health-related behaviors, are important for public health policies. This article presents a novel approach to estimate mask wearing using geotagged Twitter image data from March through September, 2020 in the United States. We validate our measure using public opinion survey data and extend the analysis to investigate county-level differences in mask wearing. We find a strong association between mask mandates and mask wearing—an average increase of 20%. Moreover, this association is greatest in Republican-leaning counties. The findings have important implications for understanding how governmental policies shape and monitor citizen responses to public health crises.

Widespread mask wearing greatly reduces COVID-19 transmission^{1–6}. Accurate and continuous measures of mask use are therefore necessary for public health agencies to understand and predict outbreaks, identify susceptible populations, and formulate timely policy responses. Throughout the COVID-19 pandemic, health and government officials, as well as the general public, received real-time access to important information such as hospitalizations, deaths, and vaccination rates. Yet, data on preventative behavior is largely retrospective or unknown.

Our study addresses these shortcomings by presenting a novel way to measure individual-level behaviors in real time using geotagged social media images. The contribution is threefold. First, we develop an automated image classifier using a convolutional neural network (CNN) to detect images of people wearing masks and apply this classifier to geotagged Twitter data from March 1 through September 11, 2020. Twitter data were collected in real time and represent all publicly available (and approximately one-third of actual) geotagged tweets from the United States at this time. Second, we demonstrate that social media behavior closely tracks survey data using YouGov's COVID-19 Public Monitor and Democracy Fund + UCLA Nationscape^{7,8}. In doing so, we find that several individual-level correlates of mask wearing documented in observational research—age, gender, and partisanship—are also predictive of mask wearing on Twitter. Third, we investigate county-level factors that may impact widespread mask wearing. With 18,968,038 total geotagged tweets with images from 1,451,591 unique users, 73,489 of which posted mask wearing images, Twitter data are vast and allow us to examine less populated areas commonly underrepresented in surveys. We combine information on local mask wearing with county-level data on mask mandate policies, COVID-19 death rates, election returns, and individual mobility, as well as national media attention. While we find that a county's 2016 GOP presidential vote share is negatively associated with mask wearing overall, the introduction of mask mandates in Republican-leaning counties is associated with larger increases in mask-wearing images than in counties that lean Democratic. National media attention and urbanization are also predictive of mask wearing, though local measures of population density and mobility are not.

Benefits and drawbacks to using Twitter

This study highlights myriad advantages of using Twitter to measure the effect of political interventions on public health. First, since Twitter data are collected passively, they often lend themselves to long-term temporal estimates better than survey data. This feature is especially important for surprising events because it takes time to generate, test, and administer a survey. Twitter is able to generate estimates of mask wearing much earlier than surveys,

¹Department of Statistics, University of California - Los Angeles, Los Angeles, CA 90024, USA. ²Department of Communication, University of California - Los Angeles, Los Angeles, CA 90024, USA. ³Department of Political Science, University of California - Los Angeles, Los Angeles, CA 90024, USA. ⁴Department of Public Policy, Luskin School of Public Affairs, University of California - Los Angeles, Los Angeles, CA 90024, USA. ✉email: gkernell@ucla.edu

even ones conducted online such as the U.S. COVID-19 Trends and Impact Survey (CTIS)⁹. Second, Twitter data enables the measurement of precise geographic areas. In this case, we focus on the county-level, where the majority of mask mandates took place. Third, survey respondents may misrepresent their true behaviors due to social desirability bias, a problem especially concerning for politically-sensitive behaviors. Since Twitter users do not think of themselves as survey respondents, measurements using the social media platform are less likely to exhibit bias. Last, surveys may not ask about mask wearing: CTIS, for example, did not start asking about masks until September 8, at which point mask mandates were already pervasive. The large, passively monitored, and less strategic behavior on Twitter compared to surveys has proved fruitful for a wide range of social and political studies, such as of crime, natural disasters, mass gatherings, and international conflict^{10–16}.

At the same time, Twitter data have disadvantages stemming from non-representativeness and non-probability sampling. Surveys and academic studies show that Twitter users are not representative samples of the offline population^{17–19}. Users must geotag a tweet in order for it to appear in this paper's sample, introducing concerns about bias in users who employ this feature compared with those who do not. While these concerns are valid, it is worth noting that previous research finds Twitter users who share images without location information are not measurably different than those who share geotagged images²⁰, suggesting that relying on users who geotag face mask images may not bias results compared to the larger Twitter population.

Because Twitter requires no identifying information for accounts, probability sampling is not feasible. This paper's use of non-probability sampling is therefore problematic if the goal is to produce population estimates (Note: CTIS also collected data online via non-probability sampling using Facebook surveys.) We acknowledge and accept this limitation, given the fact that surveys do not provide sufficient geographic detail to measure the impacts of mask mandates. By applying computer vision techniques to a novel type of data, we have chosen a sampling design fit for this purpose²¹. However, the point estimates in this paper are most important for the trends they reveal and not as estimates of true rates of mask adherence.

Materials and Methods

Geotagged Twitter database. We compiled a dataset of geotagged publicly-available tweets from March 1 through September 11, 2020, collected in real time from the Twitter Application Programming Interface (API), using Twitter's POST statuses/filter streaming endpoint. Approximately 3% of tweets in English are geocoded, and Twitter's API (v1.1) returns approximately 1/3 of all geotagged tweets^{22,23}, resulting in 170,014,835 tweets from 2,603,654 U.S. users. Of these, 18,968,038 tweets contained an image, and 1,451,591 unique users posted at least one image.

Each geotagged tweet contains a global bounding box of coordinates identifying the tweet's location. However, the size of the bounding box varies because users can specify one of five levels to geotag: country, state, city, neighborhood, or point of interest. We dropped all tweets with boxes above the city level (7% of tweets), because they were too large to provide county-level information. For cities that span multiple counties (14% of tweets that specify the city level), we assign the tweet to the county that covers the greatest share of the city's population²⁴. All Twitter images analyzed in this study are publicly available and considered part of the public domain. (The use of public geotagged Twitter data was authorized by the University of California, Los Angeles Institutional Review Board, approval number #18-001354. All methods were performed in accordance with the UCLA IRB regulations and guidelines as well as the Declaration of Helsinki.)

Mask detection. To automatically detect mask-wearing images, we developed an image classifier using a convolutional neural network (CNN) with the ResNet-50 architecture²⁵. CNNs have been widely used for automated visual content analysis including facial mask detection^{25–28}. We took a supervised learning approach to collect and annotate images with mask-wearing labels to train and evaluate our model. The quality of training data is critical to ensure the model's performance. Therefore, we used an iterative approach to collect diverse and challenging data to make the final model robust.

To this end, we first collected 8000 images using Google's image search API for three mask-related keyword phrases ('wearing masks,' 'face covering,' and 'mask selfie'), as well as three keyword phrases not related to mask wearing ('selfie,' 'hangout,' and 'stay at home'). Approximately 1300 images were collected for each phrase.

We then manually annotated each of the 8000 images as mask-wearing or non-mask-wearing. To be coded as mask-wearing, an image must show at least one human face occupying at least 5% of the image's area wearing a medical or cloth mask over the mouth and nose. The rest of the images are negative samples that contain, for example, no face, faces without masks, faces occupying less than 5% of the image's area, or faces in advertisements for masks. Such "hard" negatives samples present visuals similar to mask-wearing scenes and thus train a more robust classifier (See the work of Shrivastava et al.²⁹, for example).

We use an active learning approach for robust model training that takes advantage of iterative model training and hard-data mining. Specifically, we first trained an initial model using the above dataset and applied it to a large set of unlabeled images, sourced from geotagged Twitter images from March 1–5, April 1–5, and May 20–25. Among these, 9391 images were classified positive. We manually verified their correctness and added the images and labels to the original set, resulting in 17,391 images. The optimal model parameters in the ResNet-50 classifier were obtained by stochastic gradient descent (SGD) with a binary cross entropy loss function, 500 training epochs, learning rate = 0.005, weight decay = 0.0004, and momentum = 0.9, as well as initial weights pre-trained on Imagenet data³⁰. We did not include Twitter images predicted as negative by the initial classifier because there were so many of them, and the recall was very low. Finally, we randomly divided the annotated images in a training set (80%, 13,913 images) and a validation set (20%, 3478 images). These labeled images further fine tune the initial classifier. All manual labeling mentioned above was completed by one human coder. Since these annotations are not used for measurements in actual analysis, it is acceptable to have only one coder,

Threshold	Precision	Recall	F1 Score
0.50	0.77	0.87	0.82
0.60	0.81	0.85	0.83
0.70	0.86	0.83	0.85
0.80	0.92	0.81	0.86
0.90	0.93	0.77	0.85

Table 1. Validation accuracy of mask prediction model from images with different decision thresholds. Precision is calculated as the proportion of predicted mask images that actually show a mask. Recall is the proportion of actual mask images predicted as a mask image. F1 score = $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$.

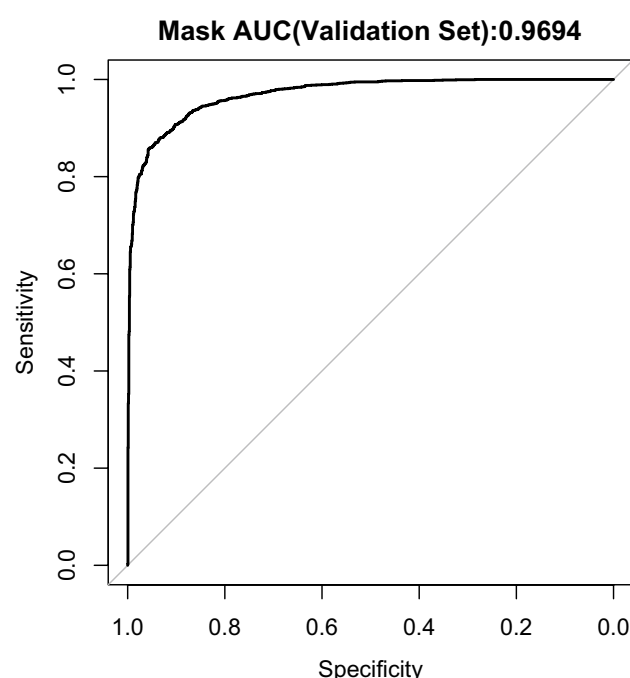


Figure 1. Receiver Operating Characteristics (ROC) curve for mask classifier on the validation set. Sensitivity is the fraction of actual positive data points that the model correctly classifies as positive. Specificity is the fraction of actual negative data points that the model correctly classifies as negative. Higher Area Under the Curve (AUC) values are associated with a greater ability to distinguish positive from negative examples.

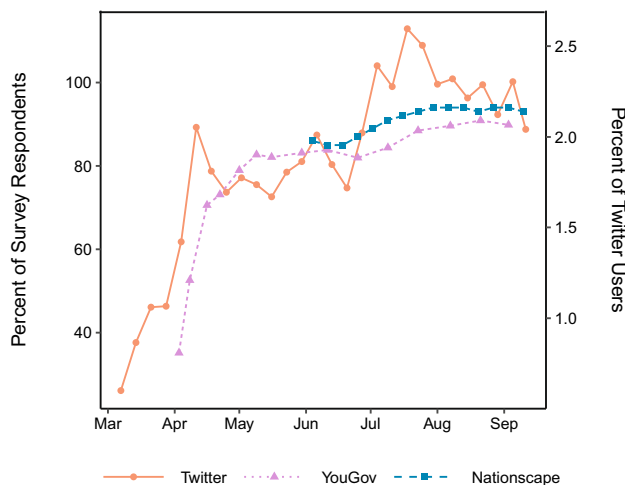
as long as the model performance is satisfactory. Two other coders verified the main coder's work by annotating a 1000-image random subset of the training set. The Krippendorff's alphas between the primary coder and the two additional coders were 0.865 and 0.870, showing strong inter-rater reliability.

Last, we chose an optimal decision threshold. Naively choosing 0.5 as the threshold may still yield an undesirable result (e.g., a too high false positive rate). Table 1 displays the classifier performance when using different decision thresholds on the same set. Since thresholds 0.7, 0.8, and 0.9 produce similarly high F1 scores (0.85, 0.86, and 0.85, respectively), we use a threshold of 0.9 in order to maximize precision (0.93). (Among all Twitter images, the proportion showing a mask is low. We manually verified a random subset of 1000 geotagged Twitter images posted on Aug 1, 2020 and found only 11 positive cases. To maximize our model precision in the face of potentially high false positives, we used the 0.9 threshold). With this threshold, the model has high classification accuracy. The model still demonstrates strong predictive power across other thresholds, as shown in the Receiver Operating Characteristic graph in Fig. 1.

Mask mandate coding. For the county-level analysis, the unit of observation is the county-week dyad, which covers a Monday-Sunday week. Each observation is coded as having a mask mandate in place if a mandate started on or before the Thursday in that week.

Case-control sampling. For the individual-level analyses, we perform a case-controlled sampling design by collecting all tweets with mask-wearing images, as well as a random sample of 1000 tweets per day from the set of non-mask images³¹. Then we extract demographic features of users who posted these tweets to calculate odds ratios.

	Min	Median	Mean	Q3	95% percentile	Max	Std
Image users	1	16	110.19	60	443	16,410	450.8
Mask users	0	0	1.98	1	8	343	9.2

Table 2. Weekly number of users who posted images/mask-wearing images by county.**Figure 2.** Mask wearing from March through September, 2020. The left vertical axis shows percentage of survey respondents who reported wearing a mask in public. The right vertical axis shows the percentage of Twitter users who posted a mask image (among those who posted at least one image in a geotagged tweet).

Sample size determination. The Twitter data is imbalanced across geographical regions. Table 2 displays key summary statistics for the weekly number of Twitter images and users who posted images in the geotagged Twitter dataset. In some less populated counties, the sample size become too small for our analysis to have sufficient statistical power. To determine the number of tweets with images needed per county, the following calculation is performed. The null and alternative hypothesis proportions are $p_\mu = 0.005$ and $p_\alpha = 0.02$, respectively, which are equal to the approximate minimum (p_μ) and maximum (p_α) values of the nationally aggregated mask wearing on Twitter. Let the level of significance $\alpha = 0.05$ and the desired power $= 1 - \beta = 0.9$, where $\beta = 0.1$ is the probability of committing a Type II error. Using Cohen's arcsine transformation³², the effect size is $h = 2 * (\arcsin \sqrt{p_\alpha} - \arcsin \sqrt{p_\mu}) = 0.142$. Under the null hypothesis, μ_0 , the decision boundary $\hat{p}$ should have z-score Z_0 , where $P(Z < Z_0) = 1 - \alpha / 2 = 0.975$ for $Z \sim N(0, 1)$, so $Z_0 = 1.959$. Under the alternative hypothesis, μ_a , $\hat{X}$ should have a z-score Z_a , where $P(Z < Z_a) = \beta = 0.1$, so $Z_a = -1.281$. Thus the required sample size $n = \left(\frac{Z_0 - Z_a}{h} \right)^2 = 519$. In our sample, 463 of 3006 counties have at least 519 unique Twitter users with images in the entire collected geotagged dataset. These counties account for 77% of the U.S. population.

Results

Validating Twitter image data using national surveys. To assess how well Twitter image data captures mask-wearing trends across the nation, we compare aggregate levels of mask wearing in our sample with self-reports from two nationally-representative surveys run during this period: YouGov's COVID-19 Public Monitor⁷ and UCLA Nationscape⁸. These online surveys asked the public whether or not they wore a mask in public spaces during the past week. (YouGov's survey stated: "Thinking about the last 7 days, how often have you worn a face mask outside your home to protect yourself or others from coronavirus (COVID-19)?" Response options included "always," "frequently," "sometimes," "rarely," and "not at all." The UCLA Nationscape survey asked respondents "Have you done any of the following in the past week?" with "Worn a mask when going out in public" as one of the possible categories and a binary response option of "Yes" or "No.") The Twitter measure is the fraction of users posting any image that also posted a mask image over a given week.

Figure 2 reveals a strong association between the estimates using Twitter images and both surveys. (Note the dual vertical axes to accommodate different magnitudes from surveys and Twitter.) Both Twitter and YouGov reveal similar increases in public mask wearing from April to May. (Nationscape did not ask about mask wearing for the first half of the data.) All estimates increased at a similar level from June to September. High correlations between mask usage on Twitter and self-reports from YouGov ($r = 0.61, p < 0.01, N = 18$) and Nationscape ($r = 0.79, p < 0.01, N = 10$) suggest that social media images are highly predictive of similar behaviors offline. (In addition, the Twitter data correlate with a different YouGov study running from March through September 2020 at $r = 0.84$. We do not include this study in Fig. 2 because the individual-level data is not publicly available³³.) We also took the first differences of time series to address non-stationarity. The resulting stationary

Twitter time series is still significantly correlated with YouGov ($r = 0.63, p < 0.01, N = 17$) and Nationscape ($r = 0.80, p < 0.01, N = 9$). (We present correlations in the time-series by state in Supplementary Information Section 6.) County-level data allows for more precise estimates of how mask mandates and partisanship interact. (The only survey we know of that identifies mask wearing metrics at the county level is one conducted by *The New York Times* in July of 2020³⁴. That survey, however, only lasted for a 12-day period (July 2–July 14) and thus did not facilitate the same observational study overtime.)

Individual-level determinants. Previous research examining individual-level factors finds that older individuals are more likely to wear masks than their younger counterparts^{8,34–37} (although see the work of MacIntyre et al.³⁸ and Howard³⁹). Females are also more likely than males to wear a mask^{36,39–42}, and Democrats are consistently more likely than Republicans to not only to wear a mask, but to support mask mandates^{39–42}. To test the validity of our data, we compare mask wearing across age, gender, and political ideology for Twitter users. A user's age and gender are estimated by applying a deep neural network facial classifier trained by FairFace⁴³ on the user's profile image. *Conservative Leaning* measures a user's ideology based on the number of influential partisan-leaning Twitter accounts that individual follows. Users with a *Conservative Leaning* score greater than 0 are generally considered Republicans⁴⁴. Previous research identifies partisanship as the strongest predictor of mask use¹⁵.

Table S4 reports odds ratios for each demographic group in relation to the specified reference group. In line with previous research, we find that seniors, females, and Democrats demonstrate significantly higher likelihoods of mask wearing. (FairFace is also capable of classifying race, but the accuracy on Twitter profiles is unsatisfactory—64% on a random sample of 1000 profiles. The odds ratios between race groups classified by FairFace are not significant in our data.)

Similar to previous studies which reported yield rates when analyzing Twitter users^{44,46}, we are able to infer gender and age from 20.4% of users and partisan leaning from 39.4% of users. Accuracy rates for the algorithms we use that assign socio-demographic and political-leaning characteristics are extremely high. Previous work finds that the average classification accuracy for the algorithms we use is 94.4% for gender and 60.2% for age⁴³. Scholars have validated the Twitter partisanship measure employed here by matching it with campaign contribution records using name and zip code. The correlation between campaign contributions and political leaning is also very high ($r = 0.80$)⁴⁴. To provide further evidence that our partisanship estimates match the actual population, we correlated Republican vote share in the 2016 election with the share of Twitter users we infer to be Republicans—among those for whom we can infer political leaning—and found they are highly correlated. Across all weeks, they correlate at 0.63 ($N = 100, p < 0.01$). Across all counties, the correlation is 0.29, $N = 3009, p < 0.01$.

Many studies find that the inferred socio-demographic and political leaning variables are supported in survey data. The canonical article about Twitter and ideology⁴⁴ shows that the latter is recoverable from follower relationships on Twitter, and that this measure sorts accounts accurately⁴⁷. Comparing survey respondents' self-described Twitter behaviors with their actual Twitter behavior reveals that the two are positively correlated⁴⁷. Other work⁴⁸ connects Twitter users with voter files and shows that the demographics and ideology of Twitter users match voting administrative data. And, scholars have successfully estimated protesters' age and gender using Twitter images with surveys of actual protesters¹⁵.

Contextual determinants: mask mandates, mobility, death rates, and county partisanship. To examine contextual factors that shape mask wearing, we calculate the percent of users who post at least one mask image (among those who post any image) for each week in the data. By only including users who post at least one image of any kind, we control for unobserved heterogeneity that may reflect differences between users who share images and those who do not.

Mask mandate data come from Wright et al.⁴⁹. This database includes the earliest effective dates of local and state mask mandates for every U.S. county and runs through August 4, 2020. Figure 3 shows the share and distribution of counties, as well as the corresponding share of the U.S. population, under an active county or state mandate (or both) from May 1 through August 1, 2020. As of August 1, 2020, 66.3% of U.S. counties, and 87.4% of the U.S. population were under active mask-wearing mandates.

Table 3 presents summary statistics for each of the continuous explanatory variables of interest. Subscript i denotes different counties, t denotes different weeks, and $t - 1$ indicates a lag of one week. In selecting our control variables, we have attempted to address many of the factors that might plausibly be associated with variations in the incidence or severity of COVID-19, or with the likely local or personal response to the pandemic. For example, greater news coverage of the pandemic or a spike in the number of deaths might reasonably increase the probability an individual would choose to wear a mask, regardless of whether there is a governmental requirement to do so. Close proximity to others due to geography or the activity of individuals might also be expected to increase the precautions one might take against contracting COVID-19. Finally, there have been clear and ongoing partisan differences related to mask-wearing throughout the pandemic, as illustrated in various prior studies and surveys^{8,42,45}.

Deaths per 10k is the number of reported COVID-19 deaths in the county during the prior week per 10,000 residents⁵⁰. *GOP Vote 2016* is the share of the two-party vote that went to Donald Trump in the 2016 presidential election⁵¹. *Urban Population Percentage*⁵² is the percent of a county's population living in a Census-defined urban area. *Population Density*⁵³ is the county's population (using 2010 Census data) divided by its area in square miles. Media coverage of the pandemic may drive public attention to COVID-19 and increase mask wearing. *COVID News* measures the average number of times the word "COVID" is used per national morning or evening news show on ABC, CBS, and NBC⁵⁴. Last, images of mask wearing may decrease as people venture outside their

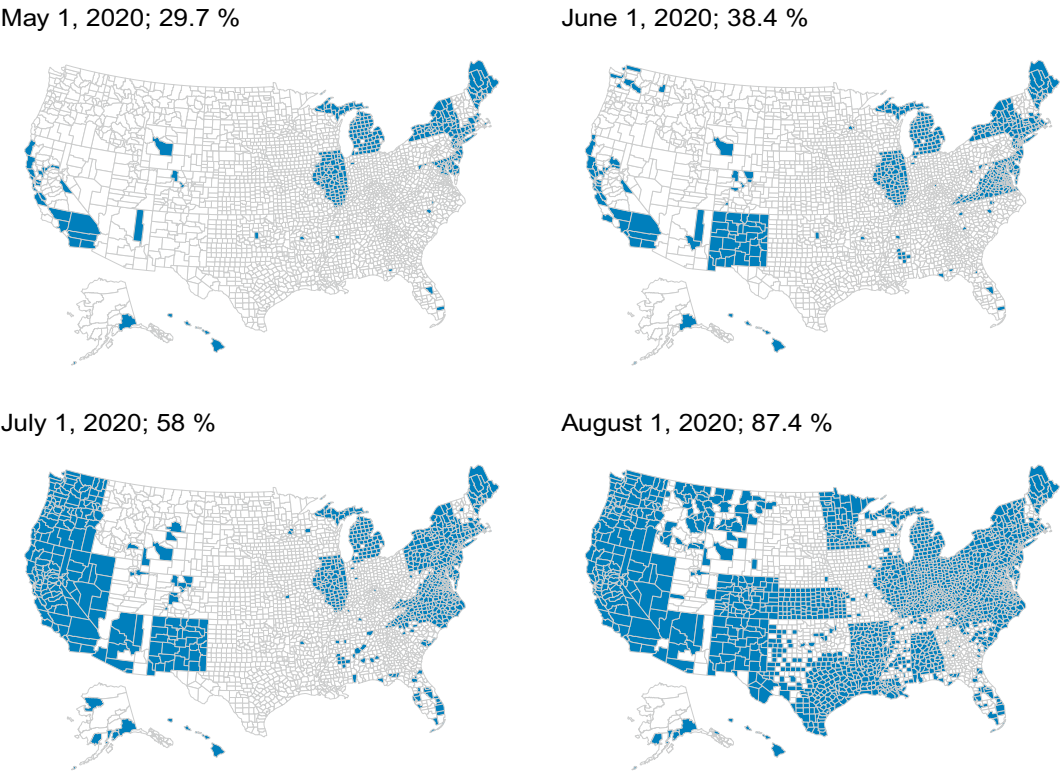


Figure 3. Counties under a mask mandate. Percentages indicate the share of the U.S. population under a mask mandate on the corresponding date. Mandate data are acquired from Wright et al.⁴⁹.

	Top 300 counties	All counties
Deaths per $10k_{i,t-1}$	0, 1.6	0, 1.0
GOP vote 2016_i	15.3, 70.1	23.7, 85.1
Population Density _{<i>i</i>}	81.7, 11,379.5	5.2, 2630.0
COVID News _{<i>t</i>}	0.9, 11	0.9, 11
Retail visit per Hundred _{<i>i,t</i>}	0.67, 4.24	0.61, 5.58

Table 3. Explanatory Variables 95% Intervals for the 300 most populous counties and all U.S. counties.

homes. We use Safegraph’s Places of Interest dataset to measure changes in activity⁵⁵. Specifically, *Retail Visit per Hundred* measures the number of visits to a grocery store, as defined by those with NAICS business codes 44 or 45, in a county during the current week per 100 residents. In addition, the following models include a *Week Counter*, which ranges from 1 to 22 and controls for unobserved time trends. See Table S7 for models with different subsets of control variables.

Like the population, tweets are imbalanced across counties. Although Twitter data generally produces samples larger than surveys in less populated areas, the sample sizes in the majority of the country’s 3,006 counties are still too small to use for a powerful statistical test (See Table S5 and S6 for sample sizes and numbers of Twitter users/survey respondents wearing masks). As a result, the main regression is run for the 300 most populous counties, which together account for approximately 64% of the United States population. All of the top 300 counties have more than 519 unique image-posting users, thus satisfying the sample size calculated in the Sample Size Determination section. There are no substantive differences found when we run the regression on the top 100, top 500, or all counties, as shown in Table S1.

Table 4 presents results from two multilevel generalized linear models⁵⁶ in the binomial family with random effects for week, county, and state. (A generalized linear model with a binomial family is ideal for modeling proportions that are otherwise prone to heteroskedasticity while accommodating lower and upper bounds at 0 and 1.) The first model examines the effect of *Mask Mandate* and *GOP Vote 2016* without an interaction effect. As we can see the mandate is positive and significant; overall, people are more likely to wear a mask when there is a mandate in place. Republican support is, as expected, negative and significant. More conservative counties on the whole see fewer individuals posting Twitter images with masks.

Variable	No interaction	With interaction
	Estimate(SE)	Estimate(SE)
Intercept	−5.129*** (0.257)	−5.065*** (0.256)
Mask mandate	0.058*** (0.014)	−0.055* (0.028)
GOP vote 2016	−0.005*** (0.001)	−0.006*** (0.001)
Mask mandate * GOP vote 2016		0.003*** (0.001)
Deaths per 10k	0.014 (0.007)	0.011 (0.007)
Urban population percentage	0.004* (0.002)	0.004* (0.002)
Population density	0.000 (0.000)	0.000 (0.000)
COVID News	0.066*** (0.014)	0.064*** (0.014)
Retail visit per hundred	0.021 (0.016)	0.008 (0.016)
Week counter	0.028*** (0.006)	0.029*** (0.006)
AIC	31989.304	31967.327
BIC	32070.837	32055.653
Log likelihood	−15982.652	−15970.663
County weeks	6600	6600
States (Including D.C)	45	45
Num. groups: week	22	22
County intercept	0.038	0.037
State intercept	0.014	0.014
Week intercept	0.027	0.027

Table 4. Contextual predictors of mask-wearing images. The dependent variable is the proportion of Twitter users who posted mask-wearing images among users that posted any tweets with images. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Now let us turn to the interaction model. (Because we include an interaction between *Mask Mandate* and *GOP Vote 2016*, we must interpret the effects of the two variables jointly. For example, the negative coefficient on *Mask Mandate* in the interaction model can only be interpreted for a county with zero percent GOP vote, which is never the case.) Under no mask mandate, county-level Republican leaning shows a negative, significant association with mask wearing: for a one standard deviation increase (12.8 percentage points) in GOP vote share in the 2016 presidential election, the expected level of mask wearing decreases 7.4 percent. Not surprisingly, mask wearing is lower in Republican strongholds, all else equal. This negative association is consistent with previous findings⁴⁵.

The introduction of a mandate, however, is strongly associated with an increase in mask wearing. Moreover, this association is especially large in counties that favored Donald Trump in 2016. To illustrate these interactive effects, Fig. 4 shows the predicted marginal effect of a mandate for varying levels of GOP vote share. Moving from the county with the lowest percentage vote for the GOP to the county with the highest vote increases the expected effect of mandates by 25 percentage points. Although mask mandates have a greater predicted effect in Republican counties, the overall level of mask wearing remains slightly higher in Democratic counties. Indeed, Republican counties with mask mandates display similar levels of mask wearing to Democratic counties without mandates. These results suggest that in majority-Democratic counties individuals are more likely to wear a mask regardless of whether there is a legal mandate requiring them to do so or not, whereas in majority-Republican counties mandates may prompt people to put on a mask, thus narrowing the gap caused by partisanship. (Using a binned estimator⁵⁷ to evaluate flexible marginal effects produces similar results.)

News coverage is positively associated with mask wearing images: when the average occurrence of the word “COVID” increases by one standard deviation (2.9 times), mask wearing is predicted to increase by 20.7%–21.4% across the models. To a large extent, the news variable mirrors what people are tweeting about. In particular, COVID-19 news decreased by approximately 60% from week 13 to week 16 during the protests following George Floyd’s death, and mask wearing images decreased during this time as well. (Table S2 shows the same regression models using cable rather than network news coverage and finds the same positive and significant association with mask wearing.)

Mask wearing also increases with urbanization: A one standard deviation rise in urbanization (25 percentage points) corresponds to a 10.5% increase in mask images. The remaining population variables - deaths and density - are not statistically significant in most models. Mask-wearing images also do not decrease with higher rates of grocery store visits. Not surprisingly, the positive coefficient on the linear week counter demonstrates

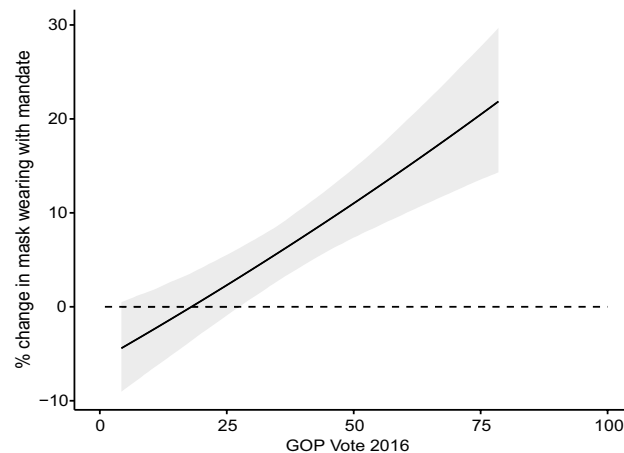


Figure 4. Marginal Effect of Mask Mandates for Varying GOP County Vote. Shaded areas represent 95% confidence intervals.

an overall increasing trend in public mask wearing. As time goes by, mask wearing becomes a habit for many individuals. Table S3 shows the estimation with a logged week counter and demonstrates the same effect of time.

Respondents in Nationscape are identified at both the state and the congressional district level, while respondents in YouGov are only identified at the state level. Thus, we performed another check of the hypotheses at the individual level focusing only on Nationscape respondents. The dependent variable indicates whether or not the person stated that they wore a mask in the previous week. The key independent variables are respondent Party ID (ranging from 1 = “strong Democrat” to 7 = “strong Republican”), Mask Mandate (indicating if there was a mandate in that congressional district in the week prior to the survey), and their interaction. We controlled for individual demographic variables including age, gender, ethnicity, household income, and education level. We also controlled for district-level features including population density and urban population proportion. As in the county-level analysis, we included a week counter to control for a general time trend, and national level COVID News variables. The regression output is reported in Table S8 in the Supplementary Information. We found the same trend as indicated on the county-level analyses: while higher GOP partisanship itself is correlated with lower mask wearing, the interaction of GOP partisanship and mask mandate is positive and significant, indicating that Republicans may be more likely to change their behavior during mask mandates than Democrats.

Discussion

Using Twitter, and digital trace data more broadly, confers a number of benefits. In contrast with surveys, Twitter data do not rely on retrospective responses, which often suffer due to memory loss and social desirability biases. Twitter data provide a measure that is free and can be collected and stored on demand in real time. The new Academic Search Product provides any published tweet that is still public, facilitating post hoc studies. In addition, the sheer volume of Twitter data allows researchers to estimate effects across groups that would otherwise be prohibitively costly. Though Twitter is not a representative sample of Americans¹⁹, it can serve as a meaningful thermometer for public opinion and health behavior, despite sometimes drastically different levels of measurement¹⁴.

Moreover, social media image data provide an invaluable and unobtrusive way to measure public health behaviors and trends^{58,59}. To our knowledge, this is among the earliest work to collect COVID-19 Twitter image data and apply deep learning based image classification⁶⁰, and the first to measure public health behaviors related to COVID-19 with such methods. Future research may employ social media images to identify other behaviors or attributes, such as smoking, alcohol or drug use, obesity, and seat-belt compliance.

Despite the advantages of using geotagged tweets to study public health behavior, there are drawbacks. One difficulty is class imbalance across outcomes (there are many more images without facemasks than with them), which could cause a classifier to fit out of sample data poorly⁶¹. We avoid this problem by constructing training datasets with an equal number of facemask and non-facemask images and later choose a classification threshold high enough to generate a precision of 0.93. A second drawback concerns the potential bias in user behavior during this time period. Given that people likely stay home more during mask mandates, they may also tweet less or remove masks when taking photos. Either behavior could cause us to underestimate changes in mask wearing. Future research may incorporate the content of tweets to better identify a person's location or to ascertain information about other user characteristics of interest, such as education or income. Because most geotagged tweets are only located at the level of neighborhoods or larger geographical regions, our research was unable to identify whether or not photos are taken in public settings. “Selfie” images taken indoors do not provide sufficient information to distinguish between public and private locations. (We also considered surveillance camera video data as an alternative indicator of mask-wearing behavior. Unfortunately, these data simply do not exist in a way that provides the coverage necessary for this study. Publicly-available surveillance data rely on footage from security cameras that use non-password-protected internet connections. Putting the ethical issues aside,

these feeds tend to not be taken in environments that are conducive to the study, such as residential buildings rather than commercial settings.)

The principle finding that mandates are more closely associated with mask wearing in areas with more Republican voters suggests that governmental policies have the ability to equalize health behaviors, and potentially health outcomes. We know that conservative voters are less likely to vaccinate against COVID-19⁶², and early research suggests that vaccine mandates are effective at increasing compliance^{63,64}. Although systematic experimental analyses have yet to be conducted, our study suggests that vaccine mandates may have had the greatest effect in GOP strongholds.

Last, it is worth noting that Twitter postings are not simply a random slice of life, but might also reflect individuals' strategic choices aimed at influencing friends or public opinion. These decisions likely change over time in response to the prevalence and politicization of mask wearing, and future research would benefit from exploring a dynamic relationship between mask images, mandates, and the political climate.

Data availability

All aggregate-level data, deidentified individual-level data, and replication files are available on Dataverse doi:10.7910/DVN/1ZBK71. The image classifier code and trained model are available through the Github repository <https://github.com/Bernardo1998/Mask-Wearing-in-Geolocated-Twitter>.

Received: 18 February 2022; Accepted: 31 October 2022

Published online: 09 December 2022

References

- Howard, J. *et al.* An evidence review of face masks against COVID-19. *Proc. Natl. Acad. Sci.* **118**(4), e2014564118 (2021).
- Brooks, J. T. & Butler, J. C. Effectiveness of mask wearing to control community spread of SARS-CoV-2. *JAMA* **325**, 998–999 (2021).
- National Center for Immunization and Respiratory Diseases, Division of Viral Diseases. Scientific Brief: Community Use of Cloth Masks to Control the Spread of SARS-CoV-2. <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/masking-science-sars-cov2.html> (2021).
- Costantino, V. & Raina MacIntyre, C. The impact of universal mask use on SARS-COV-2 in Victoria, Australia on the epidemic trajectory of COVID-19. *Front. Public Health* **9**, 307 (2021).
- Lyu, W. & Wehby, G. L. Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US. *Health Aff.* **39**, 1419–1425 (2020).
- Abaluck, J. *et al.* The impact of community masking on COVID-19: A cluster randomized trial in Bangladesh. <https://elischolar.library.yale.edu/cowles-discussion-paper-series/2642> (2021).
- Jones, S. P. Imperial College London YouGov COVID Data Hub. <https://github.com/YouGov-Data/covid-19-tracker>. Imperial College London Big Data Analytical Unit and YouGov Plc (2020).
- Tausanovitch, C. & Vavreck, L. Democracy Fund + UCLA Nationscape. <https://www.voterstudygroup.org/covid-19-updates#> (2020).
- Salomon, J. A. *et al.* The US COVID-19 Trends and Impact Survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proc. Natl. Acad. Sci.* **118**(5), e211454118 (2021).
- Gerber, M. S. Predicting crime using Twitter and kernel density estimation. *Decis. Support Syst.* **61**, 115–125 (2014).
- Sakaki, T., Okazaki, M. & Matsuo, Y. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web* 851–860 (2010).
- Chang, K.-C., Hobbs, W. R., Roberts, M. E. & Steinert-Threlkeld, Z. C. COVID-19 increased censorship circumvention and access to sensitive topics in China. *Proc. Natl. Acad. Sci.* **119**(4), e2102818119 (2022).
- Botta, F., Moat, H. S. & Preis, T. Quantifying crowd size with mobile phone and Twitter data. *R. Soc. Open Sci.* **2**, 150162 (2015).
- Sobolev, A., Chen, M. K., Joo, J. & Steinert-Threlkeld, Z. C. News and geolocated social media accurately measure protest size variation. *Am. Pol. Sci. Rev.* **114**, 1343–1351 (2020).
- Barrie, C. & Frey, A. Faces in the crowd: Twitter as an alternative to protest surveys. *PLoS ONE* **(16)**:11, e0259972 (2021).
- Driscoll, J. & Steinert-Threlkeld, Z. C. Social media and Russian territorial irredentism: Some facts and a conjecture. *Post Sov. Aff.* **36**(2), 101–121 (2020).
- Sloan, L., Morgan, J., Burnap, P. & Williams, M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *PLoS ONE* **10**(3), e0115545 (2015).
- Mellon, J. & Prosser, C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Res. Pol.* **4**(3), 205316801772000 (2017).
- Wojcik, S. & Hughes, A. Sizing up Twitter users. Pew Research Center (2019).
- Steinert-Threlkeld, Z. C., Chan, A. & Joo, J. How State and protester violence affect protest dynamics. *J. Pol.* **84**(2), 798–813 (2022).
- Kohler, U., Kreuter, F. & Stuard, E. A. Nonprobability sampling and causal analysis. *Annu. Rev. Stat. Appl.* **6**, 149–172 (2019).
- Steinert-Threlkeld, Z. C. *Twitter as Data (Elements in Quantitative and Computational Methods for the Social Sciences)* (Cambridge University Press, Cambridge, 2018).
- Huang, B., & Carley, K. M. A large-scale empirical study of geotagging behavior on Twitter. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 365–373 (2019).
- Indiana Business Research Center, City-to-County Finder. <https://www.statsamerica.org/CityCountyFinder/> (2021).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
- Bhuiyan, M. R., Khushbu, S. A. & Islam, M. S. A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. In *2020 11th International Conference on Computing, Communication and Networking Technologies* 1–5 (2020).
- Sethi, S., Kathuria, M. & Kaushik, T. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *J. Biomed. Inform.* **120**, 103848 (2021).
- Loey, M., Manogaran, G., Taha, M. H. N. & Khalifa, N. E. M. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustain. Cities Soc.* **65**, 102600 (2021).
- Shrivastava, A., Gupta, A. & Girshick, R. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 761–769 (2016).
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (2009).

31. LaMorte, W. W. Case-control studies. https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_analyticoverview/EP713_AnalyticOverview5.html#headingtaglink_1 (2017).
32. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, New York, 1988).
33. YouGov. Personal measures taken to avoid COVID-19. <https://today.yougov.com/topics/international/articles-reports/2020/03/17/personal-measures-taken-avoid-covid-19> (2021).
34. Katz, J., Katz, M. S. & Quealy, K. Faced with losing their jobs, even the most hesitant are getting vaccinated. *The New York Times* (2020).
35. Hutchins, H. J. *et al.* COVID-19 mitigation behaviors by age group—United States, April–June 2020. *Morb. Mortal. Wkly Rep.* **69**, 1584–1590 (2020).
36. Haischer, M. H. *et al.* Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic. *PLoS ONE* **15**, e0240785 (2020).
37. Knotek, II E. S., Schoenle, R. S., Dietrich, A. M., Müller, G. J., Myrseth, K. O., Weber, M. Consumers and COVID-19: Survey results on mask-wearing behaviors and beliefs. *Economic Commentary* (Federal Reserve Bank of Cleveland) 1–7 (2020).
38. MacIntyre, C. R. *et al.* Mask use, risk-mitigation behaviours and pandemic fatigue during the COVID-19 pandemic in five cities in Australia, the UK and USA: A cross-sectional survey. *Int. J. Infect. Dis.* **106**, 199–207 (2021).
39. Howard, M. C. The relations between age, face mask perceptions and face mask wearing. *J. Public Health* **fdab018** (2021).
40. Kramer, S. More Americans say they are regularly wearing masks in stores and other businesses. <https://pewrsr.ch/32trRi> (2020).
41. Naeim, A. *et al.* Effects of age, gender, health status, and political party on COVID-19-related concerns and prevention behaviors: Results of a large, longitudinal cross-sectional survey. *JMIR Public Health Surveill.* **7**, e24277 (2021).
42. Brenan, M. Americans' face mask usage varies greatly by demographics. <https://news.gallup.com/poll/315590/americans-face-mask-usage-varies-greatly-demographics.aspx> (2020).
43. Karkkainen, K. & Joo, J. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 1548–1558 (2021).
44. Barberá, P. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Pol. Anal.* **23**, 76–91 (2015).
45. Milosh, M., Painter, M., Sonin, K., Dijcke, D. V. & Wright, A. L. Unmasking partisanship: Polarization undermines public response to collective risk. *J. Public Econ.* **107**, 104538 (2020).
46. Jung, S. G., An, J., Kwak, H., Salminen, J. & Jansen, B. J. Inferring social media users' demographics from profile pictures: A Face++ analysis on Twitter users. In *International Conference on Electronic Business* 140–145 (2017).
47. Guess, A., Munger, K., Nagler, J. & Tucker, J. How accurate are survey responses on social media and politics?. *Polit. Commun.* **36**(2), 241–258 (2019).
48. Hughes, A. G. *et al.* Using Administrative records and survey data to construct samples of Tweeters and tweets. *Public Opin. Q.* **85**, 323–346 (2021).
49. Wright, A. L., Chawla, G., Chen, L. & Farmer, A. Tracking mask mandates during the COVID-19 pandemic. University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-104. Available at SSRN: <https://ssrn.com/abstract=3667149> (2020).
50. Ponce, M. & Sandhel, A. covid19.analytics: An R package to obtain, analyze and visualize data from the Coronavirus disease pandemic. [arXiv:2009.01091](https://arxiv.org/abs/2009.01091) (2020).
51. McGovern, T., Larson, S., Morris, B., Ro, J. & Hodges, M. U.S. county-level presidential election results. https://github.com/tonmcg/US_County_Level_Election_Results_08-20 (2020).
52. U.S. Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria. <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html> (2010).
53. Magdel, E. U.S. population density. <https://github.com/camillol/cs424p3> (2011).
54. Joo, J., Sison, F. F. & Turner, M. Red Hen Lab: Dataset and tools for multimodal human communication research. *KI-Künstliche Intell.* **31**, 357–361 (2017).
55. Safegraph. Weekly patterns. <https://docs.safegraph.com/v4.0/docs/weekly-patterns> (2020).
56. Crawley, M. *The R Book* (John Wiley & Sons, New Jersey, 2012).
57. Hainmueller, J., Mummolo, J. & Xu, Y. How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Pol. Anal.* **27**, 163–192 (2019).
58. Aiello, A. E., Renson, A. & Zivich, P. Social media-and internet-based disease surveillance for public health. *Annu. Rev. Public Health* **41**, 101–118 (2020).
59. Sinnenberg, L. *et al.* Twitter as a tool for health research: A systematic review. *Am. J. Public Health* **107**, e1 (2017).
60. Khurana, S., Chopra, R. & Khurana, B. Automated processing of social media content for radiologists: Applied deep learning to radiological content on twitter during COVID-19 pandemic. *Emerg. Radiol.* **28**, 477–483 (2021).
61. Ghosh, K., Banerjee, A., Chatterjee, S. & Sen, S. Imbalanced Twitter sentiment analysis using minority oversampling. In *Proceedings of the 2019 IEEE 10th International Conference on Awareness Science and Technology* 1–5 (2019).
62. Stolberg, S. G. O.P. Seethes at Biden Mandate, even in States requiring other vaccines. <https://www.nytimes.com/2021/09/12/us/politics/vaccine-mandates-republicans.html> (2021).
63. Blake, A. The evidence is building: Vaccine mandates work - and well. <https://www.washingtonpost.com/politics/2021/09/29/evidence-is-building-vaccine-mandates-work-well/> (2021).
64. Hsu, A. Faced with losing their jobs, even the most hesitant are getting vaccinated. <https://www.npr.org/2021/10/07/1043332198/> (2021).

Author contributions

Conceptualization: X.L., G.K., T.G., J.J., J.L., and Z.C.S.T. Research and Writing: X.L. and G.K. Analyses and Figures: X.L. Image Classifier: X.L. Twitter Data Acquisition: Z.C.S.T. Discussion: X.L., G.K., T.G., J.J., J.L., and Z.C.S.T.

Funding

This work was supported by National Science Foundation SBE/SMA #1831848.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-23368-6>.

Correspondence and requests for materials should be addressed to G.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022