

## RESEARCH ARTICLE

# Risk factors for increased COVID-19 case-fatality in the United States: A county-level analysis during the first wave

Jess A. Millar<sup>1\*</sup>, Hanh Dung N. Dao<sup>2</sup>, Marianne E. Stefopoulos<sup>3</sup>, Camila G. Estevam<sup>4</sup>, Katharine Fagan-Garcia<sup>5</sup>, Diana H. Taft<sup>6</sup>, Christopher Park<sup>7</sup>, Amaal Alruwaily<sup>8</sup>, Angel N. Desai<sup>9\*</sup>, Maimuna S. Majumder<sup>10\*</sup>

**1** Department of Epidemiology, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States of America, **2** Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States of America, **3** Child Health Evaluative Sciences Program, SickKids Hospital, Toronto, Canada, **4** Department of Public Health, State University of Campinas, Campinas, SP, Brazil, **5** Department of Medicine, University of Alberta, Edmonton, AB, Canada, **6** Department of Food Science and Technology, University of California Davis, Davis, CA, United States of America, **7** College of Global Public Health, New York University, New York, NY, United States of America, **8** Independent Scholar, Riyadh, Saudi Arabia, **9** Division of Infectious Disease, Department of Internal Medicine, University of California Davis Medical Center, Sacramento, CA, United States of America, **10** Harvard Medical School and Boston Children's Hospital, Boston, MA, United States of America

\* [jamillar@umich.edu](mailto:jamillar@umich.edu) (JAM); [andesai@ad3.ucdavis.edu](mailto:andesai@ad3.ucdavis.edu) (AND); [maimuna.majumder@childrens.harvard.edu](mailto:maimuna.majumder@childrens.harvard.edu) (MSM)



## OPEN ACCESS

**Citation:** Millar JA, Dao HDN, Stefopoulos ME, Estevam CG, Fagan-Garcia K, Taft DH, et al. (2021) Risk factors for increased COVID-19 case-fatality in the United States: A county-level analysis during the first wave. *PLoS ONE* 16(10): e0258308. <https://doi.org/10.1371/journal.pone.0258308>

**Editor:** Wen-Jun Tu, Chinese Academy of Medical Sciences and Peking Union Medical College, CHINA

**Received:** March 5, 2021

**Accepted:** September 23, 2021

**Published:** October 14, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0258308>

**Copyright:** © 2021 Millar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying the results presented in the study are available

## Abstract

The ongoing COVID-19 pandemic is causing significant morbidity and mortality across the US. In this ecological study, we identified county-level variables associated with the COVID-19 case-fatality rate (CFR) using publicly available datasets and a negative binomial generalized linear model. Variables associated with decreased CFR included a greater number of hospitals per 10,000 people, banning religious gatherings, a higher percentage of people living in mobile homes, and a higher percentage of uninsured people. Variables associated with increased CFR included a higher percentage of the population over age 65, a higher percentage of Black or African Americans, a higher asthma prevalence, and a greater number of hospitals in a county. By identifying factors that are associated with COVID-19 CFR in US counties, we hope to help officials target public health interventions and healthcare resources to locations that are at increased risk of COVID-19 fatalities.

## Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) originated in Wuhan, China in November 2019 and has since spread to 210 countries worldwide [1]. By the last day considered for inclusion in the present work, June 12th, 2020, SARS-CoV-2 had caused over 2 million Coronavirus Disease 2019 (COVID-19) cases and 114,753 deaths in the United States (US) [2, 3]. The excess mortality from COVID-19 is likely to be underestimated, and recent

from (cited within the manuscript) [https://github.com/jmillar201/COVID19\\_CFR](https://github.com/jmillar201/COVID19_CFR).

**Funding:** This study was supported by the National Science Foundation in the form of a grant awarded to JAM (DGE-1256260) and the National Institute of Child Health and Human Development in the form of grants awarded to DHT (1F32HD093185-03) and MSM (T32HD040128). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

work estimates 912,345 deaths from COVID-19 between March 2020 and May 2021, compared to the officially reported 578,555 deaths [4]. Early work on COVID-19 has highlighted patient characteristics that increase an individual's risk of death [5, 6], however it is unclear to which individual risk factors are best suited to understanding which populations are most at risk of high fatality rates from COVID-19. The distribution of infected cases and fatalities in the US has been heterogeneous across counties [7], and identification of sub-populations at risk of increased morbidity and mortality remains crucial to effective response efforts by federal, state, and local governments [8]. Counties where governing officials are aware that their populations are at a higher risk of COVID-19 mortality, meaning the population experiences a higher case-fatality rate, may opt to tailor state policies or take earlier action to curtail the spread of SARS-CoV-2. Additionally, the federal government may opt to target vaccine resources to counties experiencing higher COVID-19 mortality rates.

The case-fatality rate (CFR) is defined as the number of deaths divided by the total number of confirmed cases from a given disease [9]. When a disease is non-endemic, the CFR fluctuates over time. During the beginning of an epidemic, there is often a lag when counting the number of deaths compared to cases and hospitalizations, leading to an underestimation of the CFR. Furthermore, CFR will fluctuate rapidly early in an epidemic when each additional case or death has an excessive impact on calculating CFR. It is important to not only account for the lag between cases and deaths (i.e., lag-adjusted CFR), but also to ensure that the CFR is no longer fluctuating.

In this study, our objective was to use a lag-adjusted CFR to conduct a county-level mortality risk factor analysis of demographic, socioeconomic, and health-related variables in the US during the first wave of the COVID-19 pandemic (March 28, 2020 to June 12, 2020). This will provide critical information on what population characteristics are most informative to identify counties at high risk of experiencing high COVID-19 mortality rates. We expand upon prior work by considering possible risk factors of an increased CFR from multiple categories (e.g., non-pharmaceutical interventions such as shelter-in-place orders [10], prevalence of pre-existing conditions such as cardiovascular disease [11], and socio-economic circumstances such as hospital accessibility [12]) in a single model. This is also the first paper to focus on this range of risk factors during the first wave of the pandemic, so that results from this can be used for targeted intervention at the county level at the beginning of a pandemic.

## Methods

All code for our work can be found on our GitHub repository [13].

### Study population

We conducted a cross-sectional ecological study to assess risk factors associated with an increased COVID-19 lag-adjusted CFR in US counties. Our study population included 3,004 counties or county-equivalents with Federal Information Processing Standards (FIPS), a unique code for US federal identification (Appendix A1 in [S1 Appendix](#)). Only publicly available aggregate data were used; therefore, no IRB approval was required.

### County-level variables

We identified potential risk factors across several different categories: demographic, socioeconomic, healthcare accessibility, comorbidity prevalence, and non-pharmaceutical interventions. Each category-targeted risk factor relevant to the risk of COVID-19 mortality by conducting a comprehensive review of existing literature by March 28, 2020 supplemented with variables relevant to other respiratory epidemics [14–16]. Appendix A2 in [S1 Appendix](#)

provides detailed justifications for the inclusion of each risk factor. Only variables with publicly available data sources at the county- or state-level were included. Appendix A3 in [S1 Appendix](#) listed data sources, variable descriptions, and manipulations (if applicable). We directly imported and cleaned the datasets using R (v3.6.3).

We included five demographic variables: total population, population density, the percentage of the population over age 65, the percentage of population 17 or younger, and race/ethnicity. All demographic variable data were from the 2018 American Community Survey 5-Year Data from the US Census annual survey, except for race/ethnicity data from the U.S. Census Populations with Bridged Race Categories [17].

We included 13 socioeconomic variables, with their data primarily from the 2018 American Community Survey 5-Year Data [18]. In addition to the commonly used socioeconomic variables, we included certain variables contributing to the composite Social Vulnerability Index (SVI). The SVI was created by the Centers for Disease Control and Prevention (CDC) to describe US geographic areas by their social vulnerability and has been validated by multiple studies within and outside of the CDC [19–24]. Social vulnerability is defined as “the characteristics of a person or community that affect their capacity to anticipate, confront, repair, and recover from the effects of a disaster.” [19] We included individual SVI variables on socioeconomic status, household composition and disability, minority status and language, and housing and transportation. We preferred to use the individual variables rather than overall SVI or by theme because we were most interested in understanding which components of social vulnerability contributed to increased CFR.

We included 5 healthcare-related variables: number of hospitals per capita, number of ICU beds per capita, number of primary care physicians per capita, percentage of residents without health insurance, and percentage of Medicaid eligible residents. Variable data was from the Kaiser Health News [25], the Heart Disease and Stroke Atlas [26], and the 2018 American Community Survey 5-Year Data [18].

We included 18 comorbidity variables: diagnosed diabetes prevalence; diagnosed obesity prevalence; hypertension hospitalization and death prevalence, cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD), asthma, and cancer; Medicare beneficiaries with heart disease percentage, current smokers prevalence, and stroke-related hospitalization and mortality prevalence. Variable data was from the US Diabetes Surveillance System [27], the Heart Disease and Stroke Atlas [26], the Behavioral Risk Factor Surveillance System [28], and the State Cancer Profiles by the National Cancer Institute [29].

Non-pharmaceutical intervention data (including information on closing of public venues such as restaurants, gathering size limits, complete lockdown of non-essential activity in the county, if religious gatherings were included in gathering size limits, shelter-in-place orders, and social distancing mandates) were extracted from the COVID-19-intervention GitHub page, an open source data-sharing platform and compiled by Keystone Strategy [30]. However, this resource does not cover all counties, thus missing data was supplemented from a variety of governmental executive orders and news articles detailed in the supplementary code. Variables with dates were transformed to how many days the event occurred after the first case in a county. States where an intervention never occurred were given a zero. Since 47% of all counties did not ban religious gatherings, data on when religious gatherings were banned in a county was transformed into an indicator variable (1 if the ban occurred, 0 if not).

### Lag adjusted case-fatality rate (CFR) data and calculation

To calculate CFR during the first COVID-19 wave in the US, we obtained open access county-level COVID-19 data from the New York Times through June 12, 2020, the date the CDC

released guidance for easing restrictions as states began to reopen [2, 31]. Only data that contained FIPS county codes to identify case and death locations were included. County-level data for New York City, NY was accessed from the New York City Department of Health and Mental Hygiene [32]. To calculate lag-adjusted CFR (laCFR), we used Nishiura et al.'s method, expanded upon by Russell et al., to account for the delay between COVID-19 diagnoses and deaths [33, 34]. We updated this approach by using time-from-hospitalization-to-death from the US population [34, 35]. The final dataset included 1,779 counties with 1,968,739 cases and 106,279 deaths, comprising 96.8% of national cases and 96.8% of national deaths as of June 12, 2020.

During the first wave of the pandemic, SARS-CoV-2 was non-endemic, leading the case-fatality rate (CFR) to fluctuate over time. This is due to a lag when counting the number of deaths compared to cases and hospitalizations, leading to an underestimation of the CFR. The CFR continues to fluctuate rapidly early in an epidemic when each additional case or death has an excessive impact on calculating CFR. It is important to not only account for the lag between cases and deaths (i.e., lag-adjusted CFR), but also to ensure that the CFR is no longer fluctuating.

To do this, we use a method developed by Nishiura et al. and expanded upon by Russell et al., where case and death incidence data are used to estimate the number of cases with known outcomes, i.e. cases where the resolution, death or recovery, is known to have occurred [33, 34]:

$$u_t = \frac{\sum_{i=0}^t \sum_{j=0}^{\infty} c_{i-j} f_j}{\sum_{i=0}^t c_i}$$

where  $c_t$  is the daily case incidence at time  $t$ , (with time measured in calendar days),  $f_t$  is the proportion of cases with delay  $t$  between onset or hospitalization and death;  $u_t$  represents the underestimation of the known outcomes and is used to scale the value of the cumulative number of cases in the denominator in the calculation of the laCFR. Russell et al. used the estimated distribution in Linton et al., based on data from China up until the end of January 2020. For this study, we instead used United States centric data from Lewnard et al., which estimates the distribution of time from hospitalization to death based on data from Washington and California [35].

Lewnard et al., fits the distribution conditionally on age resulting in a Weibull distribution for each age group [35]. The overall distribution was obtained empirically by weighting the densities at time  $t$  across all age groups. Because of this, the overall distribution doesn't have its own shape/scale parameters. However, we were able to estimate what these parameters would be by fitting a Weibull distribution that captures the 2.5, 25, 50, 75, and 97.5 percentiles (1.6, 7.3, 12.7, 19.8, 37.4), as well as the average (14.5).

Use of the laCFR assumes the measure has stabilized [33]. Counties where the laCFR is still rapidly changing cannot be used in the study as these are not unbiased estimates of the true CFR. laCFRs were calculated incrementally for each day and assessed whether they changed on average less than 1% a week for the last two weeks of available data. The final calculation based on all data available was used as the laCFR in our model.

## Statistical analysis

To reduce multicollinearity, we eliminated linear combinations and variables with correlations  $>0.5$  using the R package *caret* (v6.0.86). Remaining variables were screened for missingness and missing values were imputed using five imputations in the R package *mice* (v3.8.0) [36].

Data were randomly split into training (1,186 counties) and testing sets (593 counties) to assess generalizability (a table of the characteristics can be seen in Appendix A4 in [S1 Appendix](#)). A negative binomial linear model with an offset for the number of COVID19 cases per county was chosen based on Kolmogorov-Smirnov and dispersion tests found in the R package DHARMA (v0.3.1). Variable selection was conducted using purposeful selection, an iterative process in which covariates are removed from the model if they are neither significant nor confounders [37, 38]. With clinical risk factors, purposeful selection outperforms other variable selection procedures and tests for the presence of confounders [37]. Removing highly correlated variables beforehand reduces the chance of multicollinearity between non-significant variables that may have been retained in purposeful selection due to confounding effects. Per Bursac et al., we used the 0.1  $\alpha$ -level for initial selection using bivariate models and a change of >20% in any remaining model coefficients compared with the full multivariate model for confounding evaluation [37]. All variables in the final model were significant at the 0.05  $\alpha$ -level, and no statistical confounders were included in the final model.

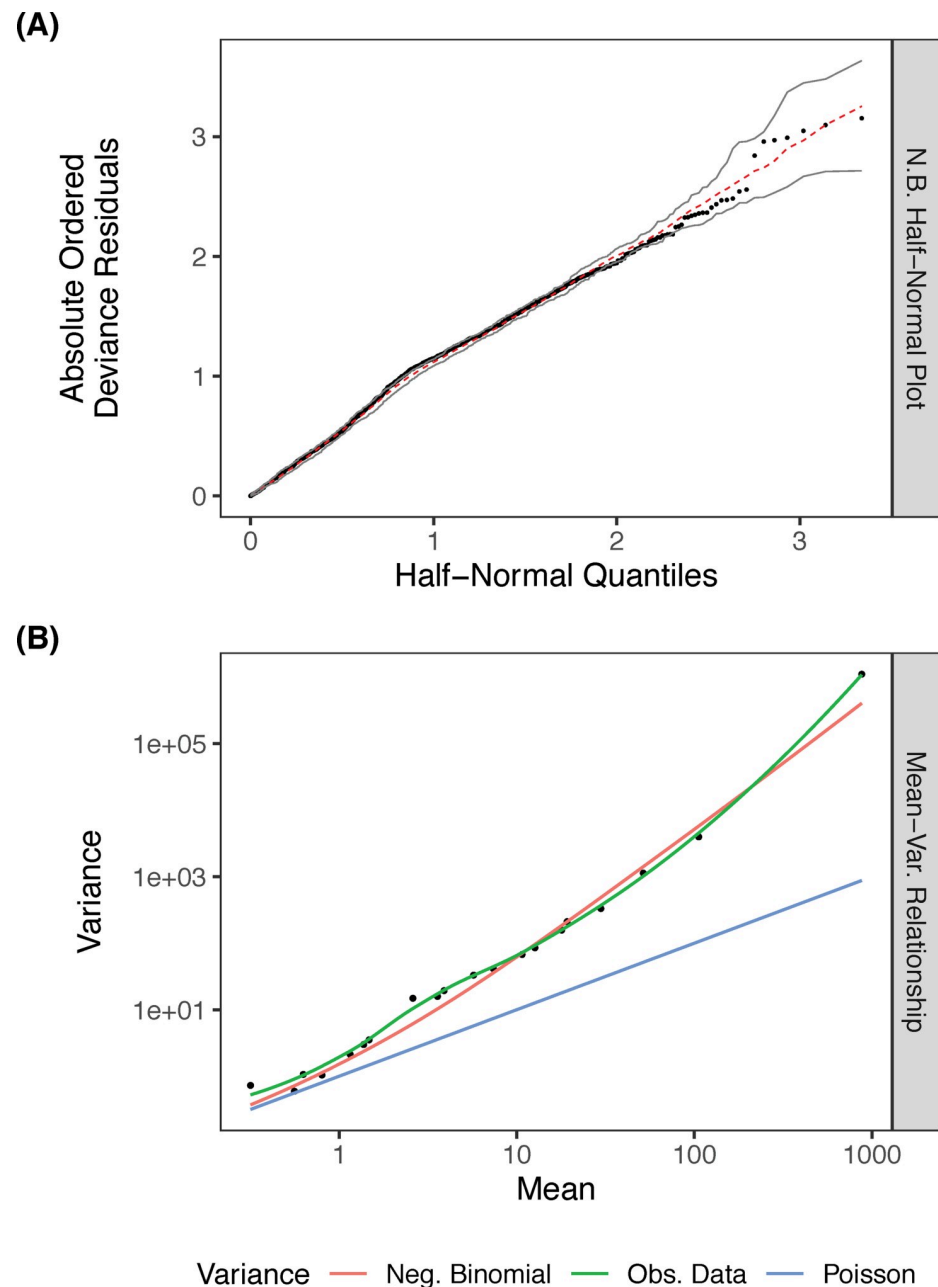
### Model fit

We observed the fit of the model using a half-normal plot ([Fig 1A](#)). The simulated envelope for the deviance residuals in the half-normal plot serves as a guide of what to expect under a well-fitted model, with most of our model's deviance residuals lying within [39]. We compared the mean and variance seen within our model predictions to the theoretical mean and variance expected in a Poisson and negative binomial model. After grouping the fitted predictions into 20 quantiles and calculating their means and variances, we saw the negative binomial model captures our data variance well [40]. Loess smooth was used for the empirical mean ([Fig 1A](#)). As an additional check, we calculated the ratio of Pearson residuals to degrees of freedom, which was 1.04, indicating we accounted for most of the over-dispersion in laCFR using the negative binomial model. The Cox and Snell Pseudo R2 for our model was 0.86, which accounts for the majority of the variance present in our outcome variable. All variables had a variance inflation factor of less than 2, indicating collinearity was not an issue with our variables (Appendix 3 in [S1 Appendix](#)).

We checked the coverage, which is the probability that our model outcomes are found within our prediction interval. To estimate our predictive coverage (empirical coverage), we simulated a prediction interval. The coverage was 0.9730 for the training data and 0.9713 for testing data ([Fig 2A](#)). Similar to ROC, a gain curve plot measures how well the model score sorts the data compared to the true outcome value [41]. When the predictions sort in exactly the same order, the relative Gini coefficient is 1. When the model sorts poorly, the relative Gini coefficient is close to zero, or even negative. The relative Gini scores were high for both our training set and testing set. (0.9840 and 0.9829, respectively, [Fig 2B](#)).

### Results

Of the 64 variables collected, 22 were retained for analysis after minimizing correlation (Appendix A2-A5 in [S1 Appendix](#)). Multiple imputation was used to correct for missingness (less than 2%) in two of the retained variables, neither of which appeared in the final model. Fifteen variables were significant in bivariate models in the first step of purposeful selection, and were included in the initial multivariate model. Eight variables were significant in the initial multivariate model and were retained in the final model. Including variables that were non-significant in the bivariate models with these eight variables did not significantly change the performance of the model, as determined by the Likelihood Ratio Test. No potential

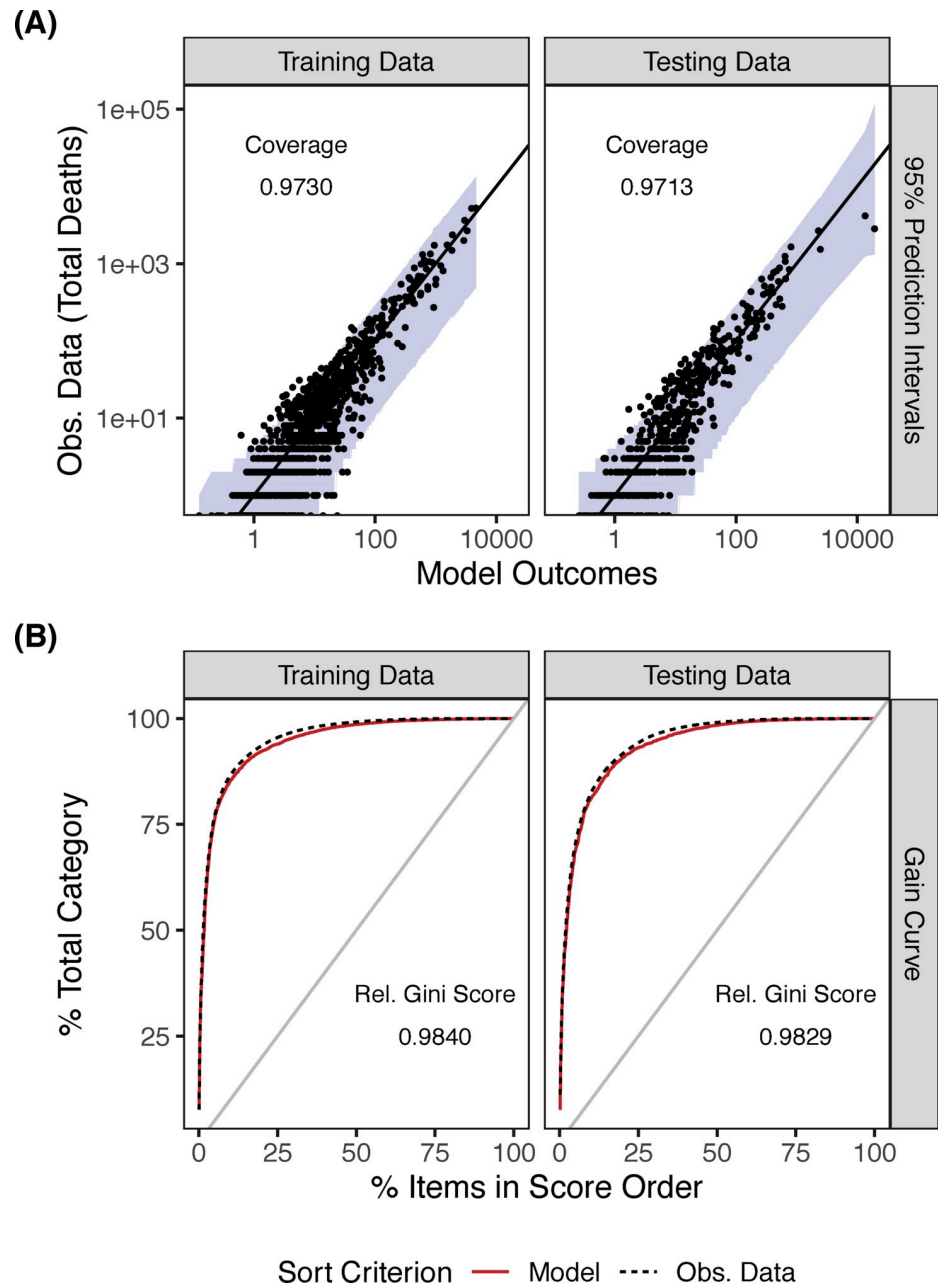


**Fig 1. Assessing model fit.** Plots showing (A) half-normal residuals and (B) mean-variance relationship of the observed county-level COVID-19 laCFRs.

<https://doi.org/10.1371/journal.pone.0258308.g001>

confounders were identified among the correlation minimized variables that were previously discarded due to non-significance in the models.

The final model is shown in Table 1. The negative binomial model appears to be a good fit, capturing the mean-variance relationship observed in the data and displaying expected residuals (Fig 1A and 1B). The model was well-calibrated, with the training and testing model having comparable coverage and relative Gini score (Fig 2A and 2B). Since we used a negative binomial model with an offset, the exponentiated coefficients represent the change in laCFR observed for a one-unit increase in each continuous variable, assuming all other variables in the model are



**Fig 2. Assessing model generality.** Plots showing (A) model outcomes found within the prediction intervals for training data and testing data for the county-level COVID-19 laCFRs and (B) gain curves for training data and testing data for the county-level COVID-19 laCFRs.

<https://doi.org/10.1371/journal.pone.0258308.g002>

held constant. Four variables were inversely associated with laCFR: number of hospitals per 10,000 people (-32% laCFR per additional hospital per 10,000), banning religious gatherings during the initial state or county shutdown (-13% laCFR if religious gatherings were banned), percentage of housing units that were mobile homes (-0.79% laCFR per 1% increase in the proportion of mobile homes), and percentage of population without health insurance (-1.5% laCFR per 1% increase in percentage uninsured). Four variables were directly associated with laCFR: percentage over age 65 (+4.5% laCFR per 1% increase in population over age 65), percentage

**Table 1. Parameter estimates for the final multivariate model of laCFR.**

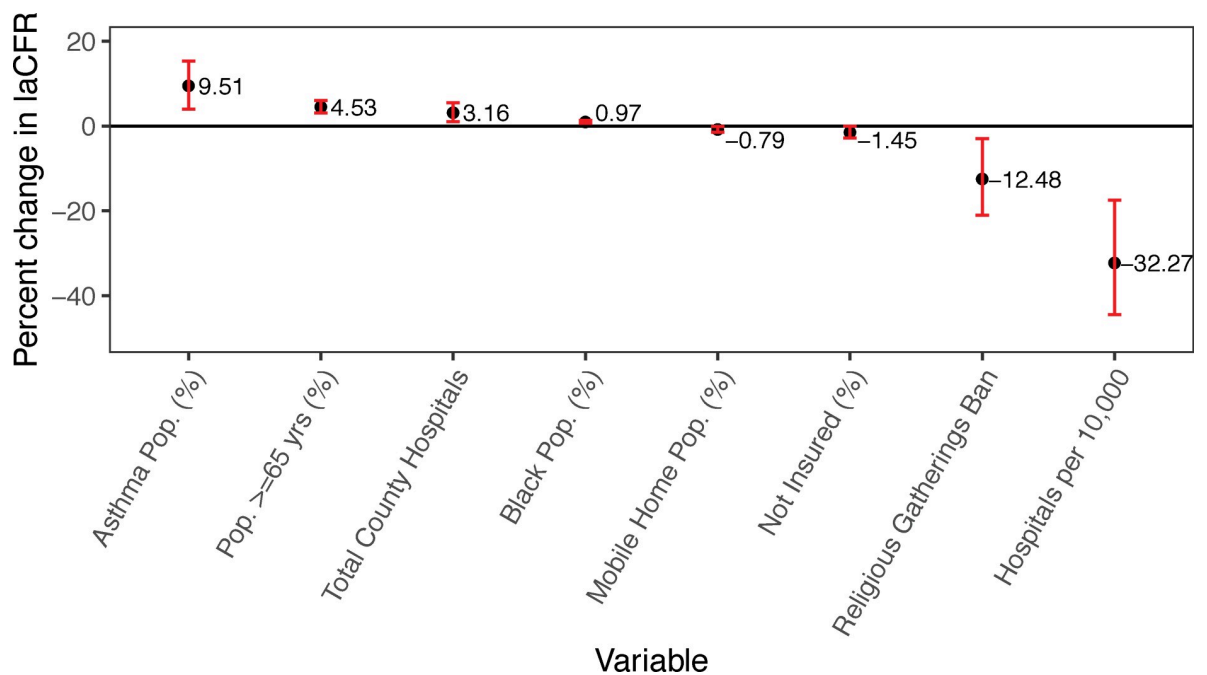
Variable	Coefficient	95% CI	p-value
Intercept	-4.5	(-5.1, -3.9)	<0.001
Percentage population aged 65+	0.044	(0.030, 0.059)	<0.001
Percentage population Black or African American	0.0097	(0.0063, 0.013)	<0.001
Hospitals per 10,000 persons	-0.39	(-0.59, -0.19)	<0.001
Asthma prevalence	0.091	(0.039, 0.14)	<0.001
Total number of hospitals	0.031	(0.0099, 0.054)	0.0017
Ban on religious gatherings indicator	-0.13	(-0.24, -0.030)	0.011
Percentage housing stock that were mobile homes	-0.0079	(-0.015, -0.0011)	0.024
Percentage population without health insurance	-0.015	(-0.029, -0.00021)	0.052

<https://doi.org/10.1371/journal.pone.0258308.t001>

Black or African American (BAA) (+0.97% laCFR per 1% increase in BAA population), percentage with asthma (+9.5% laCFR per 1% increase in asthma prevalence), and number of hospitals (+3.2% laCFR per one additional hospital). Fig 3 demonstrates the relationship between each variable and the laCFR over a range of values. We have stratified these variables further for comparison, and results can be found in Appendix A6 in S1 Appendix.

### Discussion

In this ecological study of mortality due to SARS-CoV-2 infection during the first wave of COVID-19 in the US, we found that county-level laCFR was significantly associated with eight variables. Four variables—banning religious gathering, proportion of mobile homes, hospitals per 10,000 persons, and proportion of uninsured individuals in a county—were associated with decreased laCFR. Four variables—percentage of population over age 65, total number of



**Fig 3. Percentage change in COVID-19 laCFR given a 1 unit increase in the variable for each individual variable (shown in black dots) and 95% confidence interval (shown in red), using training data.**

<https://doi.org/10.1371/journal.pone.0258308.g003>



hospitals per county, prevalence of asthma, and percentage of population BAA—were associated with increased laCFR. Each variable provided unique insights into factors that may be worth considering for county-level COVID-19 response efforts.

### Inverse association with case-fatality rate

Our model indicated a 13% reduction in the average laCFR for counties that banned religious gatherings compared to counties that did not. Gatherings often involve dense mixing of people in a confined space, sometimes over long periods of time [42], which drives COVID-19 transmission [43]. Interventions targeting increased physical distancing and limiting contact were introduced in some countries, including the closure of schools, places of worship, malls, and offices [42]. Our model suggests that specifically exempting religious gatherings from bans may increase the laCFR, consistent with the combination of findings that [1] religious gatherings across the globe were linked to COVID-19 superspreader events [44] and [2] older Americans (who are more likely to attend religious services than younger Americans [45]) are at increased risk of death due to COVID-19.

The percentage of the population living in mobile homes was also associated with a decrease in laCFR. A 1% increase in mobile home living was associated with a 0.79% decrease in laCFR. While a small difference at first glance, it becomes more meaningful when considering the large variation in mobile home living across counties. Between counties at the 25th percentile of percentage living in mobile homes (4%) and 75th percentile (18%), the difference in the percentage of mobile home living correlated with an 11% decrease in laCFR. This might represent a built-environment effect, given that mobile homes have separate plumbing and ventilation unlike apartments and other multi-family residences. Recent work suggests that fecal aerosol transmission of SARS-CoV2 can occur [46]. Ventilation patterns in apartment complexes represent additional opportunities for transmission [47]. The benefit of separate units such as mobile homes may be especially important to low-income workers who are both more at risk of death from COVID19 due to increased chance of having a co-morbid condition and more likely to live in multi-family housing with maintenance issues [48].

The number of hospitals per 10,000 was also inversely associated with laCFR. We found that for each additional hospital per 10,000 inhabitants, the laCFR decreased by 32%, despite the exclusion of other healthcare-related variables due to non-significance (e.g., ICU bed availability). Prior work demonstrated that the percentage of ICU and non-ICU beds occupied by COVID-19 patients directly correlated with COVID-19 deaths [49] and a county with more hospitals per 10,000 inhabitants may be able to cope with more COVID-19 cases before reaching the same percentage of hospitals beds occupied as a county with fewer hospitals per 10,000 inhabitants. Furthermore, because adding beds requires fewer resources than adding hospitals, the number of hospitals per 10,000 persons in a county might represent a greater ability to expand capacity. As a result, using hospitals per 10,000 may be a better indicator of healthcare capacity than the number of ICU beds early on in the pandemic. Because healthcare resources in the US correlate with community wealth [50], the rate of hospitals per 10,000 may also reflect increased community wealth and the protective effect of higher socio-economic status on health. More hospitals per 10,000 persons may also represent increased competition for patients, which is associated with decreased mortality from community-acquired pneumonia [51].

Unexpectedly, the percentage uninsured was inversely associated with laCFR. We found a 1.5% reduction in laCFR for every 1% increase in uninsured inhabitants. Prior studies found longer travel times to COVID-19 testing facilities were directly associated with percentage uninsured [52, 53]. Because uninsured persons may be unable to readily access testing, this

finding may relate to incomplete reporting, such that only individuals who survive long enough are tested for COVID-19, leading to a potential undercount of deaths attributable to SARS-CoV-2 infection.

### Direct association with case-fatality rate

In our model, a 1% increase in the population over 65 years old was associated with a 4.5% increase in average laCFR. This is consistent with recent epidemiological studies demonstrating an association between the severity of COVID-19 infection and age. According to provisional death data from the National Center of Health Statistics, people aged 65 and older have a 90- to 630-fold higher risk of mortality due to COVID-19 than 18-29-year olds [54].

Also, directly associated with laCFR was the total number of hospitals per county, with an observed increase of 3.2% in average laCFR per additional hospital. This variable was strongly correlated with total population ( $r = 0.92$ ). Given that the number of hospitals per 10,000 was associated with decreased laCFR, this correlation suggests that total hospitals might be a proxy indicator for total population. Previous work assessed population density as a risk factor for increased laCFR, but not total population [43]. Since our analysis focused on the first wave of COVID-19, this variable could reflect overwhelmed healthcare systems in highly populated counties where most of the COVID-19 cases initially occurred [55].

Asthma prevalence was also directly associated with laCFR. A 1% increase in asthma prevalence was associated with a 9.5% increase in laCFR. Evidence regarding asthma as a risk factor in COVID-19 is mixed. Although the US CDC has determined that patients with moderate to severe asthma belong to a high-risk group [56], the Chinese CDC indicated that asthma was not a risk factor for severe COVID-19 [57]. One study showed that COVID-19 patients with asthma were of older age and had an increased prevalence of multiple comorbidities compared to those without asthma [58], but that the presence of asthma alone was not a risk factor for increased mortality [58]. Thus, despite our findings, it is unclear whether asthma has a direct impact on COVID-19 disease or if other factors may be associated with both asthma and COVID-19. One such potential confounder is exposure to air pollution, as air pollution is associated with both asthma and risk of death from COVID-19 [59].

Finally, laCFR was directly associated with the percentage of the population identifying as BAA in a county. Our model showed that a 1% increase in BAA was associated with a 0.97% increase in the laCFR. This likely reflects the effects of structural racism in the US, where BAAs have fewer economic and educational opportunities than White Americans and as a result are exposed to increased risk of morbidity and mortality from COVID-19 [60]. Dalsania et al. also found that the social determinants of health contributed to an unequal impact of the COVID-19 pandemic for BAA at the county level [61]. A study by Golestaneh showed that US counties with BAA as the majority had three times the rate of infection and almost six times the rate of death as majority White counties [62]. Factors underlying this trend include years of structural racism resulting in a lack of financial resources, increased reliance on public transportation, housing instability, and dependence on low-paying retail jobs [63]. Our approach considered several other variables that might explain the effect but were either non-significant (e.g., household crowding, percentage of households without a vehicle, and county land area) or were correlated with percentage BAA (e.g. percentage single parent households and percentage living in poverty), further emphasizing the role of systemic racism in COVID-19 laCFR.

### Excluded predictor variables

In reducing multicollinearity and using purposeful selection, several variables were surprisingly excluded. One of these excluded variables was population density, although higher

population density had been hypothesized to increase contact rate and non-adherence with physical distancing [43]. Diabetes and cardiovascular disease were excluded, despite multiple studies reporting these conditions as risk factors for COVID-19 mortality [64, 65]. While these factors are important at an individual level to assess the mortality risk, our model suggests that other variables may be more informative at the county-level, underscoring the value of ecological studies.

### Study strengths and limitations

This study had several strengths besides the benefits of an ecological design when considering population interventions. First, the data were nationally representative, including over 50% of all US counties. Our model captured the variability in the data and accounted for the observed data distribution. The model also captured almost all outcomes within the prediction interval for both training and testing data sets, with similar accuracy between them, which indicates that our model is generalizable within the US. Additionally, our model based laCFR calculations on the distribution of times from hospitalization to death from US data [35], which differed from earlier Chinese data [57]. Using US-based distribution of times likely improved our laCFR estimation for this study. The final model included several variables previously attributed to higher laCFR (such as older age) [54] and included a variable unique to the pandemic shutdown, i.e., banning religious gatherings, giving more nuanced insights into heterogeneous COVID-19 mortality rates across counties.

Despite these strengths, our study had several limitations. First, under-reporting of cases might affect the accuracy of CFR calculation [66]. The reported cases and deaths we used likely underestimated the true COVID-19 parameters. This underestimation was more among the asymptomatic and mild cases due to limited testing capacity and changes in testing practice; hence, the laCFR might have appeared inflated. Second, the type and timing of the tests used may have impacted the measured laCFR. Samples collected early during the infection can yield higher false negatives with RT-PCR tests [67]. False negatives in critically-ill patients who later die could decrease the measured laCFR unless probable COVID-19 deaths are reported, while false negatives in mild cases who are not retested later could increase the measured laCFR as survivable cases go undetected. These are challenges for any CFR study and highlight the ongoing need for improved COVID-19 testing. Third, COVID-19 reporting practices vary widely by state. For example, Florida was found to report fewer COVID-19 deaths in the official tally than the Medical Examiners Commission [68]. In addition to deliberate underreporting of deaths, states also vary in reporting of probable cases and deaths [69]. Without national standards in the COVID-19 response, comparing case counts and deaths across state–let alone county–is deterred by lack of clarity about how these data differ [69].

Beyond these, our study was also limited by the fact that relevant data were frequently unavailable, including data on non-pharmaceutical interventions (NPI) and comorbidities. To limit missingness in the NPI data, we used state-level data when available given that counties also enforce state-level orders. However, there may be heterogeneity between county- and state-level information making this a less effective approach. Other variables of interest were not available at the state- or county-level, including information on contact tracing efforts and community compliance with public health mandates. Funding to collect public health information on more variables at a granular level would improve the information available to guide decision-making during emergencies. Another limitation was the highly correlated nature of the 64 variables considered for inclusion. Multicollinearity greatly affects the interpretability of coefficients and is rarely accounted for in epidemiologic studies [70]. Highly correlated variables in a model are unstable and can bias standard errors, leading to unreliable p-values and

unrealistic interpretations [70]. Because we ensured our model interpretability by excluding highly correlated variables, not all of our collected 64 variables were screened for inclusion in the final model.

Finally, our study period ended in mid-June. Recent work has divided the COVID-19 pandemic in the USA into three waves, with the first wave running from late March 2020 until mid-June 2020 [71]. The exact day of June 12, 2020 was chosen because [1] enough cases had occurred in the US to obtain reliable estimates of I<sub>a</sub>CFR by county and [2] it preceded CDC reopening guidance and a shift in reporting to the HHS Protect system, which is less readily available to the public than the prior CDC reporting system [72]. The decision by the government to switch to the HHS Protect system hinders the ability of academic scientists to aid in the response to the on-going pandemic [72]. Making these data more readily available to the public would permit inclusion of additional data for future research.

## Conclusion

This study highlights several variables that were associated with county-level I<sub>a</sub>CFR during the first wave of COVID-19 in the US. Though further research is needed to examine the effects of additional NPIs, our work provides insights that may aid in targeting response and vaccination efforts for improved outcomes in subsequent waves.

## Supporting information

**S1 Appendix.**  
(PDF)

## Acknowledgments

This project is part of the COVID-19 Dispersed Volunteer Research Network (COVID-19-DVRN) led by M. Majumder and A. Desai. The authors thank Marie Charpignon, Catherine Pollack, and Emily Ricotta for their thoughtful feedback and review of the manuscript.

## Author Contributions

**Conceptualization:** Jess A. Millar, Hanh Dung N. Dao, Marianne E. Stefopoulos, Camila G. Estevam, Katharine Fagan-Garcia, Diana H. Taft, Christopher Park, Angel N. Desai, Maimuna S. Majumder.

**Data curation:** Jess A. Millar, Hanh Dung N. Dao, Marianne E. Stefopoulos, Camila G. Estevam, Katharine Fagan-Garcia, Diana H. Taft, Christopher Park, Amaal Alruwaily.

**Formal analysis:** Jess A. Millar.

**Investigation:** Jess A. Millar, Hanh Dung N. Dao, Marianne E. Stefopoulos, Camila G. Estevam, Katharine Fagan-Garcia, Diana H. Taft, Christopher Park, Amaal Alruwaily, Angel N. Desai, Maimuna S. Majumder.

**Methodology:** Jess A. Millar, Hanh Dung N. Dao.

**Project administration:** Diana H. Taft.

**Software:** Jess A. Millar, Hanh Dung N. Dao.

**Supervision:** Angel N. Desai, Maimuna S. Majumder.

**Validation:** Jess A. Millar.

**Visualization:** Jess A. Millar.

**Writing – original draft:** Jess A. Millar, Hanh Dung N. Dao, Marianne E. Stefopoulos, Camila G. Estevam, Katharine Fagan-Garcia, Diana H. Taft, Christopher Park, Amaal Alruwaily.

**Writing – review & editing:** Jess A. Millar, Hanh Dung N. Dao, Marianne E. Stefopoulos, Camila G. Estevam, Katharine Fagan-Garcia, Diana H. Taft, Christopher Park, Amaal Alruwaily, Angel N. Desai, Maimuna S. Majumder.

## References

1. Ortiz-Prado E, Simbaña-Rivera K, Gómez-Barreno L, Rubio-Neira M, Guaman LP, Kyriakidis NC, et al. Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. *Diagn Microbiol Infect Dis*. 2020; 98(1): 115094. <https://doi.org/10.1016/j.diagmicrobio.2020.115094> PMID: 32623267
2. GitHub—nytimes/covid-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S.; 2020 [cited 2020 Aug 31] Repository: GitHub [Internet]. Available from: <https://web.archive.org/web/20200901034011/https://github.com/nytimes/covid-19-data>
3. COVID-19 Map—Johns Hopkins Coronavirus Resource Center. 2020 [cited 2 Sep 2020]. In: Johns Hopkins University [Internet]. Available from: <https://web.archive.org/web/20200831235220/https://coronavirus.jhu.edu/map.html>
4. Estimation of total mortality due to COVID-19 [cited 19 May 2021]. In: Institute for Health Metrics and Evaluation [Internet]. Available from: <http://www.healthdata.org/special-analysis/estimation-excess-mortality-due-covid-19-and-scalars-reported-covid-19-deaths>
5. Cao J, Tu W-J, Cheng W, Yu L, Liu Y-K, Hu X, et al. Clinical features and short-term outcomes of 102 patients with Corona Virus Disease 2019 in Wuhan, China. *Clin Infect Dis*. 2020; 71(15): 748–755. <https://doi.org/10.1093/cid/ciaa243> PMID: 32239127
6. Tu W-J, Cao J, Yu L, Hu X, Liu Q. Clinicolaboratory study of 25 fatal cases of COVID-19 in Wuhan. *Intensive Care Med*. 2020; 46(6): 1117–1120. <https://doi.org/10.1007/s00134-020-06023-4> PMID: 32253448
7. Basu A. Estimating the infection fatality rate among symptomatic COVID-19 cases in the United States. *Health Aff*. 2020; 39(7): 1229–1236. <https://doi.org/10.1377/hlthaff.2020.00455> PMID: 32379502
8. Hutchins SS, Truman BI, Merlin TL, Redd SC. Protecting vulnerable populations from pandemic influenza in the United States: A strategic imperative. *Am J Public Health*. 2009; 99 Suppl 2: S243–S438. <https://doi.org/10.2105/AJPH.2009.164814> PMID: 19797737
9. Celentano DD, Mhs S, Szklo M. Gordis. *Epidemiología*. 6th ed. Philadelphia: Elsevier; 2019.
10. Lyu W, Wehby GL. Shelter-in-place orders reduced COVID-19 mortality and reduced the rate of growth in hospitalizations. *Health Aff*. 2020; 39(9): 1615–1623. <https://doi.org/10.1377/hlthaff.2020.00719> PMID: 32644825
11. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: A systematic review and meta-analysis. *Int J Infect Dis*. 2020; 94: 91–95. <https://doi.org/10.1016/j.ijid.2020.03.017> PMID: 32173574
12. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob Health*. 2020; 8(4): e480. [https://doi.org/10.1016/S2214-109X\(20\)30068-1](https://doi.org/10.1016/S2214-109X(20)30068-1) PMID: 32109372
13. Millar J. GitHub—jmillar201. COVID19\_CFR: Repository of data and code for calculation COVID-19 county level CFRs; 2020 [cited 2020 Nov 13] Repository: GitHub [Internet] Available from: [https://web.archive.org/web/20210922153139/https://github.com/jmillar201/COVID19\\_CFR](https://web.archive.org/web/20210922153139/https://github.com/jmillar201/COVID19_CFR)
14. Nguyen-Van-Tam JS, Openshaw PJM, Hashim A, Gadd EM, Lim WS, Semple MG, et al. Risk factors for hospitalization and poor outcome with pandemic A/H1N1 influenza: United Kingdom first wave (May–September 2009). *Thorax*. 2010; 65(7): 645–651. <https://doi.org/10.1136/thx.2010.135210> PMID: 20627925
15. Chan-Yeung M, Xu R-H. SARS: Epidemiology. *Respirology*. 2003; 8 Suppl: S9–S14. <https://doi.org/10.1046/j.1440-1843.2003.00518.x> PMID: 15018127
16. Park J-E, Jung S, Kim A, Park J-E. MERS transmission and risk factors: A systematic review. *BMC Public Health*. 2018; 18(1): 574. <https://doi.org/10.1186/s12889-018-5484-8> PMID: 29716568
17. National Center for Health Statistics. Vintage 2018 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010–July 1, 2018), by year, county, single-year of age (0, 1, 2, . . . , 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative

- arrangement with the U.S. Census Bureau. [Internet]. 2019 [cited 2020 Mar 29]. Available from: [https://web.archive.org/web/20200903035551/https://www.cdc.gov/nchs/nvss/bridged\\_race.htm](https://web.archive.org/web/20200903035551/https://www.cdc.gov/nchs/nvss/bridged_race.htm).
18. United States Census Bureau. American Community Survey 5-Year Data (2009–2018) [Internet]. 2019 [cited 2020 May 9]. Available from: <https://web.archive.org/web/20200904163926/https://www.census.gov/data/developers/data-sets/acs-5year.html>.
  19. Flanagan BE, Hallisey EJ, Adams E, Lavery A. Measuring community vulnerability to natural and anthropogenic hazards: The centers for disease control and prevention's social vulnerability index. *J Environ Health*. 2018 Jun; 80(10):34–6. Available from: [https://web.archive.org/web/20210330172523/https://svi.cdc.gov/Documents/Publications/CDC\\_ATSDR\\_SVI\\_Materials/JEH2018.pdf](https://web.archive.org/web/20210330172523/https://svi.cdc.gov/Documents/Publications/CDC_ATSDR_SVI_Materials/JEH2018.pdf) PMID: 32327766
  20. Flanagan BE, Gregory EW, Hallisey EJ, Heitgerd JL, Lewis B. A social vulnerability index for disaster management. *J Homel Secur Emerg Manag*. 2011 Jan 5; 8(1):3. <https://doi.org/10.2202/1547-7355.1792>
  21. Adams E. Social vulnerability and disaster-related health outcomes. Poster presented at: Esri User Conference; 2016 Nov 18 [cited 2020 Nov 15]; San Diego, CA. Available from: [https://web.archive.org/web/20210922151810/https://stacks.cdc.gov/view/cdc/45924/cdc\\_45924\\_DS1.pdf](https://web.archive.org/web/20210922151810/https://stacks.cdc.gov/view/cdc/45924/cdc_45924_DS1.pdf).
  22. Lavery A. Mapping mortalities following Hurricane Harvey, Harris County, TX, August–September 2017 [Internet]. Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry GIS Day; 2017 Nov 15 [cited 2020 Sep 26]; Atlanta, GA. Available from: [https://web.archive.org/web/20210326131259/https://www.cdc.gov/gis/docs/Full\\_Agenda\\_2017.pdf](https://web.archive.org/web/20210326131259/https://www.cdc.gov/gis/docs/Full_Agenda_2017.pdf).
  23. Kolling J, Wilt G, Berens A, Strosnider H, Devine O. Social and environmental risk factors associated with county-level asthma emergency department visits [Internet]. American Public Health Association Annual Meeting; 2017 Nov [cited 2020 Sep 26]; Atlanta, GA. Available from: [https://web.archive.org/web/20210922151959/https://svi.cdc.gov/Documents/Publications/CDC\\_ATSDR\\_SVI\\_Materials/APHApaperV7\\_TOPPRINT.pdf](https://web.archive.org/web/20210922151959/https://svi.cdc.gov/Documents/Publications/CDC_ATSDR_SVI_Materials/APHApaperV7_TOPPRINT.pdf).
  24. Bakkensen LA, Fox-Lent C, Read LK, Linkov I. Validating resilience and vulnerability indices in the context of natural disasters. *Risk Anal*. 2017; 37(5):982–1004. <https://doi.org/10.1111/risa.12677> PMID: 27577104
  25. Millions of Older Americans Live in Counties with no ICU Beds as Pandemic Intensifies | Kaiser Health News [Internet]. [cited 2020 Aug 31]. Available from: <https://archive.ph/Xapvq>.
  26. Interactive Atlas of Heart Disease and Stroke [Internet]. [cited 2020 Aug 31]. Available from: <https://web.archive.org/web/20201016161641/https://nccd.cdc.gov/DHDSAtlas/?state=County>
  27. U.S. Diabetes Surveillance System [Internet]. [cited 2020 Aug 31]. Available from: <https://web.archive.org/web/20200905140951/https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html>
  28. CDC—BRFSS [Internet]. [cited 2020 Aug 31]. Available from: <https://web.archive.org/web/20200901141556/https://www.cdc.gov/brfss/index.html>
  29. State Cancer Profiles [Internet]. [cited 2020 Aug 31]. Available from: <https://web.archive.org/web/20200919084932/https://statecancerprofiles.cancer.gov/>
  30. Keystone Strategy PA. Covid19-intervention-data [Internet]. GitHub repository. 2020 [cited 2020 Sep 13]. Available from: <https://web.archive.org/web/20201114171705/https://github.com/Keystone-Strategy/covid19-intervention-data>
  31. Galvin G. CDC issues new guidance for Americans as states reopen from Coronavirus closures. *U.S. News and World Report*. *US News and World Report*. 2020 Jun 12 [cited 2020 Dec 19]. Available from: <https://web.archive.org/web/20200831050041/https://www.usnews.com/news/health-news/articles/2020-06-12/cdc-issues-new-guidance-for-americans-as-states-reopen-from-coronavirus-closures>
  32. COVID-19: Data by Borough—NYC Health; 2020 [cited 2020 Aug 31]. Repository: NYC Health [Internet]. Available from: <https://web.archive.org/web/20200831165703/https://www1.nyc.gov/site/doh/covid/covid-19-data-boroughs.page>
  33. Nishiura H, Klippenberg D, Roberts M, Heesterbeek JAP. Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS ONE*. 2009; 4(8): e6852. <https://doi.org/10.1371/journal.pone.0006852> PMID: 19718434
  34. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill*. 2020; 25(12). <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256> PMID: 32234121
  35. Lewnard JA, Liu VX, Jackson ML, Schmidt MA, Jewell BL, Flores JP, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study. *BMJ*. 2020; 369: m1923. <https://doi.org/10.1136/bmj.m1923> PMID: 32444358
  36. Wulf JN. Multiple imputation by chained equations in praxis: Guidelines and review. *Electron J Bus Res*. 2017; 15(1): 41–56.

37. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source Code Biol Med*. 2008; 3: 17. <https://doi.org/10.1186/1751-0473-3-17> PMID: 19087314
38. Chen Y, Millar JA. Machine learning techniques in cancer prognostic modeling and performance assessment. In: Matsui S, Crowley J, editors. *Frontiers of biostatistical methods and applications in clinical oncology*. Singapore: Springer Singapore; 2017. pp. 193–230. [https://doi.org/10.1007/978-981-10-0126-0\\_13](https://doi.org/10.1007/978-981-10-0126-0_13)
39. Friendly M. *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*. Chapman and Hall/CRC; 2015.
40. Moral RA, Hinde J, Demétrio CGB. Half-normal plots and overdispersed models in R: Thehnp package. *J Stat Softw*. 2017; 81(10).
41. Nisbet R, Miner G, Yale K. Model evaluation and enhancement. In: Miner G, Yale K, Nisbet R, editors. *Handbook of statistical analysis and data mining applications*. Elsevier San Diego, CA; 2018. pp. 215–234. <https://doi.org/10.1016/B978-0-12-416632-5.00011-6>
42. Yezli S, Khan A. COVID-19 pandemic: It is time to temporarily close places of worship and to suspend religious gatherings. *J Travel Med*. 2021; 28(2): taaa065. <https://doi.org/10.1093/jtm/taaa065> PMID: 32339236
43. Rocklöv J, Sjödin H. High population densities catalyze the spread of COVID-19. *J Travel Med*. 2020; 27(3): taaa038. <https://doi.org/10.1093/jtm/taaa038> PMID: 32227186
44. Saidan MN, Shbool MA, Arabeyyat OS, Al-Shihabi ST, Abdallat YA, Barghash MA, et al. Estimation of the probable outbreak size of novel coronavirus (COVID-19) in social gathering events and industrial activities. *Int J Infect Dis*. 2020; 98: 321–327. <https://doi.org/10.1016/j.ijid.2020.06.105> PMID: 32634588
45. Bengtson VL, Silverstein M, Putney NM, Harris SC. Does religiousness increase with age? Age changes and generational differences over 35 years. *J Sci Study Relig*. 2015; 54(2): 363–379. <https://doi.org/10.1111/jssr.12183>
46. Kang M, Wei J, Yuan J, Guo J, Zhang Y, Hang J, et al. Probable evidence of fecal aerosol transmission of SARS-CoV-2 in a high-rise building. *Ann Intern Med*. 2020; 173(12): 974–980. <https://doi.org/10.7326/M20-0928> PMID: 32870707
47. Niu J, Tung TCW. On-site quantification of re-entry ratio of ventilation exhausts in multi-family residential buildings and implications. *Indoor Air*. 2008; 18(1): 12–26. <https://doi.org/10.1111/j.1600-0668.2007.00500.x> PMID: 18093125
48. Flores EO, Padilla A. Hidden threat: California COVID-19 surges and worker distress. Community and labor center at the University of California Merced; 2020 Jul [cited 26 Sep 2020]. In: Community and Labor Center at the University of California Merced [Internet]. Available from: [https://web.archive.org/web/20210402085324/https://clc.ucmerced.edu.672elmp01.blackmesh.com/sites/clc.ucmerced.edu/files/page/documents/hidden\\_threat\\_july\\_12.pdf](https://web.archive.org/web/20210402085324/https://clc.ucmerced.edu.672elmp01.blackmesh.com/sites/clc.ucmerced.edu/files/page/documents/hidden_threat_july_12.pdf)
49. Karaca-Mandic P, Sen S, Georgiou A, Zhu Y, Basu A. Association of COVID-19-related hospital use and overall COVID-19 mortality in the USA. *J Gen Intern Med*. 2020. <https://doi.org/10.1007/s11606-020-06084-7> PMID: 32815058
50. Rushing WA, Wade GT. Community-structure constraints on distribution of physicians. *Health Serv Res*. 1973; 8(4): 283–297. PMID: 4783752
51. Gozvriskaran G, Town RJ. Competition, payers, and hospital quality. *Health Serv Res*. 2003; 38(6 Pt 1): 1403–1421. <https://doi.org/10.1111/j.1475-6773.2003.00185.x> PMID: 14727780
52. Rader B, Astley CM, Sy KTL, Sewalk K, Hsuen Y, Brownstein JS, et al. Geographic access to United States SARS-CoV-2 testing sites highlights healthcare disparities and may bias transmission estimates. *J Travel Med*. 2020; 27(7). <https://doi.org/10.1093/jtm/taaa076> PMID: 32412064
53. Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environ Res*. 2020; 188: 109890. <https://doi.org/10.1016/j.envres.2020.109890> PMID: 32846654
54. COVID-19 Hospitalization and Death by Age | CDC. 2020 [cited 14 Sep 2020]. In: Centers for Disease Control and Prevention [Internet]. Available from: <https://web.archive.org/web/20200914100504/https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html>
55. Oster AM, Kang GJ, Cha AE, Beresovsky V, Rose CE, Rainisch G, et al. Trends in Number and Distribution of COVID-19 Hotspot Counties—United States, March 8–July 15, 2020. *MMWR Morb Mortal Wkly Rep*. 2020; 69(33): 1127–1132. <https://doi.org/10.15585/mmwr.mm6933e2> PMID: 32817606
56. People with Moderate to Severe Asthma | CDC. 2020 [cited 13 Sep 2020]. In: Centers for Disease Control and Prevention [Internet]. Available from: <https://web.archive.org/web/20200913052116/https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/asthma.html>

57. Wu Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020; 323(13): 1239–1242. <https://doi.org/10.1001/jama.2020.2648> PMID: 32091533
58. Chhiba KD, Patel GB, Vu THT, Chen MM, Guo A, Kudlaty E, et al. Prevalence and characterization of asthma in hospitalized and nonhospitalized patients with COVID-19. *J Allergy Clin Immunol*. 2020; 146(2): 307–314.e4. <https://doi.org/10.1016/j.jaci.2020.06.010> PMID: 32554082
59. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Sci Adv*. 2020; 6(45): eabd4049. <https://doi.org/10.1126/sciadv.abd4049> PMID: 33148655
60. Coronavirus (COVID-19). 2020 [cited 5 Sep 2020]. In: National Association of Counties [Internet]. Available from: <https://web.archive.org/web/20200908233445/https://www.naco.org/resources/covid19>
61. Dalsania AK, Fastiggi MJ, Kahlam A, Shah R, Patel K, Shiao S, et al. The relationship between social determinants of health and racial disparities in COVID-19 mortality. *J Racial Ethn Health Disparities*. 2021. <https://doi.org/10.1007/s40615-020-00952-y> PMID: 33403652
62. Golestaneh L, Neugarten J, Fisher M, Billett HH, Gil MR, Johns T, et al. The association of race and COVID-19 mortality. *EClinicalMedicine*. 2020; 25: 100455. <https://doi.org/10.1016/j.eclinm.2020.100455> PMID: 32838233
63. Egede LE, Walker RJ. Structural racism, social risk factors, and Covid-19—A dangerous convergence for Black Americans. *N Engl J Med*. 2020; 383(12): e77. <https://doi.org/10.1056/NEJMp2023616> PMID: 32706952
64. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol*. 2020; 109(5): 531–538. <https://doi.org/10.1007/s00392-020-01626-9> PMID: 32161990
65. Apicella M, Campopiano MC, Mantuano M, Mazoni L, Coppelli A, Del Prato S. COVID-19 in people with diabetes: understanding the reasons for worse outcomes. *Lancet Diabetes Endocrinol*. 2020; 8(9): 782–792. [https://doi.org/10.1016/S2213-8587\(20\)30238-2](https://doi.org/10.1016/S2213-8587(20)30238-2) PMID: 32687793
66. Battagay M, Kuehl R, Tschudin-Sutter S, Hirsch HH, Widmer AF, Neher RA. 2019-novel Coronavirus (2019-nCoV): estimating the case fatality rate—a word of caution. *Swiss Med Wkly*. 2020; 150: w20203. <https://doi.org/10.4414/smw.2020.20203> PMID: 32031234
67. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann Intern Med*. 2020. <https://doi.org/10.7326/M20-1495> PMID: 32422057
68. Florida medical examiners were releasing coronavirus death data. The state made them stop. *Tampa Bay Times*. 2020 Apr 29 [cited 2020 Sep 9]. Available from: <https://web.archive.org/web/20210901112018/https://www.tampabay.com/news/health/2020/04/29/florida-medical-examiners-were-releasing-coronavirus-death-data-the-state-made-them-stop/>
69. Dyer O. Covid-19: US states are not reporting vital data, says former CDC chief. *BMJ*. 2020; 370: m2993. <https://doi.org/10.1136/bmj.m2993> PMID: 32719009
70. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology*. 2016; 6(2). <https://doi.org/10.4172/2161-1165.1000227> PMID: 27274911
71. Hale T, Angrist N, Hale AJ, Kira B, Majumdar S, et al. Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLoS One*. 2021; 16(7): e0253116. <https://doi.org/10.1371/journal.pone.0253116> PMID: 34242239
72. How new hospital data reporting rules will affect U.S. Covid-19 response. *STAT News*. 2020 Jul 16 [cited 2020 Sep 9]. Available from: <https://web.archive.org/web/20210815215335/https://www.statnews.com/2020/07/16/hospital-data-reporting-covid-19/>