Factors Contributing to the Insufficient Understanding of how COVID-19 is Intertwined with Race and Ethnicity in the United States

by

Elora Corrine Kalix

BS Public Health, West Chester University, 2019

Submitted to the Graduate Faculty of the

Department of Epidemiology

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

Elora Corrine Kalix

on

April 30, 2021

and approved by

Essay Advisor:
Wilbert van Panhuis, MD, PhD
Associate Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh
Associate Professor, Biomedical Informatics, University of Pittsburgh School of Medicine
Director, MIDAS Coordination Center

Michael Jonah Fine, MD, MSc
Professor of Medicine, University of Pittsburgh School of Medicine
Director, VA HSR&D Center for Health Equity Research and Promotion

Samar R. El Khoudary, PhD, MPH, BPharm, FAHA
Associate Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh
Associate Professor, Clinical Translational and Science Institute
Vice Chair for Education, Epidemiology, Graduate School of Public Health

Tiffany L. Gary-Webb, PhD, MHS
Associate Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh
Associate Director, Center for Health Equity
Associate Dean for Diversity and Inclusion, Office of the Dean

Factors Contributing to the Insufficient Understanding of how COVID-19 is Intertwined with Race and Ethnicity in the United States

Elora Corrine Kalix, MPH

University of Pittsburgh, 2021

Abstract

Background: The United States is currently experiencing how the COVID-19 pandemic is disproportionately affecting many non-white communities, yet there continues to be difficulty in accurately analyzing and interpreting racial and ethnic disparities, as the methods for collecting and reporting these data vary dramatically across the nation. The goals of this study were to provide an overview of how states were reporting their COVID-19 race and ethnicity data and to explore if there were state-level factors associated with how Hispanic ethnicity was being reported.

Methods: Data on all fifty states' reporting of race and ethnicity distributions of COVID-19 cases, hospitalizations, and deaths were collected from state health departments' publicly available COVID-19 data. State factors that included sociodemographics, healthcare, and politics were collected and used in univariate and multivariate logistic regression analyses assessing associations with the outcome of reporting ethnicity as a variable separate from or included with race.

Results: As of February 2021, there were 49 (98%) states publicly reporting race and ethnicity data for COVID-19 cases, 17 (34%) for COVID-19 hospitalizations, and 45 (90%) for COVID-19

deaths. Indicators used for race and ethnicity varied across states. For all states, missing race and/or ethnicity of COVID-19 cases and deaths data ranged from 5% to 67.8% for cases and 0% to 43.3% for deaths. In the multivariate logistic regression model with case data, two predictors were statistically significantly associated with lower odds of reporting ethnicity separate from race: higher state proportion of Hispanic population (OR 0.81; 95%CI 0.68, 0.97) and higher state median income (OR 0.70 95%CI 0.52, 0.92). In the multivariate logistic regression with death data, the only factor that was statistically significant was median income (OR 0.78; 95%CI 0.65, 0.94).

Conclusion: Across the United States, data on race and ethnicity in cases, hospitalizations, and deaths from COVID-19 vary in their reporting. It is imperative that the collection and reporting of COVID-19 race and ethnicity data be improved, as the ability to have meaningful impact on this public health concern is contingent on high-quality data.

Table of Contents

List of Tables

List of Figures

Preface

On March 10, 2020, I saw a tweet that finally put me over the edge in anger and frustration about the xenophobia that was already quite rampant stemming from the spread of COVID-19. I wrote on Twitter right then, that that tweet decided for me that I would write my master's essay about racism in a pandemic. Of course, I had no idea then just how much worse it was going to get, so thank you to the racists, the xenophobes, and all the structural factors in the United States that allow this, for giving me the opportunity for this interesting, albeit angering, perspective into epidemiologic concepts about data reporting.

A genuine thank you to my committee for their involvement and guidance, as well as to my academic (and honestly, life sometimes) advisor, Dr Nancy Glynn.

Most of all, a final thank you to the global pandemic for providing the most ironic and unbelievable situation to get a degree in epidemiology during.

1.0 Introduction

On December 31, 2019, the World Health Organization (WHO) first learned of a cluster of viral pneumonia with unknown cause in the city of Wuhan, Hubei Province, People's Republic of China, and within days, investigations were underway.[1] The following week, WHO reported that Chinese researchers had determined the cause of the outbreak to be a novel coronavirus, the viral genome was sequenced, and the first related death was reported.[1] The United States reported its first case of this novel coronavirus on January 21, 2020, marking the first case in the Americas.[1] The WHO announced the official name of this novel coronavirus disease on February 11, 2020: COVID-19.[1] An abbreviation of coronavirus disease 2019, COVID-19 was chosen using the relatively recently updated disease naming best practices, wherein attention is paid to avoid stigmatization and inaccuracies associated with geographic location or groups of people.[2]

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a betacoronavirus that shares similarities to previously epidemic-producing betacoronaviruses SARS-CoV-1 and Middle Eastern respiratory syndrome coronavirus (MERS-CoV).[3] Surface glycoproteins give the coronavirus a crown-like structure that provides the name for the Coronaviridae family.[3] The virus is highly transmissible through respiratory droplets, which are expelled through actions of speaking, coughing, or sneezing.[3,4,5] The spectrum of severity of COVID-19 varies from asymptomatic to critically ill.[3] Common symptoms of mild illness are similar to what is typical of upper respiratory infections and consist of fever, headache, dry cough, and fatigue.[3,5] Other symptoms reported in mild to moderate disease include rashes, loss of taste and/or smell, and gastrointestinal symptoms like nausea or diarrhea.[3,5] More serious COVID-19 illness can lead to hypoxia requiring mechanical ventilation and sepsis requiring vasopressor support.[3]

As the virus continued to rapidly spread, WHO declared the situation a pandemic on March 11, 2020.[1] Infection prevention campaigns were put into action, and high-level suggestions from the WHO and the United States Centers for Disease Control and Prevention (CDC) were similar and comprised practicing proper hand hygiene, physical distancing of at least six feet, disinfecting frequently touched surfaces, and staying home if feeling ill.[4,5] As the pandemic and our understanding of the virus progressed, prevention and mitigation efforts were modified accordingly. In the United States, some of these prevention and mitigation efforts were at the federal level, but most fell largely at the discretion of individual states.

At the federal level, President Donald Trump declared the COVID-19 outbreak a national emergency on March 13, 2020, two days after the WHO officially recognized the outbreak as a pandemic.[6] This national emergency declaration, which was dated beginning March 1, 2020, facilitated the following disease mitigation strategies: establishing mandatory quarantines for infected and exposed individuals, provisioning personal protective equipment for health care providers, and preparing laboratories with tools for detecting SARS-CoV-2. The declaration also marked the denial of entry to the United States for foreign nationals from areas where COVID-19 was more rapidly spreading, including China, Iran, and the Schengen Area of Europe. National legislation, most notably the Coronavirus Aid, Relief, and Economic Security (CARES) Act, was passed on March 28, 2020 and provided extended unemployment and stimulus payments.[7,8]

The responses from individual states varied dramatically in terms of how quickly measures were enacted and how extensive the mitigation efforts were. Some state governors closed non-essential businesses and mandated mask wearing in public, whereas others were more reserved in their issuance of these measures.[9] States with governors more politically aligned with President Trump, particularly Florida, Georgia, and Texas, were slower in implementing their mitigation

measures.[9] While the political party line of state governors was not the only factor affecting states' responses, there were concerns that federal allocation of resources to states was influenced by the political affiliation of the governor. For example, early in the pandemic, there were concerns about the availability of personal protective equipment availability, and a federal stockpile housed equipment that could be distributed to states. Some states with governors considered political foes of President Trump (i.e., Michigan, Oregon, and New York) reported receiving far less equipment than requested, while other states with governors considered allies of the president received all equipment requested.[9] Following the passage of the CARES Act, distribution of funds was, again, seen as disproportionately in favor of political allies over where resources were in greatest need.[9]

In the general population of the United States, the pandemic quickly became politicized. Likewise, network news and newspapers polarized their reporting of COVID-19 favoring the positions and perspectives of either Democrats or Republicans.[10] One study identified significant differences by self-reported political affiliation in the opinions of the public on if news media coverage of the pandemic was accurate and providing necessary information.[11] After the first few weeks of beginning mitigation measures in the United States, a national survey explored partisan views of the response to the pandemic thus far; there was a statistically significant difference in the proportions of self-identified liberals and conservatives who believed the government was not doing enough in response and who believed the public was over- or underreacting.[12] There has also been evidence supporting the reemergence of the political and xenophobic idea of "Yellow Peril," defined as Western fear of Asians in American, evidenced by sharply increasing reports of race-driven hate crimes and sentiments focused on Asian-Americans in 2020.[13-17]

This spike in anti-Asian attacks related to COVID-19 is not the first time that racist or xenophobia attacks have occurred in response to a health crisis. In 1900, San Francisco's

Chinatown neighborhood experienced an outbreak of bubonic plague that led to serious discrimination of these Chinese Americans.[18] During the early 1900's tuberculosis outbreak in the Southern United States, Black people were labeled by public officials as unhygienic and likely to partake in activities that could increase tuberculosis risk, calling them "incorrigible."[19] The 1918 "Spanish flu" influenza pandemic did not even originate in Spain, but its name led to the stigmatization of Spanish nationals.[13] The native population in the Southwestern United States experienced stigmatization and discrimination in 1993 after an outbreak of hantavirus in the region was referred to as "Navajo disease."[18] In 2009, during the outbreak of H1N1 influenza that originated in Mexico, Mexicans and others of Hispanic descent faced increased stigma from others in the United States.[13] In the United States, immigrants from countries with endemic tuberculosis are often reluctant to be tested for tuberculosis out of fear of social stigmatization.[18] The 2014 outbreak of Ebola in Western Africa led to Africans in the United States being targeted.[20] The 2003 SARS epidemic, the viral precursor to COVID-19, which also originated in China, resulted in many Chinese Americans and others of East Asian descent experiencing targeted stigmatization; this resulted in articles written recounting the experiences similarly to what has been occurring with COVID-19.[18,20]

Unfortunately, within the current COVID-19 pandemic, this racism and xenophobia towards Asians and Asian Americans has continued. Chinese people have generally been the target of this, given the virus's origination in China, but hate has spilled into other Asian subgroups, likely due to the documented phenomenon where non-Asians conflate various Asian groups, suggesting that "all Asians look alike."[21] Participating in this hate has become easier with the internet, particularly through sites like Twitter and 4chan, both of which have been focuses of studies during the COVID-19 pandemic about their use in Asian discrimination.[21,22] One study,

which analyzed COVID-19-related tweets around the start of the outbreak, found that tweets about Asians that were negative increased by 68.4% between November 2019 and March 2020 while negative tweets about other racial and ethnic groups remained stable.[21] A thematic analysis of COVID-19 tweets in this same study showed that about 20% of the COVID-19 tweets expressed racism and/or blame towards Asians and/or Asian Americans; 18% of COVID-19 tweets during the same time mentioned Donald Trump. It has been noted that just because a lot of this hate has been occurring on the internet does not mean it does not have serious and tangible effects on Asian and Asian-American communities.[16]

The use of terms like "Wuhan virus" and "China virus" were frequent early in the outbreak, where they were Googled extensively in January 2020, although at this time, the geographic location was the main information to go off of for searching purposes.[21] These terms and their variants became the catalyst for many debates on Twitter.[21] By February 2020, when the WHO released its naming guidelines of this novel virus, specific in avoiding potential geographic stigmatization, searches and use of these terms online slowed. However, Google searches of these terms spiked again in March, rising 650%, after their frequent use by some United States politicians.[21] These include US Representative from Arizona Paul Gosar, former US Secretary of State Mike Pompeo, and most notably, then-President Donald Trump.[21] The participation in and defense of the use of these terms continued. A joint statement, written by the US State Department, from the Group of Seven (G7) member nations regarding the virus used the term "Wuhan virus."[23] By this time in late March 2020, the WHO had already officially named the virus, sparking criticisms and separate statements made by other G7 nations. The State Department countered that the use of "Wuhan virus" is meant to be informative, though it has otherwise been referred to as inflammatory, especially given the spiking reports of racist and xenophobic incidents involving

Asians and Asian Americans in the United States around this time.[23] Secretary of State Mike Pompeo continued to publicly target China.[23]

It did not take much time before the United States faced international criticism for its initial handling of the pandemic and use of such terms. However, plenty of criticism was coming from within the nation as well, particularly once there were reported trends of minority populations in the United States being disproportionately affected by COVID-19. In early April 2020, the states of Michigan and Wisconsin released state COVID-19 data that showed disparate effects of rates of COVID-19 infection by race.[19] Soon after, authorities from the cities of Chicago and New York City reported that their Black and Hispanic/Latino populations had higher case rates and mortality than white individuals.[24] Racial and ethnic disparities in COVID-19 infection and mortality continued to be reported in other locations, including front-page articles in newspapers such as the *New York Times* and the *Los Angeles Times*.[19] With concerns about the quality of collected and reported demographic race and ethnicity data, two Democratic elected United States legislators from Massachusetts, Senator Elizabeth Warren and Representative Ayanna Pressley, faulted the government for the lack of publicly-reported racial and ethnic demographic data on COVID-19 testing and infections. Following their open letter, several states and municipalities began to incorporate such demographic data in publicly available COVID-19 datasets.[19] Members of a House of Representatives health subcommittee met with Dr. Robert R. Redfield, the director of the CDC, and openly criticized the agency's anticipation and response to the pandemic and its disproportionate effects on racial and ethnic minority communities.[25]

These criticisms of the CDC and federal response to the pandemic at large resulted in new data reporting requirements from the Trump Administration on June 4, 2020.[26] For COVID-19 test results, data on race and ethnicity, age, and sex became required for reporting to local and state

health departments, which would subsequently report these to the CDC.[25,26] While many states otherwise also report about their COVID-19 hospitalizations and deaths, collecting and/or reporting race and ethnicity for hospitalizations or deaths was not mandated. After the first few months of the demographics reporting, race and ethnicity data for COVID-19 cases as reported by the CDC were sparce. For example, as of September 16, 2020, 50% of the reported cases had missing data for race and ethnicity.[26] One early report of the missing race and ethnicity data for COVID-19 cases noted that these data were missing for up to a third of cases in states where some then-leaders of the US government call home, notably the President (Florida), the Vice President (Indiana), the Senate Majority Leader (Kentucky), and the Speaker of the House of Representatives (California).[26] Missing data linking the demographic characteristics of those with COVID-19 infection and their medical outcomes have been a major limitation of studies focused on variation in COVID-19 infection, COVID-19 medical outcomes, and the quality and equity of COVID-19 management in the United States.[24,26-31]

For some racial groups in the United States, the difficulties of drawing conclusions are compounded by the fact that some race categories are often left out. For the minority racial populations in the United States who comprise much smaller proportions of the population, specifically American Indians, Alaska Natives, and Pacific Islanders, the respective category for race may not be an option, as suggested in studies of racial disparities of COVID-19.[24,27,28] Moreover, given the available race category options available to choose from on demographics forms in other epidemiologic studies or healthcare settings, people who are American Indian or Alaska Native are very commonly misclassified as another non-American Indian or Alaska Native race, adding to the inability to accurately estimate rates of morbidity and mortality in American Indian and Alaska Native populations.[28]

Among these concerns about misclassifying an individual's race depending on what race options are available to choose from, there is the fact that a standard of collecting race and ethnicity data in the United States does exist. These standards, the Race and Ethnic Standards for Federal Statistics and Administrative Reporting, set forth originally by the United States Office of Management and Budget in 1977, outline the categories, definitions, use, reporting, and presentation of race and ethnicity data in federal contexts.[34] These standards identify five race categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.[34] There are two ethnicity categories: Hispanic or Latino and Not Hispanic or Latino.[34] These standards do indicate that ethnicity and race are, and should be, considered two separate variables, but the standards do also provide guidance on including race and ethnicity in one variable, even though this may result in incomplete data. The United States Census Bureau of the United States uses the same main categorization of race and ethnicity to quantify the racial and ethnic makeup of each state.[35] Individual states have no mandate to follow these standards, allowing for the indicators used in collection and reporting to be at states' discretion. The inability to effectively compare COVID-19 rates by race to the state proportion of that race, that some studies exploring these racial disparities have encountered thus far, can be explained in part by these inconsistent methods used for reporting race and ethnicity data. The current definitions of these federal race and ethnicity standards categories are shown in Table 1.

Even with the poor data, what are available do indicate 3.5 times higher confirmed COVID-19 cases among American Indian or Alaska Native individuals as compared to non-Hispanic whites.[28,32] One explanation for the contribution of health disparities between American Indian or Alaska Native and white populations is the presence of historical trauma and continued, persistent racial inequalities, ideas which in part define the concepts of structural and systemic racism. As

Table 1. Summary of current definitions of race and ethnicity categories for federal reporting standards in the United States[34,34]

Category	Definition
American Indian or Alaska Native	A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment
Asian	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam
Black or African American	A personal having origins in any of the black racial groups of Africa
Native Hawaiian or Other Pacific Islander	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands
White	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa
Hispanic or Latino	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race

this is not an experience unique to American Indian and Alaska Natives, this could also contribute to the explanation of the effects of COVID-19 on Black and Hispanic communities that have been understood so far. From as early in the pandemic as mid-April 2020, the counties in the United States with an above national population average of Black residents (22% of United States counties; 13% of total population is Black) accounted for just over half of all cases and deaths from COVID-19 that had occurred thus far.[32] Similarly, of the counties in the United States that were labelled as "hot spots" for COVID-19, 75% had disproportionately high Hispanic/Latino populations.[32]

Past epidemics and natural disasters consistently indicate that marginalized populations, including racial and ethnic minorities, will be most disproportionately affected.[19] This fact has led

some to question how disproportionate effects of COVID-19 on racial minorities were not better anticipated, even describing the resulting data situation as "grossly inadequate," as South Carolina Representative James E Clyburn did in a House of Representatives subcommittee meeting on the pandemic.[25]

1.1 Gaps in Knowledge and Public Health Significance

Together, this has all lead to an insufficient understanding of the extent to which COVID-19 is intertwined with race and ethnicity. Across the United States, mandates and systems to collect and report race and ethnicity data of COVID-19 cases, hospitalizations, and/or deaths differ in methodologies and thoroughness. While numerous studies have begun trying to assess racial and ethnic disparities in COVID-19 infection, management, and outcomes within states, these have all been limited by the quality of the linked demographic data available. Yet, there has not been much done to identify and compile what all is actually being reported by each state. Additionally, there has not yet been much exploration into what, if any, factors might be associated with how or why data are being reported the way they are in states. Factors related to state sociodemographics, healthcare, and politics have rationale for potential importance in inclusion in these explorations.

Quality of data is vital for making accurate and appropriate inferences and conclusions in epidemiological studies. Without effective understanding of how and/or why these COVID-19 race and ethnicity data are being reported the way they are, further research, interventions, or policy change could be ineffective in improving this issue. Racism and xenophobia are significant public health concerns, so substantive research into these disparities is necessary for improving the health and well-being of the effected communities. As of late March 2021, there have been

nearly 31 million reported cases of COVID-19 in the United States, accounting for about 23% of the global case count.[33] Given the high burden of COVID-19 in the United States, it is imperative to find ways to improve this serious public health concern.

2.0 Objectives

This study had two objectives:

The first objective of this study was to explore and provide an overview of each state's publicly reported race and ethnicity data for COVID-19 cases, hospitalizations, and/or deaths.

The second objective of this study was to assess if there are state-level sociodemographic, political, or healthcare characteristics associated with the proper separation of ethnicity and race.

3.0 Methods

3.1 Study design

This cross-sectional ecological study was conducted at the state level in the United States using publicly available state health department data. Given the nature of the research question and resources available for this study, an ecological study at the state level was appropriate. This study included all fifty states and did not include Washington, DC, Puerto Rico, or the other United States territories.

3.2 Collection of reported race and ethnicity data

Data on reported race and ethnicity categorization indicators for COVID-19 cases, hospitalizations, and/or deaths were collected manually from each state's department of health's publicly available COVID-19 dashboard (Table 2). All race and ethnicity options the states were reporting were collected, including if the state was reporting ethnicity separate from or included with race. These manually collected indicators from each state of how their COVID-19 race and ethnicity data were being reported were individually and manually reviewed and compared with the categorization of the federal race and ethnicity reporting standards. Proportion of cases and/or deaths with missing race and/or ethnicity were also recorded; this was determined either by explicitly stated "unknown" or "missing" counts or by calculating the difference between total

Table 2. Each of the fifty states' health department's main COVID-19 dashboard pages

State	COVID-19 Dashboard Source	Ref #
Alabama	https://alpublichealth.maps.arcgis.com/apps/MapSeries/index.html?appid=d84846411471404c83313bfe7ab2a367	36
Alaska	https://alaska-coronavirus-vaccine-outreach-alaska-dhss.hub.arcgis.com/	37
Arizona	https://www.azdhs.gov/covid19/data/index.php	38
Arkansas	https://experience.arcgis.com/experience/c2ef4a4fcbe5458fbf2e48a21e4fece9	39
California	https://public.tableau.com/profile/ca.open.data#!/vizhome/COVID-19CasesDashboardv2_0/CaseStatistics	40
Colorado	https://covid19.colorado.gov/data	41
Connecticut	https://data.ct.gov/stories/s/q5as-kyim	42
Delaware	https://myhealthycommunity.dhss.delaware.gov/locations/state	43
Florida	https://experience.arcgis.com/experience/96dd742462124fa0b38ddedb9b25e429	44
Georgia	https://dph.georgia.gov/covid-19-daily-status-report	45
Hawai'i	https://health.hawaii.gov/coronavirusdisease2019/what-you-should-know/current-situation-in-hawaii/	46
Idaho	https://public.tableau.com/profile/idaho.division.of.public.health#!/vizhome/DPHIdahoCOVID-19Dashboard/Home	47
Illinois	https://www.dph.illinois.gov/covid19	48
Indiana	https://www.coronavirus.in.gov/2393.htm	49
Iowa	https://coronavirus.iowa.gov/	50
Kansas	https://www.coronavirus.kdheks.gov/160/COVID-19-in-Kansas	51
Kentucky	https://kygeonet.maps.arcgis.com/apps/opsdashboard/index.html#/543ac64bc40445918cf8bc34dc40e334	52
Louisiana	https://ldh.la.gov/coronavirus/	53
Maine	https://www.maine.gov/covid19/vaccines/dashboard	54
Maryland	https://coronavirus.maryland.gov/	55
Massachusetts	https://www.mass.gov/info-details/covid-19-response-reporting	56
Michigan	https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html	57
Minnesota	https://www.health.state.mn.us/diseases/coronavirus/situation.html	58
Mississippi	https://msdh.ms.gov/msdhsite/_static/14,0,420.html	59
Missouri	https://showmestrong.mo.gov/data/public-health/	60
Montana	https://montana.maps.arcgis.com/apps/MapSeries/index.html?appid=7c34f3412536439491adcc2103421d4b	61
Nebraska	https://experience.arcgis.com/experience/ece0db09da4d4ca68252c3967aa1e9dd/page/page_0/	62
Nevada	https://nvhealthresponse.nv.gov/	63
New Hampshire	https://www.covid19.nh.gov/	64
New Jersey	https://covid19.nj.gov/	65
New Mexico	https://cvprovider.nmhealth.org/public-dashboard.html	66
New York	https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n	67
North Carolina	https://covid19.ncdhhs.gov/dashboard	68
North Dakota	https://www.health.nd.gov/diseases-conditions/coronavirus/north-dakota-coronavirus-cases	69
Ohio	https://coronavirus.ohio.gov/wps/portal/gov/covid-19/dashboards	70
Oklahoma	https://oklahoma.gov/covid19.html	71
Oregon	https://public.tableau.com/profile/oregon.health.authority.covid.19#!/vizhome/OregonsCOVID-19DataDashboards-TableofContents/TableofContentsStatewide	72
Pennsylvania	https://www.health.pa.gov/topics/disease/coronavirus/Pages/Cases.aspx	73
Rhode Island	https://ri-department-of-health-covid-19-data-rihealth.hub.arcgis.com/	74
South Carolina	https://scdhec.gov/covid19/sc-testing-data-projections-covid-19	75
South Dakota	https://doh.sd.gov/COVID/Dashboard.aspx	76
Tennessee	https://www.tn.gov/health/cedep/ncov/data.html	77
Texas	https://txdshs.maps.arcgis.com/apps/dashboards/ed483ecd702b4298ab01e8b9cafc8b83	78
Utah	https://coronavirus.utah.gov/case-counts/	79
Vermont	https://www.healthvermont.gov/covid-19/current-activity/vermont-dashboard	80
Virginia	https://www.vdh.virginia.gov/coronavirus/covid-19-in-virginia/	81
Washington	https://www.doh.wa.gov/Emergencies/COVID19/DataDashboard	82
West Virginia	https://dhhr.wv.gov/COVID-19/Pages/default.aspx	83
Wisconsin	https://www.dhs.wisconsin.gov/covid-19/data.htm	84
Wyoming	https://health.wyo.gov/publichealth/infectious-disease-epidemiology-unit/disease/novel-coronavirus/covid-19-state-and-county-dashboards/	85

cases/deaths reported and the total cases/deaths with race/ethnicity reported. Data collection occurred in January and February 2021.

3.3 Collection of state-level characteristics

State factors related to sociodemographics, healthcare, and politics were also collected from secondary sources for use in analyses of associations with states who are and are not reporting ethnicity separate from race in COVID-19 cases and deaths (Table 3). Included covariates were hypothesized from previous inclusion in statistical analyses with race and ethnicity, before or during the COVID-19 pandemic.

States' racial and ethnic demographic makeup may influence decisions related to that race or ethnicity, whether positively or negatively.[86,87] The proportion of each state's population who report race as white and the proportion who report a Hispanic/Latino ethnicity were collected from the United States Census Bureau 2019 population estimates.[88]

Income of the state population may serve as an understanding of the state's economic position, which can be related to political or disparate racial factors.[29,30,87] Median household income for each state was collected from the United States Census Bureau 2015-2019 American Community Survey 5-Year Estimates.[88,89]

Per capita public health spending may be indicative of the prioritization of public health in the state.[90] The per capita public health spending by state was collected from a Kaiser Health and AP report of 2016-2018 state spending using data from the State Health Expenditure Dataset from the Optimizing Governmental Health and Social Spending Interactions project.[91,92]

Table 3. Summary of covariates used in statistical analyses

Variable	Definition	Mean (Range)	Source
Political factors			
State governor political party	The political party, either republican or democrat, of the state governor at the start of the pandemic	0 = Republican (52%) 1= Democrat (48%)	95
State 2020 presidential election majority vote	The final majority vote winner of each state after the 2020 presidential election	0 = Donald Trump (50%) 1= Joe Biden (50%)	96
Sociodemographic factors			
State white population proportion	The 2019 estimate of state population proportions whose race is white	Mean= 78.7% SD= 12.3% Range= 25.2% - 94.4%%	88
State Hispanic population proportion	The 2019 estimate of state population proportions whose ethnicity is Hispanic	Mean= 12.3% SD= 10.5% Range= 1.7% - 49.3%	88
State median household income	The 2015-2019 estimated median household income for each state, in thousands of dollars	Mean= $62.6 SD= $10.3 Range= $45.1 - $84.8	88, 89
Healthcare factors			
State per capita public health spending	The 2016-2018 average public health spending per capita for each state, in dollars	Mean= $103.30 SD= $45.92 Range= $50 - $289	91, 92
Proportion of state population under age 65 without health insurance	The 2019 proportion of states' populations under age 65 who did not have private or public health insurance	Mean= 10% SD= 3.6% Range= 3.5% - 20.8%	88, 94

The proportion of a state's population without health insurance may act as another measure of health prioritization of the state, while also incorporating potential political alignment, as some of a state's health coverage may be explained by that state's decision to participate in Medicaid expansion.[9,29,32,93] The proportion of states' populations under age 65 without health insurance

coverage was collected from the United States Census Bureau 2019 Health Insurance in the United States Report.[88,94]

As political alignment of a state and its government may influence financial or social decisions, as well as how politicized the pandemic became, general proxies of the state political alignment may contribute to explaining associations.[86] The political party of the state governor during 2020 when the pandemic began is included since most state-level mitigation measures fell at the discretion of the state governor. The 2020 presidential election was contentious in the United States, partly due to the politicization of the pandemic; a state's majority vote in this election may serve as a proxy to that state's alignment with the way the pandemic was or was not handled. State political factors included the state governor's political party during the pandemic and the state's majority vote in the 2020 presidential election, collected from the National Governors Association and the United States National Archives and Records Administration, respectively.[95,96] For the states of Maine and Nebraska, where the presidential vote of the state is split between electors, what was used for this analysis was the candidate who received the majority of the state's electors.

3.4 Statistical analysis

Descriptive statistics were used to compare state factors between the states reporting ethnicity separate from and included with race using t-tests to compare means of continuous variables and chi-square tests, or Fisher's exact as appropriate, to compare categorical variables. Statistical significance was defined at an alpha level of 0.05. Univariate and multivariate logistic regression analyses were constructed to explore any associations between the sociodemographic, healthcare, and political state factors with the state reporting ethnicity separately from or included

with race for both case data and death data. Hospitalization data were not included in analyses due to the low proportion of states reporting race and ethnicity data for COVID-19 hospitalizations. In the binary outcome of reporting ethnicity separate from race or not, states not reporting any ethnicity data were included with the states not reporting their ethnicity data separate from race. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC).

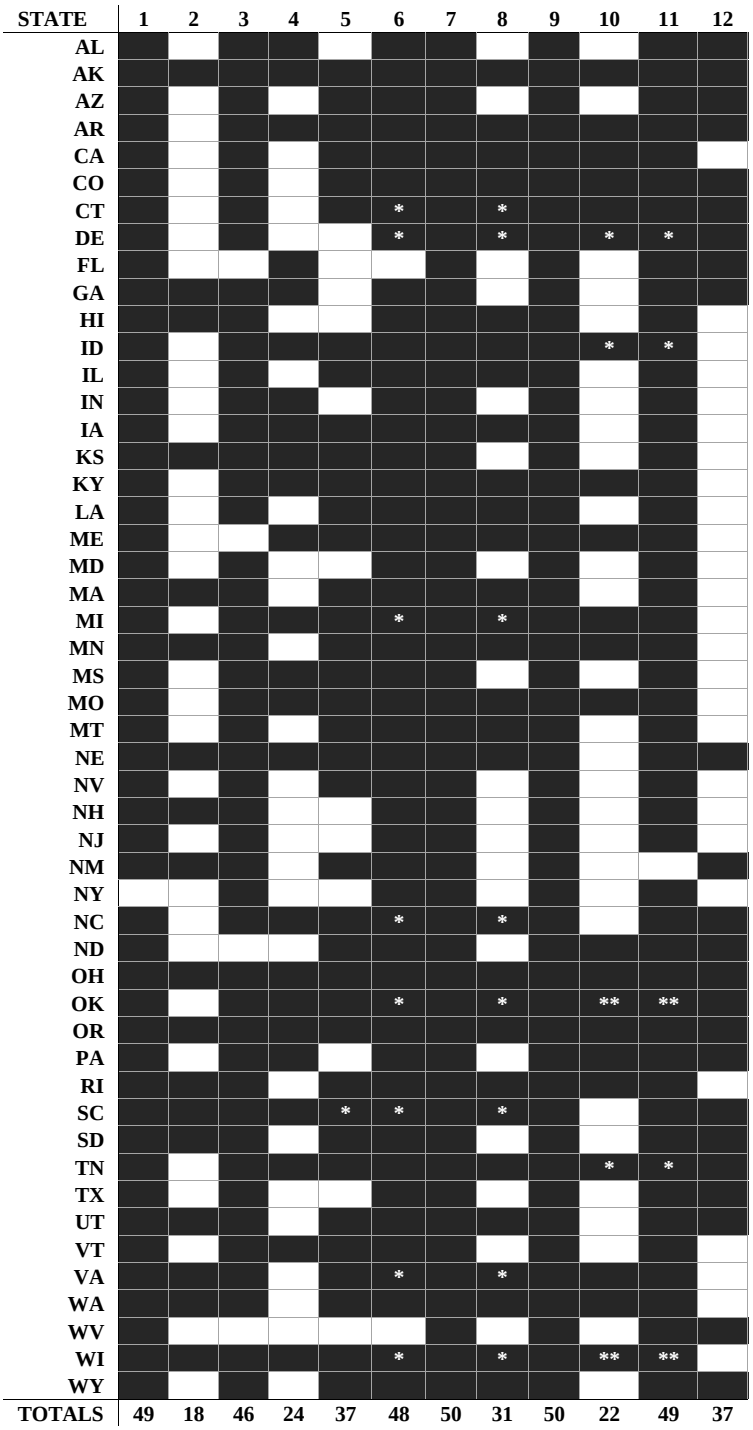
4.0 Results

4.1 Race and ethnicity indicators reported by states

As of February 2021, there were 49 (98%) states publicly reporting race and ethnicity data for COVID-19 cases, 17 (34%) states reporting these data for COVID-19 hospitalizations, and 45 (90%) reporting these data for COVID-19 deaths (Figure 1). There were 27 (54%) states following the federal race reporting standards of American Indian/Alaska Native, Asian, Black, Native Hawaiian and other Pacific Islanders, and white in their COVID-19 case and death data. Twenty-four (48%) states were reporting Hispanic/Latino ethnicity separate from race for COVID-19 case data, though 5 (10%) were not reporting ethnicity data at all (Figure 2). Twenty-three (46%) states were reporting Hispanic/Latino ethnicity separate from race for COVID-19 death data, though 6 (12%) were not reporting ethnicity data at all (Figure 3). There were 16 (32%) states whose race and ethnicity reporting closely followed the race and ethnicity reporting standards but included Hispanic/Latino ethnicity as a part of race (Figure 1).

4.2 Missing race and ethnicity data

The completeness of reported race and ethnicity data for COVID-19 cases and deaths varied greatly between and within states. For all states, missing race and ethnicity of cases and deaths ranged from 5% to 67.8% for cases and 0% to 43.3% for deaths. For states who reported race and ethnicity separately, missing race data for cases ranged from 7.2% to 38.1% and 1% to



Legend

1: Reporting race and ethnicity for cases
2: Reporting race and ethnicity for hospitalizations
3: Reporting race and ethnicity for deaths
4: Reporting ethnicity separate from race (for cases)

5: American Indian or Alaska Native
6: Asian
7: Black/African American
8: Native Hawaiian or Other Pacific Islander
9: White
10: Biracial/Multiracial
11: Other
12: Unknown/Refused/Missing

Figure 1. Race and ethnicity data publicly available from each state

Note: Columns 1-3 mark which states are reporting race and/or ethnicity for cases, hospitalizations, and/or deaths. Column 4 indicates whether the state is reporting ethnicity separate from race (filled) or included with race (not filled). Columns 5-9 indicate which race and/or ethnicity categories the state includes. Columns 10-12 are the additional race categories that may or may not be options in that state. Asterisks (*) indicate that state reported that specific category, but it was combined with the other categories with an asterisk.

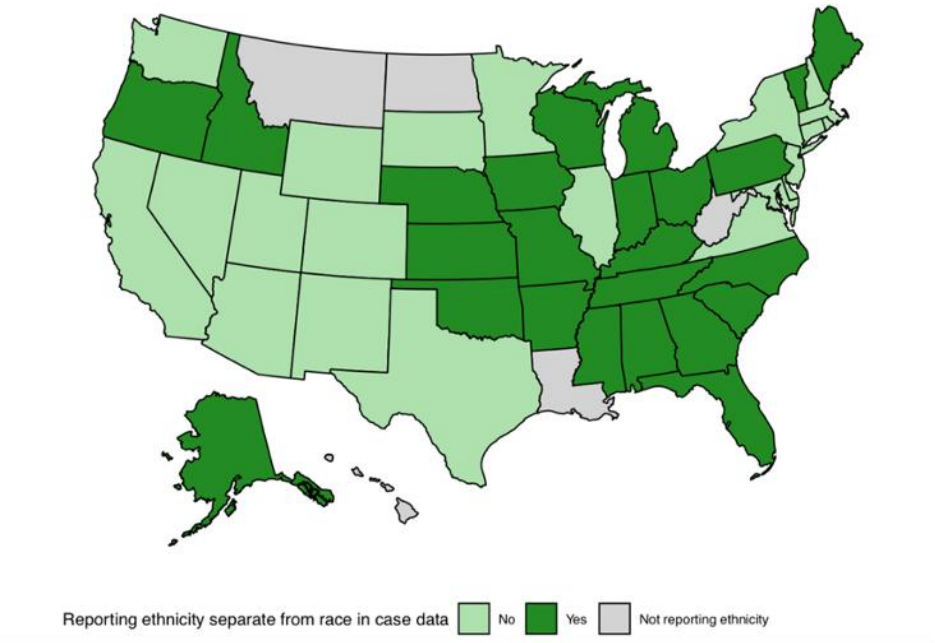


Figure 2. Geographic distribution of states' reporting of ethnicity data for case data

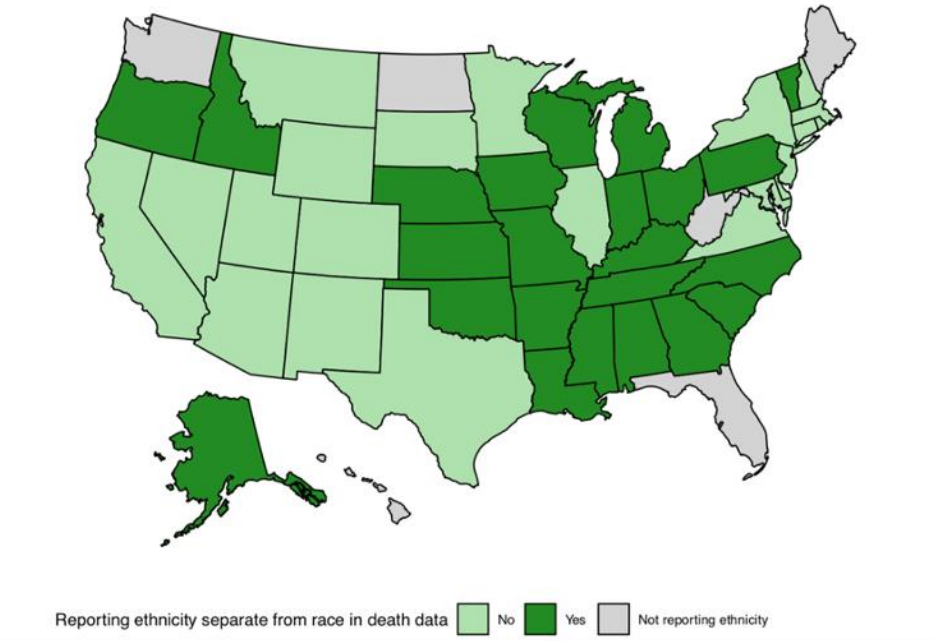


Figure 3. Geographic distribution of states' reporting of ethnicity data for death data

15.8% for deaths. Missing ethnicity data ranged from 14.7% to 67.8% for cases and 0% to 43.3% for deaths. Nationally, there was approximately 27.4% of race/ethnicity data missing for cases, 9.6% of race/ethnicity data missing for deaths, 33.4% of ethnicity-specific data missing for cases, and 14.4% of ethnicity-specific data missing for deaths. Further analysis of these data was not plausible given the inability to parse out the missing races and the missing ethnicities for the states who report these characteristics within a single variable.

4.3 Descriptive Statistics

The distribution of each included predictor characteristic— sociodemographic (white population proportion, Hispanic population proportion, and state median income), healthcare (proportion of population uninsured and per capita public health spending), and political (state governor political party and state 2020 presidential majority vote)— are visually summarized (Figure 4-Figure 8) and geographically described (Figure 9-15). These factors were also analyzed for differences between states reporting ethnicity separate from or included with race, for both case (Table 4) and death data (Table 5). Between the states that did and did not separate race and ethnicity for cases, there were statistically significant differences in the state majority vote in the 2020 presidential election ($p=0.0006$), the proportion of the population that is Hispanic ($p=0.001$), and the median household income ($p<0.0001$). Similarly, the statistically significantly different factors between states that did and did not separate race and ethnicity for deaths were the state majority vote in the 2020 presidential election ($p=0.0009$), the proportion of the population that is Hispanic ($p=0.0008$), and the median household income ($p<0.0001$). For both case and death data, differences between states separately reporting ethnicity from race were not statistically significant

for proportion of the population that is white, the state governor political affiliation, the per capita public health spending, or the proportion of the population under age 65 without health insurance.

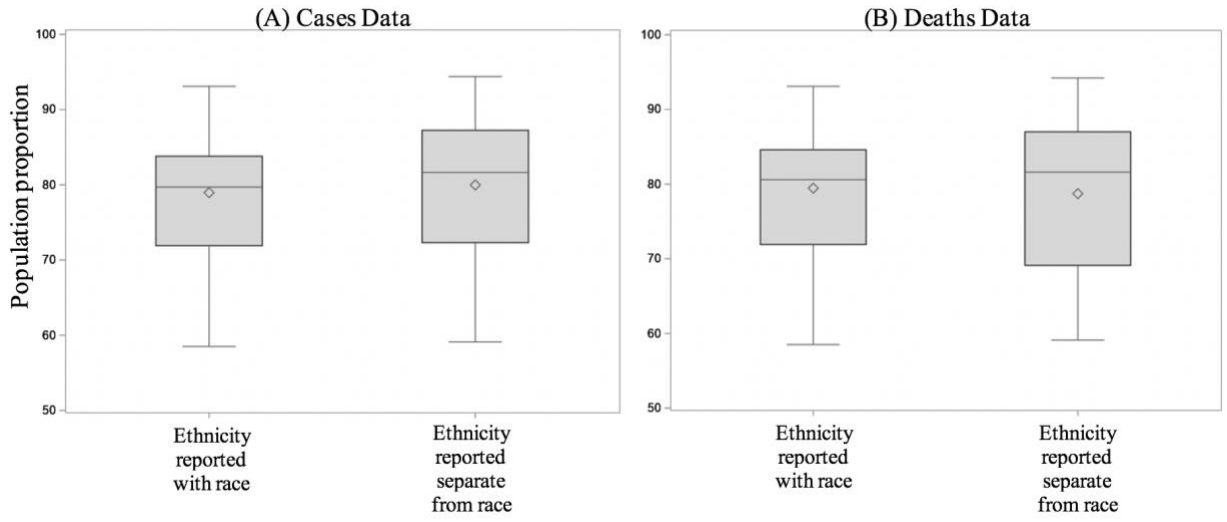


Figure 4. The distribution of white state population proportion by states reporting ethnicity separate from and included with race for both cases (A) and deaths (B)

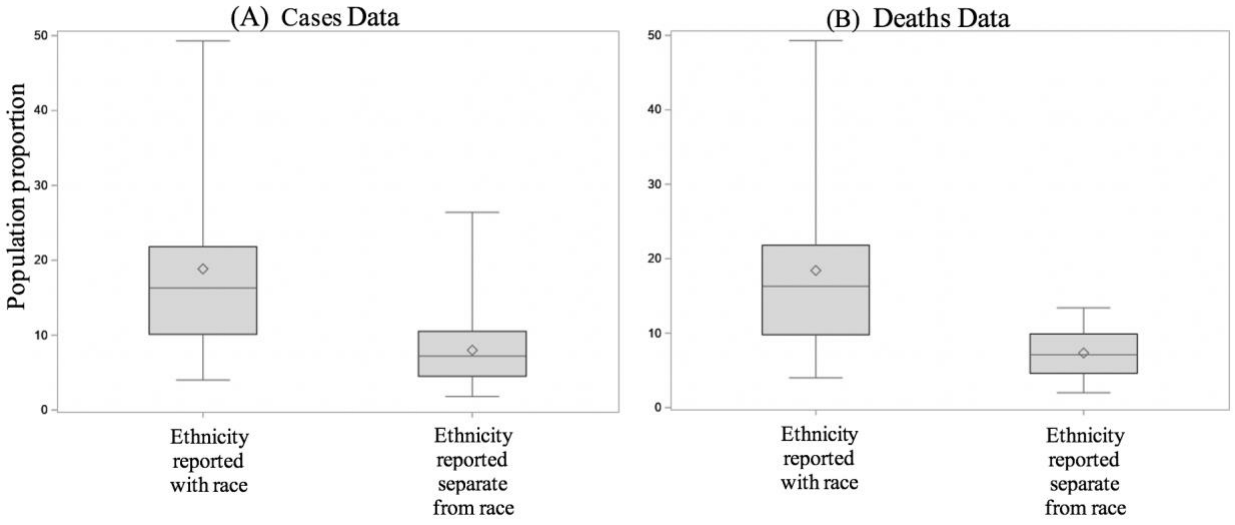


Figure 5. The distribution of Hispanic state population proportion by states reporting ethnicity separate from and included with race for both cases (A) and deaths (B)

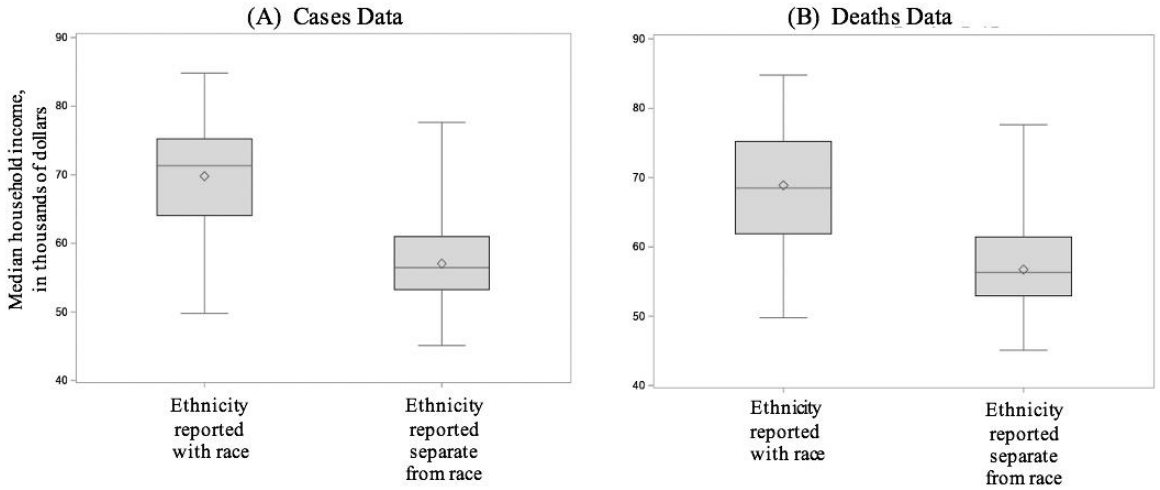


Figure 6. The distribution of state median household income by states reporting ethnicity separate from and included with race for both cases (A) and deaths (B)

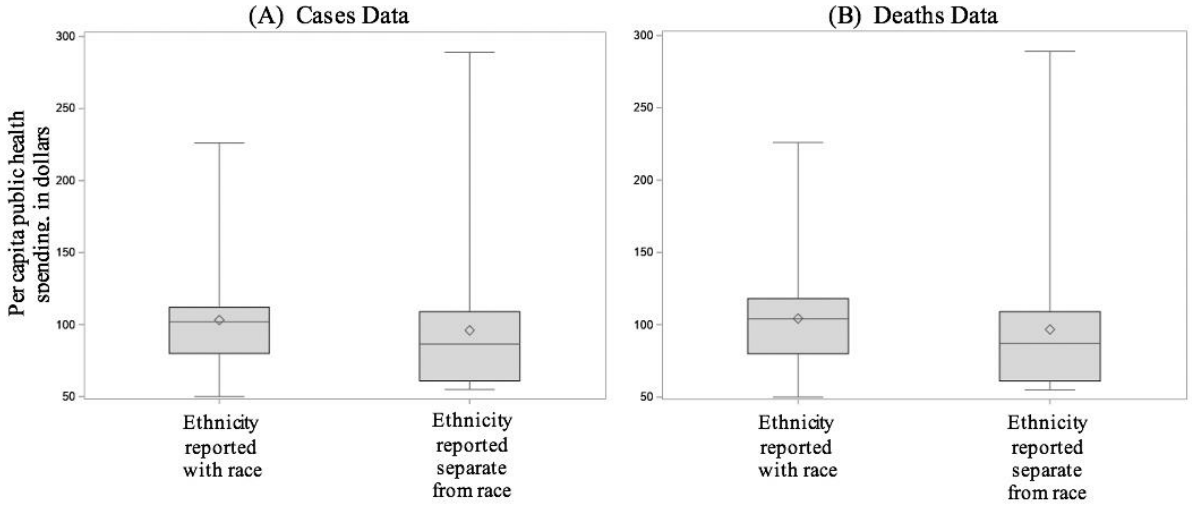


Figure 7. The distribution of state per capita public health spending by states reporting ethnicity separate from and included with race for both cases (A) and deaths (B)

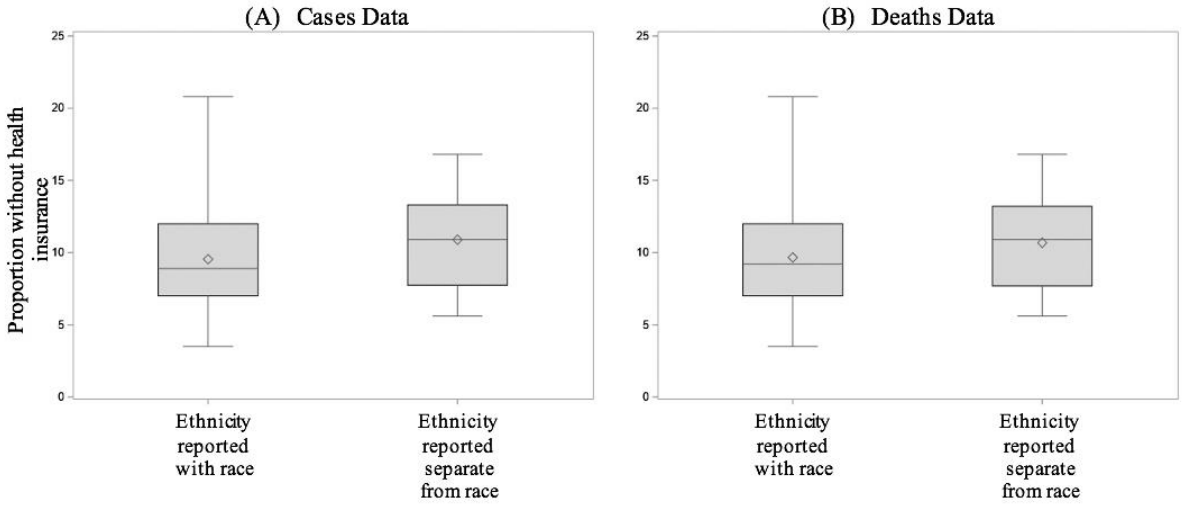


Figure 8. The distribution of state population proportions under age 65 without health insurance by states reporting ethnicity separate from and included with race for both cases (A) and deaths (B)

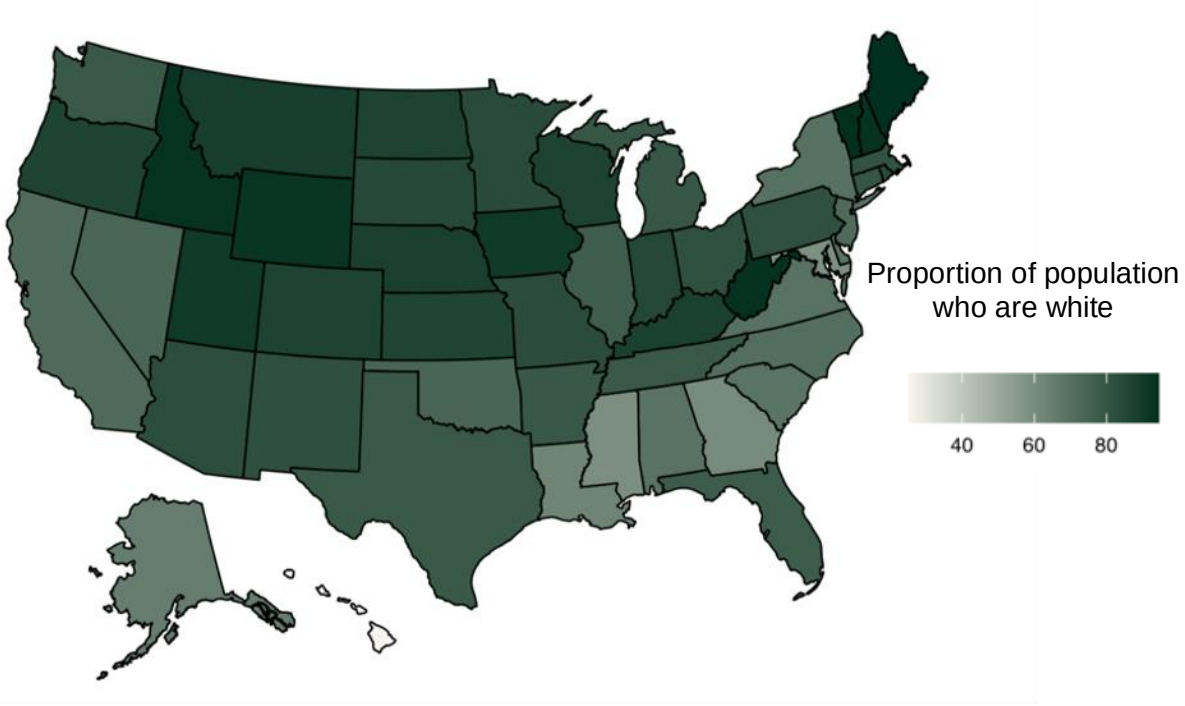


Figure 9. Geographic distribution of white proportion of state populations

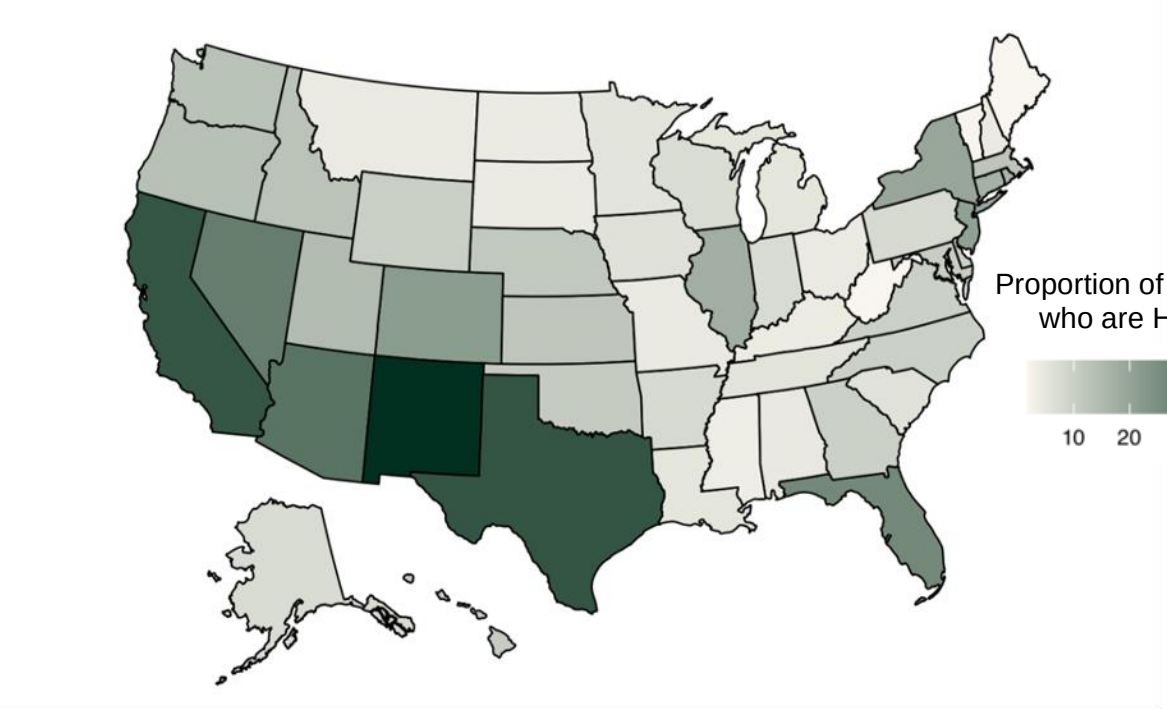


Figure 10. Geographic distribution of Hispanic proportion of state populations

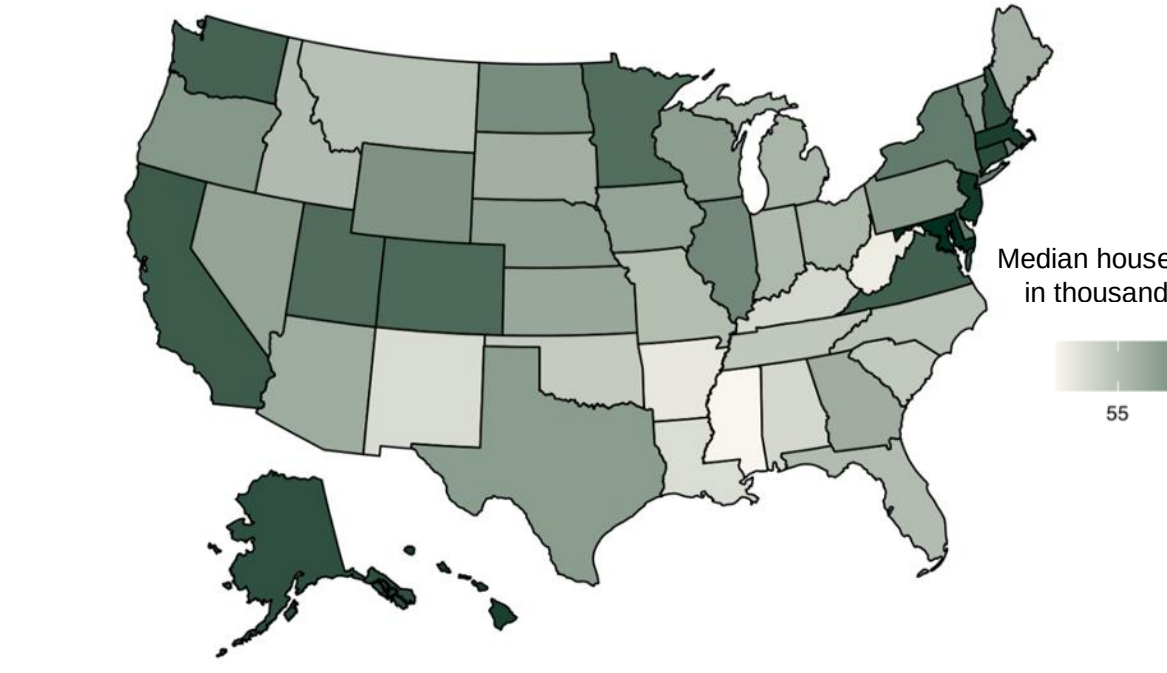


Figure 11. Geographic distribution of states' median household incomes

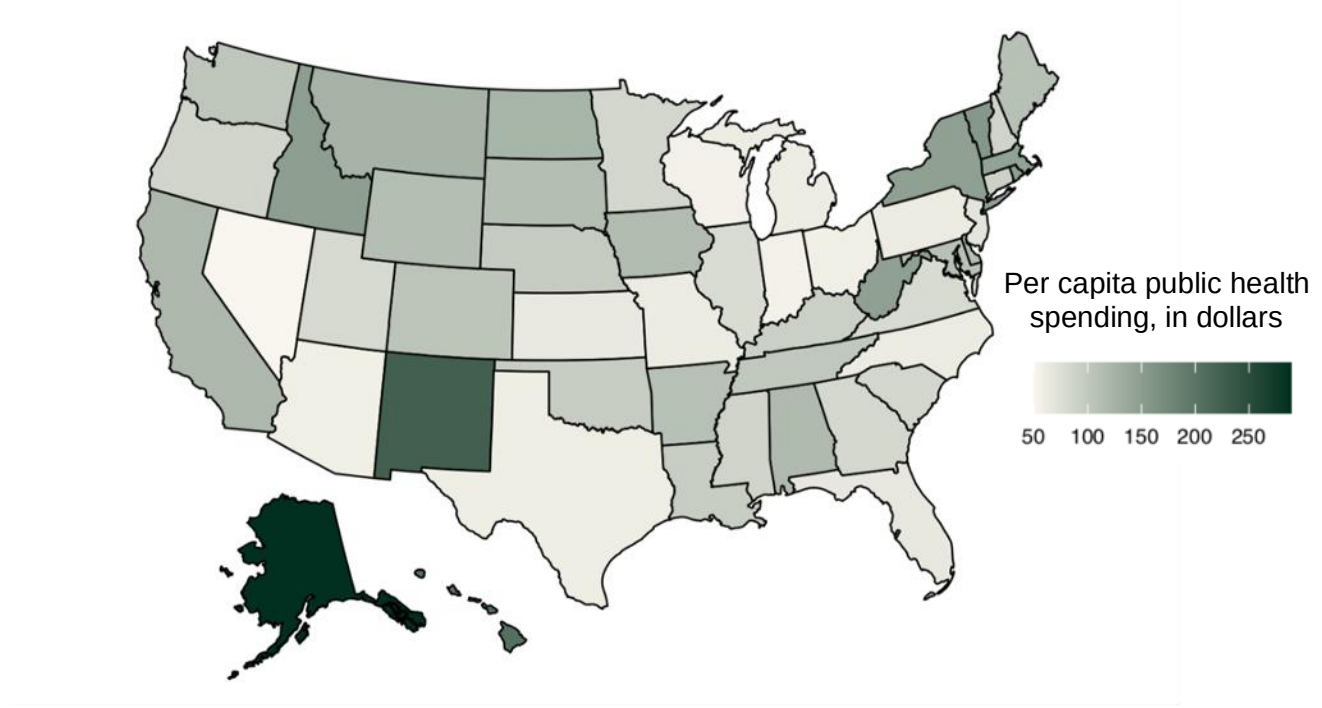


Figure 12. Geographic distribution of states' per capita public health spending

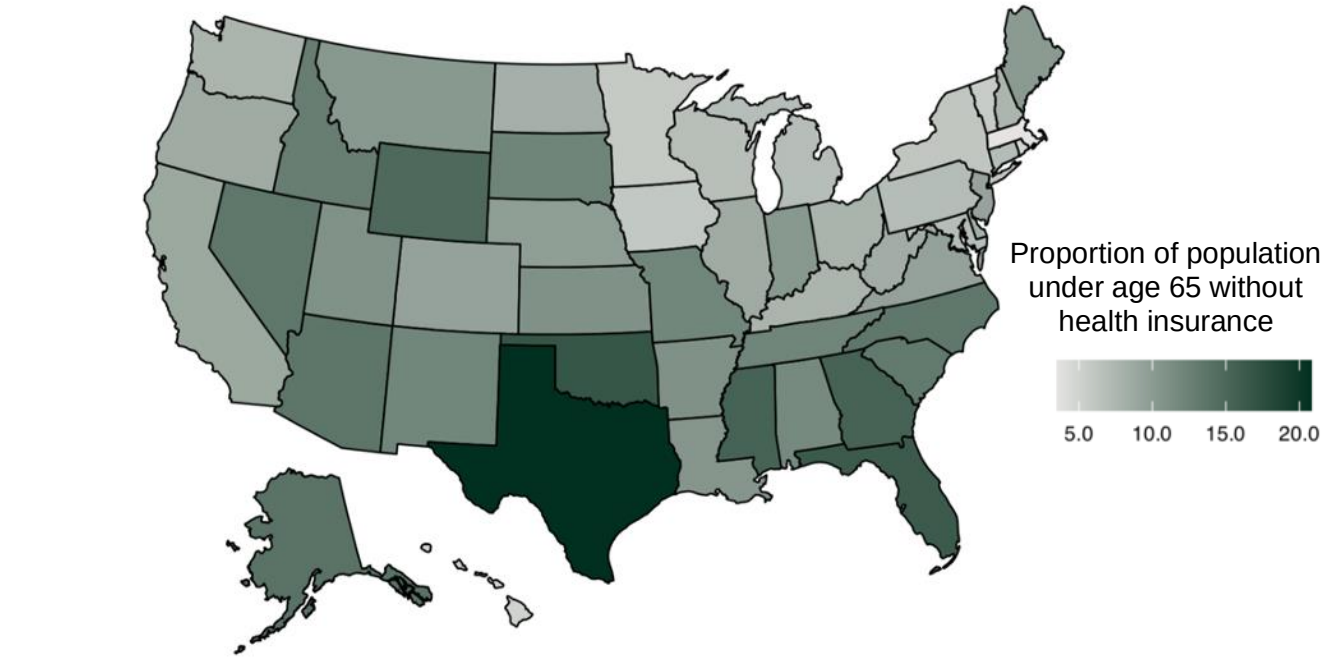


Figure 13. Geographic distribution of states' population proportions under age 65 without health insurance

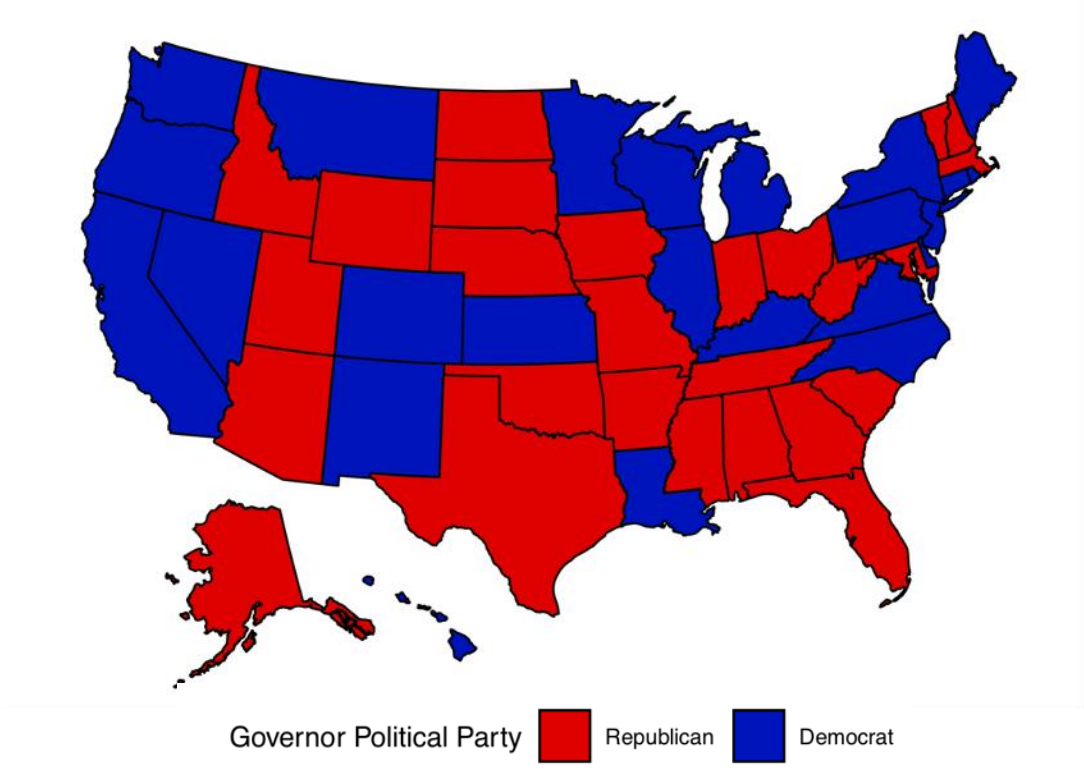


Figure 14. Geographic distribution of states' governor political party affiliation at start of COVID-19 pandemic

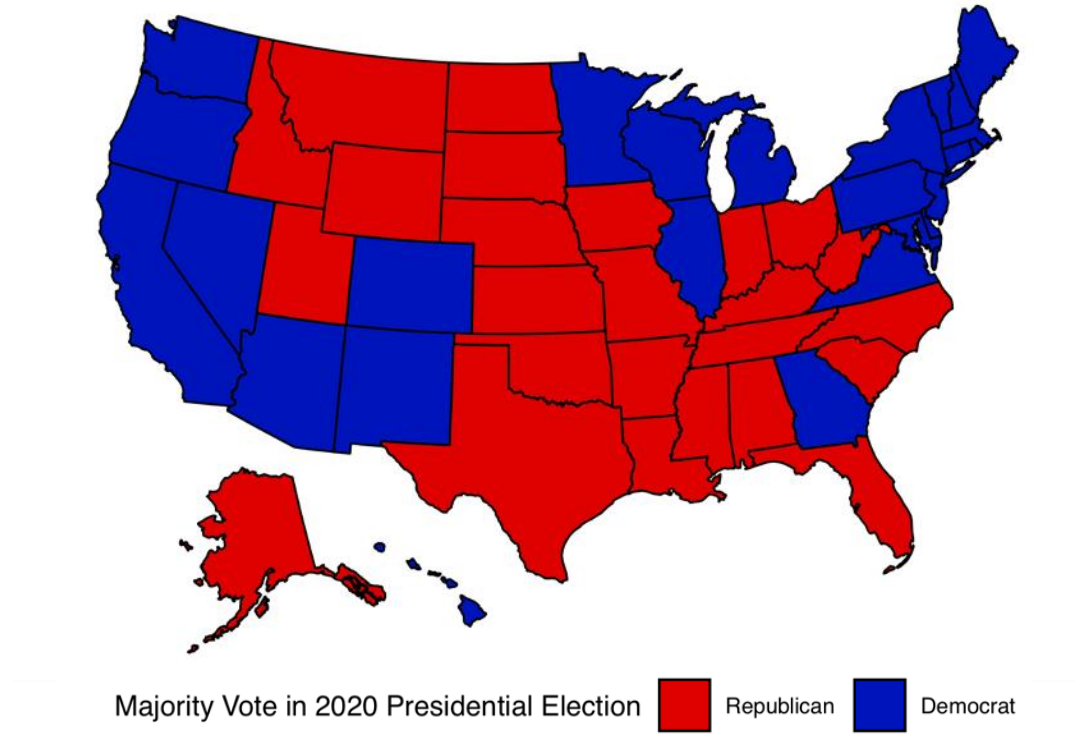


Figure 15. Geographic distribution of states' majority vote in the 2020 presidential election

Table 4. Statistical differences in state-level factors by those states who reported ethnicity separate from race or included with race for case data

Variable	States that separated race and ethnicity variables for cases n(%) or mean(SD) Yes	No	p-value
State governor political party			0.0553
Republican	16 (67)	8 (38)	
Democrat	8 (33)	13 (62)	
State 2020 presidential election majority vote			**0.0006 ***
Trump	17 (71)	4 (19)	
Biden	7 (29)	17 (81)	
State white population proportion	80.0 (10.2)	79.0 (8.7)	0.7233
State Hispanic population proportion	8.0 (5.2)	18.8 (12.5)	**0.001**
State median household income	57.0 (6.4)	69.8 (8.9)	**<0.0001**
State per capita public health spending	96.0 (49.9)	103.2 (39.7)	0.6014
Proportion of state population under age 65 without health insurance	10.9 (3.4)	9.5 (3.9)	0.2203

Note: * Fisher's exact test

Table 5. Statistical differences in state-level factors by those states who reported ethnicity separate from race or included with race for death data

Variable	States that separated race and ethnicity variables for deaths n(%) or mean(SD) Yes	No	p-value
State governor political party			0.0720
Republican	15 (65)	8 (38)	
Democrat	8 (35)	13 (62)	
State 2020 presidential election majority vote			**0.0009**
Trump	17 (74)	5 (24)	
Biden	6 (26)	16 (76)	
State white population proportion	78.7 (10.6)	79.5 (8.9)	0.8053
State Hispanic population proportion	7.3 (3.2)	18.4 (12.8)	**0.0008**
State median household income	56.7 (6.7)	68.9 (9.4)	**<0.0001**
State per capita public health spending	96.7 (50.5)	104.3 (40.0)	0.5882
Proportion of state population under age 65 without health insurance	10.7 (3.2)	9.7 (3.9)	0.3512

4.4 Univariate logistic regression of states' COVID-19 case data

In univariate logistic regression modelling of cases, three characteristics were significantly associated with the odds of reporting ethnicity separate from race: proportion of state population that is Hispanic (OR 0.84; 95%CI 0.73, 0.95), median income (OR 0.81; 95%CI 0.72, 0.92), and presidential majority vote (OR 0.10; 95%CI 0.02, 0.39) (Table 6). The proportion of the state that is white (OR 1.01), the political party of the state governor (OR 0.31), the state per capita public health spending (OR 0.97), and the proportion of the population under 65 without health insurance (OR 1.11) were not univariately statistically associated with the odds of reporting race and ethnicity separately for cases.

4.5 Multivariate logistic regression of states' COVID-19 case data

In multivariate regression modeling, two characteristics remained significant: Hispanic population proportion (OR 0.81; 95%CI 0.68, 0.97) and median income (OR 0.70 95%CI 0.52, 0.92) were statistically significant when taking each of the other characteristics into account (Table 6). Therefore, while holding the other variables constant, a one percent increase in Hispanic population proportion decreases the odds of reporting ethnicity separate from race by 19%. While accounting for the other variables, an increase in median income by $1,000 decreases the odds of reporting ethnicity separate from race by 30%.

Table 6. Results from univariate and multivariate logistic regression models analyzing case data. The modeled outcome is that ethnicity is being reported separate from race

Variable	Univariate Logistic Regression		Multivariate Logistic Regression	
	Estimate	**OR (CI)**	**Estimate**	**OR (CI)**
State governor political party Reference = Republican	-1.1786	0.31 (0.09, 1.05)	0.4819	1.62 (0.03, 98.59)
State 2020 presidential election majority vote Reference = Trump	-2.3342	**0.10 (0.02,0.39)**	-1.4110	0.24 (0.01, 16.79)
State white population proportion	0.0117	1.01 (0.95, 1.08)	-0.0847	0.92 (0.78, 1.09)
State Hispanic population proportion	-0.1808	**0.84 (0.73, 0.95)**	-0.2095	**0.81 (0.68, 0.97)**
State median household income	-0.2120	**0.81 (0.72, 0.92)**	-0.3636	**0.70 (0.52, 0.92)**
State per capita public health spending	-0.00363	1.0 (0.98, 1.01)	0.0115	1.01 (0.99, 1.04)
Proportion of state population under age 65 without health insurance	0.1076	1.11 (0.94, 1.32)	-0.0615	0.94 (0.53, 1.67)

4.6 Univariate logistic regression of states' COVID-19 death data

Logistic regression models for ethnicity reporting of deaths were similar to the models for cases (Table 7). Proportion of state population that is Hispanic (OR 0.79; 95%CI 0.67, 0.93), median income (OR 0.83; 95%CI 0.75, 0.93), and presidential majority vote (OR 0.11; 95%CI 0.28, 0.43) were all univariately statistically significantly associated with the odds of reporting ethnicity separate from race. The proportion of the state that is white (OR 0.99), the political party of the state governor (OR 0.33), the state per capita public health spending (OR 1.0), and the

proportion of the population under 65 without health insurance (OR 1.01) were not univariately statistically associated with the odds of reporting race and ethnicity separately for deaths.

4.7 Multivariate logistic regression of states' COVID-19 death data

As a multivariate logistic regression, accounting for the other variables, the only significant factor was median income (OR 0.78; 95%CI 0.65, 0.94) (Table 7). While holding the other factors constant, a one-unit increase of $1,000 in median income decreases the odds of reporting ethnicity separate from race by 22%.

Table 7. Results from univariate and multivariate logistic regression models analyzing death data. The modeled outcome is that ethnicity is being reported separate from race

Variable	Univariate Logistic Regression		Multivariate Logistic Regression	
	Estimate	**OR (CI)**	**Estimate**	**OR (CI)**
State governor political party Reference = Republican	-1.1141	0.33 (0.96, 1.22)	-1.1972	0.30 (0.02, 4.67)
State 2020 presidential election majority vote Reference = Trump	-2.2046	**0.11 (0.28, 0.43)**	-0.1352	0.87 (0.03, 26.31)
State white population proportion	-0.00799	0.99 (0.93, 1.06)	-0.0814	0.92 (0.80, 1.06)
State Hispanic population proportion	-0.2335	**0.79 (0.67, 0.93)**	-0.1529	0.86 (0.73, 1.01)
State median household income	-0.1837	**0.83 (0.75, 0.93)**	-0.2515	**0.78 (0.65, 0.94)**
State per capita public health spending	-0.00379	1.0 (0.98, 1.01)	0.00516	1.01 (0.98, 1.03)
Proportion of state population under age 65 without health insurance	0.0839	1.09 (0.91, 1.29)	-0.1458	0.86 (0.53,1.42)

5.0 Discussion

This study aimed to provide a compiled source of all fifty states in the United States' publicly reported indicators of race and ethnicity for COVID-19 cases, hospitalizations, and deaths, including whether the state was reporting ethnicity separate from or included with race. These findings were compiled in a figure to compare what race and ethnicity indicators are being reported between states and with federal race and ethnicity reporting standards. This study also aimed to explore potential associations of state sociodemographic, health, and political factors with that state's reporting of ethnicity separate from or included with race for COVID-19 cases and deaths. In the univariate logistic regressions for both COVID-19 case and death data, three of these factors were statistically significantly associated with decreased odds of reporting ethnicity separate from race: higher state population proportion who are Hispanic, higher state median income, and 2020 presidential election majority vote for Joe Biden. State population proportion who are Hispanic and state median income remained statistically associated in multivariate logistic regression, where all other factors were adjusted for, for cases. Only state median income remained statistically significant in the multivariate logistic regression for deaths.

This study allows for three major conclusions based on the main findings. First, the results of the univariate and multivariate regression models are indicative of the extent of publicly reporting of COVID-19 race and ethnicity data for cases and deaths to not be completely random.

Second, there is a lot of COVID-19 race and ethnicity data missing from the public domain, both in terms of quality and quantity. The concerns of the quantity of the data are that results can be unstable and subject to change with more complete data. The concerns with the quality of the data are that entire groups of people are likely being miscategorized, resulting in poorer

understanding of their health. Small state population proportions of minority population might make including a separate option for them seem unnecessary, but even in states with a fraction of one percent being a certain racial group, this could be a potential few thousand individuals possibly facing improper categorization. Both of these concerns, quality and quantity of race and ethnicity data for COVID-19 can be improved, though might likely require two different approaches.

The third main conclusion is that the proportion of a state's population who are Hispanic may be associated with that same state's recognition of Hispanic ethnicity as a factor separate from race. In this study, Hispanic population proportion was statistically significantly associated with the odds of reporting ethnicity separate from race for three of the four logistic regression analyses. The fourth, the multivariate analysis with death data, had the same consistent relationship as the other three analyses, but was not statistically significant. The direction of the association was an inverse relationship, meaning that higher state Hispanic population proportions were associated with lower odds of reporting ethnicity as a separate variable from race. There may have been expectation that higher Hispanic population proportions would be associated with reporting race and ethnicity as two separate variables, but this study found an inverse relationship between the two factors. This inverse relationship should be explored further, as it could be a meaningful relationship, or it could be explained by confounding by other factors.

Given the nature of this novel pandemic, there is limited research to relate the findings to. There are, however, some findings from previous outbreaks that may influence understandings in this present pandemic. The burden of various other outbreaks and epidemics in the United States have been centered on racial and/or ethnic minority populations; the current spread of HIV being significantly higher in the Black population than in the white population is just one example. There is evidence from the 2003 SARS outbreak that increasing accurate knowledge is associated with

decreased bias towards Asians, which could inform increasing knowledge about the need for more comprehensive race and ethnicity data collection.[13] It has been acknowledged that it can be complex and difficult while trying to control spread of infection in a pandemic to both effectively protect public health and prevent stigma and discrimination.[13,18] As vaccines for COVID-19 become more available, vaccine distribution from the 2009 H1N1 pandemic can inform how disparities in race of vaccination opportunities are possible, but preventable from being repeated.[97]

Socioeconomic status and its various related measurements tend to be frequently associated with health status and outcomes. As a social determinant of health, socioeconomic status is an important factor to consider, though a number of studies conducted since the start of the pandemic believe it may be a much more important factor than it has been considered. There are suggestions of socioeconomic status being available alongside race data in COVID-19 to more accurately explain racial disparities.[19,98] A report from the National Academy of Science suggests that up to 70% of health outcomes are attributable to socioeconomic factors.[32] Though socioeconomic status is not frequently collected in these same contexts of clinical settings, having this complementary socioeconomic status data could benefit in clarifying if and how race and socioeconomic status coexist to affect health.[19] In turn, the COVID-19 pandemic's effects on the economy have exacerbated poverty rates in communities who are already facing economic difficulties, communities who are disproportionately Black and Brown.[32] These findings support the consistent statistical significance of state median income in each of the univariate and multivariate regression models in the present study.

As COVID-19 has illuminated racial inequities in health, there are concurrent epidemics in the United States that are also intertwined with race: opioid drug abuse and HIV. One study comparing similarities between HIV and COVID-19 in the United States suggests that these three

overlapping epidemics in the United States could not occur the way they are by chance; rather, they are all rooted in systemic racism.[32] This supports the implications of the present study of systemic racism's involvement in COVID-19 data.

The missingness of these race and ethnicity data across states was extensive in some cases, but it is similar to findings from one study that investigated missingness of COVID-19 data in the United States between April and September 2020.[99] This study, which was assessing trends in completeness of individual data for COVID-19 cases during the first six or so months of the pandemic included a section of race and ethnicity completeness. This study found that, at the time, there was approximately 43.1% of race/ethnicity data missing; the present study found an approximate missing proportion of 27.4% as of February 2021. This may be indicative of continued improvement of completing race and ethnicity fields for COVID-19 cases. However, there is still immense variation in the range of missingness between states, which was also the case during the April to September 2020 study period; completion of race and ethnicity data at that time ranged from under 50% to over 75%, which are similar figures to the missingness observed in the present study.

There are potential explanations for the methodology of what indicators are or are not being reported by state health departments that were not explored in these analyses. This includes the possibility that health departments simply do not have the manpower, funding, or technology to have scaled up their race and ethnicity data collection and reporting. CDC Director, Dr. Robert Redfield attributed scarcity of these data to "antiquated technology in public health systems," making this issue at the state level more plausible.[25]

Besides these logistical concerns of implementing increased surveillance and data collection, the theory of simply adopting the federal race and ethnicity reporting guidelines at the

state level for COVID-19 data reporting could make comparisons between COVID-19 and published data from the Census much more meaningful. There are, however, potential objections to this. While the goal of these guidelines may be to have a clear and relatively simple method for appropriately categorizing people in the United States by race, it still is not as comprehensive and what is really necessary, especially with an increasingly racially diverse population. There have been concerns, for example, that the 'white' race category may be oversimplifying the race of descendants of the Middle East and North Africa.[84] Likewise, a report authored by two Mexican Americans suggests that Hispanic ethnicity is an obvious choice for them, but race is less obvious.[100] Since their race may be more accurately categorized as Mestizo, a Hispanic person of both European and indigenous American descent, the choice between white, American Indian, or multiracial or other, if an option, is very unclear. Their argument states that for many Hispanic/Latino individuals, choosing a race other than this ethnicity option based on available options in the United States may actually be further muddying the understanding of different racial groups in the United States.[100]

5.1 Practical Implications

Practical implications of these findings are that they have the potential to help inform choices in governments or departments of health that may be unintentionally decreasing the quality of COVID-19 race and ethnicity data. Further research to support these findings is necessary, but this study nonetheless provides an avenue for explanatory studies to make stronger conclusions in agreement or disagreement with these findings.

5.2 Strengths and limitations

This study explored potential associations between state sociodemographic, health, and political factors and the way that states are publicly reporting ethnicity of their COVID-19 case and death data. The descriptive nature of compiling all indicators that all states are reporting is valuable and can be useful in informing hypotheses in future research. Though there is potential for these race and ethnicity indicators to change by states in the future, the wrangling of hundreds of indicators provides a snapshot for what was being publicly reported in January/February 2021. This has the potential to benefit researchers in the future to explore temporal changes in states' indicators, something this study did not incorporate.

Determining any temporality in changing the methodology of reported race and ethnicity data for COVID-19 cases, hospitalizations, and/or deaths was beyond the scope of this study. Federal mandates, as well as state's choices, may have shifted how or what data were being reported throughout the pandemic, but this is not captured in analysis. There are various societal factors that have occurred throughout the pandemic that may have influenced altering race and ethnicity data reporting. As an example, in late May 2020, the murder of George Floyd in Minneapolis, Minnesota sparked a summer of protests and riots across the country. While literature has been published commenting on the potential effects of these on COVID-19 spread, little is known on how, if at all, this influenced the reporting standards of COVID-19 in any states.[25,32,101,102]

The novelty of this pandemic itself is a limiting factor to this research in that significant variables may have been overlooked because their significance to this pandemic has not yet been uncovered. The covariates that were included in these models could have unrecognized collinearity, which, if present, could limit the interpretation of the model coefficients. This research

is also limited by the sample size of fifty, though this is unavoidable when using each of the United States as its own subject. It is also possible that regional, county, or jurisdictional health departments within a state are collecting and reporting their race and ethnicity data differently than the state is, whether that be more or less comprehensive. Though this study was specifically at the state level, differences in level of granularity could contribute to more missing data than accounted for. There is also the fact that this research used only publicly reported data; it is likely that states' publicly available data is a subset of what states actually collect.

5.3 Future directions

Future work can be built from this study, especially in further understanding of what other state factors may be influencing data collection and reporting methodologies for COVID-19 race and ethnicity data. In addition, the compiled reported state indicators could be used to rank the states in how appropriate or comprehensive their reporting methods are, which could be analyzed for associations with that state's handle on the pandemic. While this study inherently could not draw conclusions about causal relationships, it can suggest that statistically significant findings may benefit from further hypothesis-drive research.

5.4 Conclusion

Overall, the concerns— and now visualizations— of the suboptimal quality of the race and ethnicity data of COVID-19 cases, hospitalizations, and deaths across the United States precludes

effective change through evidence-based interventions or policy. Knowing that minority communities are being disproportionately affected by COVID-19, but not attempting to collect or report adequate data for accurate estimations to guide interventions is a public health injustice to those populations who are being most disproportionately affected. This pandemic, as unfortunate as its effects have been, is providing an opportunity to finally and substantively address the racial health inequities that plague minority and vulnerable communities around the country. Without more appropriate evidence, it is impossible to conclude that poor quality data collection and reporting is caused by implicit or explicit racial biases in the United States, but it does not mean that the racial and xenophobic contexts surrounding the pandemic in the United States can be ignored. This context is vital to recognize, as racism and xenophobia continue to be significant public health concerns.

Bibliography

1. Timeline of WHO's response to COVID-19. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline?gclid=CjwKCAjwx6WDBhBQEiwA_dP8rbY3UHlTrOKpjH1nOC3-0E6V9Wml8lnWTMkxu_4huMztEYcrjkzLthoC1hcQAvD_BwE#. Accessed March 2021.

2. WHO issues best practices for naming new human infectious diseases. https://www.who.int/news/item/08-05-2015-who-issues-best-practices-for-naming-new-human-infectious-diseases. Accessed March 2021.

3. Neerukonda SN, Katneni U. A Review on SARS-CoV-2 Virology, Pathophysiology, Animal Models, and Anti-Viral Interventions. *Pathogens*. 2020;9(6). doi:10.3390/pathogens9060426

4. Coronavirus Disease 2019 (COVID-19). CDC. https://www.cdc.gov/dotw/covid-19/index.html. Accessed March 2021.

5. Coronavirus. https://www.who.int/health-topics/coronavirus. Accessed March 2021.

6. Trump DJ. *Proclamation 9994 of March 13, 2020: Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID–19) Outbreak*. Federal Register; 2020. http://www.govinfo.gov/content/pkg/FR-2020-03-18/pdf/2020-05794.pdf. Accessed April 2021.

7. 116th Congress. *FAMILIES FIRST CORONAVIRUS RESPONSE ACT*.; 2020. https://www.govinfo.gov/content/pkg/PLAW-116publ127/pdf/PLAW-116publ127.pdf. Accessed April 2021.

8. 116th Congress. *Coronavirus Aid, Relief, And Economic Security Act*.; 2020. https://www.govtrack.us/congress/bills/116/hr748/text. Accessed April 2021.

9. van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *Lancet*. 2020;395(10232):1243-1244. doi:10.1016/S0140-6736(20)30893-X

10. Hart PS, Chinn S, Soroka S. Politicization and Polarization in COVID-19 News Coverage. *Sci Commun*. August 2020:107554702095073. doi:10.1177/1075547020950735

11. Republicans and Democrats sharply divided in views of news media's coronavirus coverage. Pew Research Center. https://www.journalism.org/2020/05/08/americans-are-more-likely-than-not-to-think-the-news-media-are-fulfilling-key-roles-during-the-coronavirus-outbreak-but-partisans-are-starkly-divided/. Accessed April 2021.

12. Christensen SR, Pilling EB, Eyring JB, Dickerson G, Sloan CD, Magnusson BM. Political and personal reactions to COVID-19 during initial weeks of social distancing in the United States. *PLoS ONE*. 2020;15(9):e0239693. doi:10.1371/journal.pone.0239693

13. Dhanani LY, Franz B. Unexpected public health consequences of the COVID-19 pandemic: a national survey examining anti-Asian attitudes in the USA. *Int J Public Health*. 2020;65(6):747-754. doi:10.1007/s00038-020-01440-0

14. Anti-Chinese Rhetoric Tied to Racism against Asian Americans Stop AAPI Hate Report - CAA. https://caasf.org/2020/06/anti-chinese-rhetoric-tied-to-racism-against-asian-americans-stop-aapi-hate-report/. Accessed March 2021.

15. Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety - The New York Times. https://www.nytimes.com/2020/03/23/us/chinese-coronavirus-racist-attacks.html. Accessed April 4, 2021.

16. Darling-Hammond S, Michaels EK, Allen AM, et al. After "The China Virus" Went Viral: Racially Charged Coronavirus Coverage and Trends in Bias Against Asian Americans. *Health Educ Behav*. 2020;47(6):870-879. doi:10.1177/1090198120957949

17. Tessler H, Choi M, Kao G. The Anxiety of Being Asian American: Hate Crimes and Negative Biases During the COVID-19 Pandemic. *Am J Crim Justice*. June 2020:1-11. doi:10.1007/s12103-020-09541-5

18. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Infectious Diseases/SARS Community Outreach Team. Fear and stigma: the epidemic within the SARS outbreak. *Emerging Infect Dis*. 2004;10(2):358-363. doi:10.3201/eid1002.030750

19. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 - Caution and Context. *N Engl J Med*. 2020;383(3):201-203. doi:10.1056/NEJMp2012910

20. The Coronavirus's Xenophobia Problem - The Atlantic. https://www.theatlantic.com/international/archive/2020/03/coronavirus-covid19-xenophobia-racism/607816/. Accessed March 2021.

21. Nguyen TT, Criss S, Dwivedi P, et al. Exploring U.S. Shifts in Anti-Asian Sentiment with the Emergence of COVID-19. *Int J Environ Res Public Health*. 2020;17(19). doi:10.3390/ijerph17197032

22. Zannetou S, Baumgartner J, Finkelstein J, Goldenberg A. Weaponized Information Outbreak: a case study on COVID-19, bioweapon myths, and the Asian conspiracy meme—Network Contagion Research Institute. https://millercenter.rutgers.edu/wp-content/uploads/2020/10/NCRI-White-Paper-COVID-19-13-Apr-2020.pdf. Accessed March 2021.

23. US push to include "Wuhan virus" language in G7 joint statement fractures alliance - CNNPolitics. https://www.cnn.com/2020/03/25/politics/g7-coronavirus-statement/index.html. Accessed April 2021.

24. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA*. 2020;323(24):2466-2467. doi:10.1001/jama.2020.8598

25. Trump Administration Sets Demographic Requirements for Coronavirus Reports - The New York Times. https://www.nytimes.com/2020/06/04/us/politics/coronavirus-infection-demographics.html. Accessed March 2021.

26. Krieger N, Testa C, Hanage WP, Chen JT. US racial and ethnic data for COVID-19 cases: still missing in action. *Lancet*. 2020;396(10261):e81. doi:10.1016/S0140-6736(20)32220-0

27. Mackey K, Ayers CK, Kondo KK, et al. Racial and Ethnic Disparities in COVID-19-Related Infections, Hospitalizations, and Deaths: A Systematic Review. *Ann Intern Med*. 174(3):362-373. doi:10.7326/M20-6306

28. Hatcher SM, Agnew-Brune C, Anderson M, et al. COVID-19 Among American Indian and Alaska Native Persons - 23 States, January 31-July 3, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(34):1166-1169. doi:10.15585/mmwr.mm6934e1

29. Cunningham GB, Wigfall LT. Race, explicit racial attitudes, implicit racial attitudes, and COVID-19 cases and deaths: An analysis of counties in the United States. *PLoS ONE*. 2020;15(11):e0242044. doi:10.1371/journal.pone.0242044

30. Kaufman HW, Niles JK, Nash DB. Disparities in SARS-CoV-2 Positivity Rates: Associations with Race and Ethnicity. *Popul Health Manag*. 24(1):20-26. doi:10.1089/pop.2020.0163

31. Labgold K, Hamid S, Shah S, et al. Estimating the Unknown: Greater Racial and Ethnic Disparities in COVID-19 Burden After Accounting for Missing Race and Ethnicity Data. *Epidemiology*. 2021;32(2):157-161. doi:10.1097/EDE.0000000000001314

32. Millett GA. New pathogen, same disparities: why COVID-19 and HIV remain prevalent in U.S. communities of colour and implications for ending the HIV epidemic. *J Int AIDS Soc*. 2020;23(11):e25639. doi:10.1002/jia2.25639

33. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Accessed March 27, 2021.

34. Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. U.S. Department of the Interior. https://www.doi.gov/pmb/eeo/directives/race-data. Accessed March 2021.

35. Standards for the Classification of Federal Data on Race and Ethnicity. The White House. https://obamawhitehouse.archives.gov/omb/fedreg_race-ethnicity. Accessed March 2021.

36. Alabama's COVID-19 Dashboard Hub. https://alpublichealth.maps.arcgis.com/apps/MapSeries/index.html?appid=d84846411471404c83313bfe7ab2a367. Accessed February 2021.

37. Alaska Department of Health and Social Services Coronavirus Response Hub. https://alaska-coronavirus-vaccine-outreach-alaska-dhss.hub.arcgis.com/. Accessed February 2021.

38. Arizona Department of Health Services. https://www.azdhs.gov/covid19/data/index.php. Accessed February 2021.

39. The Arkansas Department of Health COVID-19 Update. https://experience.arcgis.com/experience/c2ef4a4fcbe5458fbf2e48a21e4fece9. Accessed February 2021.

40. COVID-19: California Case Statistics. https://public.tableau.com/profile/ca.open.data#!/vizhome/COVID-19CasesDashboardv2_0/CaseStatistics. Accessed February 2021.

41. Colorado COVID-19 Updates. https://covid19.colorado.gov/data. Accessed February 2021.

42. COVID-19 Daily Report | Connecticut Data. https://data.ct.gov/stories/s/q5as-kyim. Accessed February 2021.

43. State of Delaware's Latest Status. https://myhealthycommunity.dhss.delaware.gov/locations/state. Accessed February 2021.

44. Florida's COVID-19 Data and Surveillance Dashboard. https://experience.arcgis.com/experience/96dd742462124fa0b38ddedb9b25e429. Accessed February 2021.

45. COVID-19 Status Report | Georgia Department of Public Health. **Error! Hyperlink reference not valid.**. Accessed February 2021.

46. Disease Outbreak Control Division Hawaii COVID-19 Data. https://health.hawaii.gov/coronavirusdisease2019/what-you-should-know/current-situation-in-hawaii/. Accessed February 2021.

47. Idaho Department of Health and Welfare's COVID-19 Dashboard. **Error! Hyperlink reference not valid.**. Accessed February 2021.

48. Illinois Department of Health Coronavirus Disease 2019 (COVID-19). **Error! Hyperlink reference not valid.**. Accessed February 2021.

49. Indiana COVID-19 Dashboard and Map. https://www.coronavirus.in.gov/2393.htm. Accessed February 2021.

50. COVID-19 in Iowa. https://coronavirus.iowa.gov/. Accessed February 2021

51. COVID-19 Cases in Kansas. https://www.coronavirus.kdheks.gov/160/COVID-19-in-Kansas. Accessed February 2021.

52. KDPH COVID-19 Dashboard. https://kygeonet.maps.arcgis.com/apps/opsdashboard/index.html#/543ac64bc40445918cf8bc34dc40e334. Accessed February 2021.

53. Louisiana Department of Health COVID-19. https://ldh.la.gov/coronavirus/. Accessed February 2021.

54. Maine Division of Disease Surveillance Coronavirus Disease 2019 (COVID-19) Updates and Information. https://www.maine.gov/dhhs/mecdc/infectious-disease/epi/airborne/coronavirus/data.shtml. Accessed February 2021.

55. Maryland COVID-19 Data Dashboard. https://coronavirus.maryland.gov/. Accessed February 2021.

56. Massachusetts Department of Public Health COVID-19 Dashboard. https://www.mass.gov/info-details/covid-19-response-reporting. Accessed February 2021.

57. Coronavirus - Michigan Data. https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html. Accessed February 2021.

58. Situation Update for COVID-19 - Minnesota Department of Health. **Error! Hyperlink reference not valid.**. Accessed February 2021.

59. Coronavirus COVID-19 - Mississippi State Department of Health. https://msdh.ms.gov/msdhsite/_static/14,0,420.html. Accessed February 2021.

60. COVID-19 in Missouri at a Glance. https://showmestrong.mo.gov/data/public-health/. Accessed February 2021.

61. Montana Response: COVID-19. https://montana.maps.arcgis.com/apps/MapSeries/index.html?appid=7c34f3412536439491adcc2103421d4b. Accessed February 2021.

62. Nebraska Covid-19 Dashboard. https://experience.arcgis.com/experience/ece0db09da4d4ca68252c3967aa1e9dd/page/page_0/. Accessed February 2021.

63. Coronavirus (COVID-19) in Nevada | Nevada Health Response. https://nvhealthresponse.nv.gov/. Accessed February 2021.

64. NH COVID-19 Response. https://www.covid19.nh.gov/. Accessed February 2021.

65. New Jersey COVID-19 Information Hub. https://covid19.nj.gov/. Accessed February 2021.

66. NMDOH COVID-19 Public Dashboard. https://cvprovider.nmhealth.org/public-dashboard.html. Accessed February 2021.

67. NYS-COVID19-Tracker. https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n. Accessed February 2021.

68. COVID-19 North Carolina Dashboard. https://covid19.ncdhhs.gov/dashboard. Accessed February 2021.

69. North Dakota Coronavirus Cases. https://www.health.nd.gov/diseases-conditions/coronavirus/north-dakota-coronavirus-cases. Accessed February 2021.

70. Ohio Department of Health COVID-19 Dashboard. https://coronavirus.ohio.gov/wps/portal/gov/covid-19/dashboards. Accessed February 2021.

71. Oklahoma State Department of Health COVID-19 Cases. https://oklahoma.gov/covid19.html. Accessed February 2021.

72. Oregon Health Authority's COVID-19 Data Dashboards. https://public.tableau.com/profile/oregon.health.authority.covid.19#!/vizhome/OregonsCOVID-19DataDashboards-TableofContents/TableofContentsStatewide. Accessed February 2021.

73. COVID-19 Data for Pennsylvania. https://www.health.pa.gov/topics/disease/coronavirus/Pages/Cases.aspx. Accessed February 2021.

74. Rhode Island COVID-19 Response Data. https://ri-department-of-health-covid-19-data-rihealth.hub.arcgis.com/. Accessed February 2021.

75. SC Testing Data & Projections (COVID-19). https://scdhec.gov/covid19/sc-testing-data-projections-covid-19. Accessed February 2021.

76. South Dakota COVID-19 Dashboard. https://doh.sd.gov/COVID/Dashboard.aspx. Accessed February 2021.

77. Tennessee Epidemiology and Surveillance Data. https://www.tn.gov/health/cedep/ncov/data.html. Accessed February 2021.

78. Texas Department of State Health Services. https://txdshs.maps.arcgis.com/apps/dashboards/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. Accessed February 2021.

79. Utah Department of Health COVID-19 Data. https://coronavirus.utah.gov/case-counts/. Accessed February 2021.

80. Vermont Dashboard. https://www.healthvermont.gov/covid-19/current-activity/vermont-dashboard. Accessed February 2021.

81. COVID-19 In Virginia. https://www.vdh.virginia.gov/coronavirus/covid-19-in-virginia/. Accessed February 2021.

82. Washington State Department of Health COVID-19 Data Dashboard. **Error! Hyperlink reference not valid.**. Accessed February 2021.

83. West Virginia COVID-19. https://dhhr.wv.gov/COVID-19/Pages/default.aspx. Accessed February 2021.

84. COVID-19: Wisconsin Summary Data. https://www.dhs.wisconsin.gov/covid-19/data.htm. Accessed February 2021.

85. COVID-19 State and County Dashboards - Wyoming Department of Health. https://health.wyo.gov/publichealth/infectious-disease-epidemiology-unit/disease/novel-coronavirus/covid-19-state-and-county-dashboards/. Accessed February 2021.

86. Neelon B, Mutiso F, Mueller NT, Pearce JL, Benjamin-Neelon SE. Associations between governor political affiliation and COVID-19 cases and deaths in the United States. *medRxiv*. October 2020. doi:10.1101/2020.10.08.20209619

87. Yehia BR, Winegar A, Fogel R, et al. Association of Race With Mortality Among Patients Hospitalized With Coronavirus Disease 2019 (COVID-19) at 92 US Hospitals. *JAMA Netw Open*. 2020;3(8):e2018039. doi:10.1001/jamanetworkopen.2020.18039

88. U.S. Census Bureau QuickFacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045219. Accessed April 2021.

89. American Community Survey Data. https://www.census.gov/programs-surveys/acs/data.html. Accessed April 2021.

90. Himmelstein DU, Woolhandler S. Public health's falling share of US health spending. *Am J Public Health*. 2016;106(1):56-57. doi:10.2105/AJPH.2015.302908

91. Resnick B, Bishai D, Leider JP, Colrick I. State Health Expenditure Dataset (SHED), 2000-2013. Inter-university Consortium for Political and Social Research [distributor], 2017. doi:10.3886/ICPSR36741.v1

92. Hollowed-Out Public Health System Faces More Cuts Amid Virus. Kaiser Health News. https://khn.org/news/us-public-health-system-underfunded-under-threat-faces-more-cuts-amid-covid-pandemic/. Accessed April 2021.

93. Sehra ST, Fundin S, Lavery C, Baker JF. Differences in race and other state-level characteristics and associations with mortality from COVID-19 infection. *J Med Virol*. 2020;92(11):2406-2408. doi:10.1002/jmv.26095

94. Health Insurance Coverage in the United States: 2019. https://www.census.gov/library/publications/2020/demo/p60-271.html. Accessed April 2021.

95. Governors - National Governors Association. https://www.nga.org/governors/. Accessed April 2021.

96. 2020 Electoral College Results. National Archives. https://www.archives.gov/electoral-college/2020. Accessed April 2021.

97. Burger AE, Reither EN, Mamelund S-E, Lim S. Black-white disparities in 2009 H1N1 vaccination among adults in the United States: A cautionary tale for the COVID-19 pandemic. *Vaccine*. 2021;39(6):943-951. doi:10.1016/j.vaccine.2020.12.069

98. Roberts JA, David ME. The Pandemic within a Pandemic: Testing a Sequential Mediation Model to Better Understand Racial/Ethnic Disparities in COVID-19 Preventive Behavior. *Healthcare (Basel)*. 2021;9(2). doi:10.3390/healthcare9020230

99. Gold JAW, DeCuir J, Coyle JP, et al. COVID-19 Case Surveillance: Trends in Person-Level Case Data Completeness, United States, April 5-September 30, 2020. *Public Health Rep*. March 2021:333549211006973. doi:10.1177/00333549211006973

100. Separating Race from Ethnicity in Surveys Risks an Inaccurate Picture of the Latinx Community. Urban Institute. https://www.urban.org/urban-wire/separating-race-ethnicity-surveys-risks-inaccurate-picture-latinx-community. Accessed April 2021.

101. Fasehun L-KO. Reviewing COVID-19 Modelling amidst Recent United States Protests. *Ann Glob Health*. 2020;86(1):76. doi:10.5334/aogh.2970

102. Valentine R, Valentine D, Valentine JL. Relationship of George Floyd protests to increases in COVID-19 cases using event study methodology. *J Public Health (Oxf)*. 2020;42(4):696-697. doi:10.1093/pubmed/fdaa127