Racial Disparities in COVID-19 Cases and Deaths: The Role of Socio-Economic vs. Health-Related Factors*

Ljubica Ristovska

April 2023

Abstract

At the onset of the COVID-19 pandemic, non-white individuals were 3-4 times more likely to contract COVID-19 and 1.5-2 times more likely to die from COVID-19. Using comprehensive data on COVID-19 cases and deaths by race reported by the CDC, aggregated to counties and combined with extensive county data on socio-economic, health-related, and environmental factors by race, we fin d that education rates among Black individuals are strongly negatively correlated with Black case rates, but not for any other non-white race. Additionally, rates of public transportation use and household density among Hispanic individuals are important correlates of Hispanic case and mortality rates, but less so for other racial groups. Health-related factors play a smaller role in explaining observed racial disparities in COVID-19 spread, but pre-existing comorbidities are strongly positively correlated with mortality among Black individuals infected with COVID-19.

*ristovska@g.harvard.edu. This paper was previously circulated under the title "Racial Disparities in COVID-19 Cases and Deaths: Theories and Evidence". I am extremely grateful to Jason Abaluck, Raj Chetty, David Cutler, Edward Glaeser, Larry Katz, Amanda Kowalski, Sagar Saxena, Ben Sprung-Keyser, and Helen Wang for helpful discussions and comments.

1 Introduction

Throughout the COVID-19 pandemic in the U.S., large racial disparities in cases and deaths have emerged.[1] The reasons behind the high COVID-19 burden among non-white individuals are numerous. Non-white individuals are more likely to live in densely populated geographic areas or in denser living quarters due to housing segregation, high housing costs, or higher rates of incarceration, which may facilitate COVID-19 transmission. On the other hand, many COVID-19 outbreaks have occurred in nursing homes, and most nursing home residents are white and non-Hispanic (**?**).[2] Non-white individuals are more likely to work in essential occupations or occupations for which working from home is not possible (**?**). Non-white individuals may live further away from grocery stores, medical facilities, and workplaces, resulting in increased use of public transportation and higher risk of COVID-19 infection. These differences might further be amplified by or correlated with differences in income by race. Conditional on these factors, non-white individuals might be less likely to adhere to social distancing recommendations, less likely to follow guidelines for preventing infections, or be less informed about COVID-19 spread (**?**). Non-white individuals might be disproportionately affected by underlying comorbidities that increase illness severity, leading to higher COVID-19 mortality rates.[3] Pollution has also been shown to be highly correlated with COVID-19 mortality (**?**), and non-white individuals may be disproportionately more likely to live in more polluted areas. Non-white individuals may have less access to proper health care, or have access to facilities of worse quality than white individuals. Finally, racial bias in the health care system, as well as other forms of systemic racism not captured by the aforementioned factors may affect infection spread, illness severity, and health care allocation among non-white populations.

This paper aims to examine the role of large set of socio-economic, health-related, and environmental factors on racial disparities in COVID-19 spread and disease severity across counties in the U.S., with a particular emphasis on disentangling the role of socio-economic vs. health-related factors in the differential COVID-19 spread by race. We use individual-level data on COVID-19 cases and deaths in the U.S., collected by the Center for Disease Control (CDC) and the National Center for Health Statistics and aggregated to age-race-county categories, which we combine with a variety of data sources on household density, population density, income, education, public transportation use, pollution, health care quality, comorbidities, and social distancing at the age-race-county level. We exclude small counties (<100,000 population) since these do not have population denominators by race used for the calculation of COVID-19 case and mortality rates per 100,000. We focus on the initial six months of the COVID-19 pandemic to (a) minimize confounding from endogenous

[1] For an excellent and thorough review of the most recent studies on COVID-19 disparities by race, see Racial Disparities in COVID-19: Key Findings from Available Data and Analysis by the Kaiser Family Foundation.

[2] According to the Centers for Medicaid and Medicare Services, nursing home residents accounted for 3.95% of all COVID-19 cases but 25.31% of all COVID-19 deaths in the U.S. at the beginning of the COVID-19 pandemic.

[3] The CDC reports that asthma, cancer, cerebrovascular disease, chronic kidney, liver, and lung diseases, cystic fibrosis, diabetes, heart conditions, HIV, disabilities, depression, schizophrenia, dementia, obesity, physical inactivity, pregnancy, smoking, immunodeficiencies, tuberculosis, use of corticosteroids or other immunosuppressive medication, and solid organ or blood stem cell transplantation are the comorbidities with the strongest and most consistent evidence on increasing illness severity.

COVID-19 prevention policies at the individual and county level and (b) focus on factors that contribute to the differential initial impact of the pandemic by race that can be addressed by policies implemented *today*.

First, we document stark racial disparities in COVID-19 in the first six months of the COVID-19 pandemic: Black and Hispanic individuals were 2.9 and 3.9 times more likely to contract COVID-19 relative to white individuals and were 2.2 and 1.3 times more likely to die from COVID-19 after adjusting for age. Similarly, after age adjustment, AIAN individuals were 2.9 times more likely to contract COVID-19 and 1.6 times more likely to die from COVID-19 relative to white individuals. While NHPI individuals were 3.2 times more likely to have COVID-19 compared to white individuals, they were less likely to die from COVID-19. Asian individuals also had a higher case and mortality rate relative to white individuals.

We find that after controlling for other socio-economic and health characteristics, both white and non-white individuals living in denser counties and counties with higher race-specific rates of public transportation use have higher COVID-19 case and mortality rates, highlighting the importance of density-related factors in the initial spread of COVID-19. Race-specific average household income is also negatively correlated with case and mortality rates across all racial groups, as are pollution levels.

Public transportation use is particularly strongly correlated with the Hispanic case rate – counties where the share of Hispanic individuals commuting to work using public transportation is 10 percentage points higher are associated with a 13% higher average Hispanic COVID-19 case rate, controlling for other factors. Another factor that is strongly correlated with the Hispanic case rate, but not any other racial groups, is household density. Controlling for other county characteristics, ounties where Hispanic individuals are 10 percentage points more likely to live in dense households, defined as living in group quarters, multigenerational household, or multifamily households are associated with a 11% higher Hispanic COVID-19 case rate. Furthermore, the correlation between education and case and mortality rates among Black individuals is particularly striking – counties where Black individuals are 10 percentage points more likely to have a college degree have, on average, a 10% lower Black case rate after controlling for other county characteristics. The correlation between education and case rates is much smaller for other racial groups. Because most counties with AIAN and NHPI populations are small and thus excluded from our analysis, many estimates for these racial groups are noisy, although population density and average AIAN/NHPI income are statistically significantly correlated with case and mortality rates for these groups.

We also find that health care quality and comorbidity rates are uncorrelated with COVID-19 case and mortality rates across all racial groups. However, conditional on infection, counties with lower health care quality have higher COVID-19 mortality rates for white individuals. Importantly, counties with an additional 10 percentage points of Black individuals with an underlying comorbidity are associated with a 3.75% higher share of Black individuals dying from COVID-19 if infected.

Given the differences in socio-economic, health-related, and environmental characteristics between white and non-white individuals, as well as the differences in correlates for race-specific case and mortality rates, we formalize the contributions of each to the observed case and death rate differentials by race using the Oaxaca-Blinder decomposition (**?**, **?**). We decompose the average differences in COVID-19 case and mortality rates between Black and white and Hispanic and white individuals into three components: one due to differences in average characteristics by race ("differences in endowments"), differences due to differential correlations between characteristics and case/mortality rates across racial groups ("differences in coefficients"), and differences due to the interaction between the levels and coefficients for each characteristic ("differences in interactions").

The Oaxaca-Blinder decomposition suggests that if Black individuals had the same observed characteristics as white individuals, keeping correlations between characteristics and COVID-19 burden constant, the case rate and mortality rate difference between Black and white individuals would decrease by 25% and 45% (respectively) relative to the average. This reduction largely stems from equalizing education and income levels, at to a smaller extent household density, which are particularly strong correlates of case and mortality rates among Black individuals. Similarly, equalizing observed characteristics between Hispanic and white individuals would reduce the average COVID-19 case and mortality rate difference between Hispanic and white individuals by 23% and 26%, respectively, relative to the average observed case and mortality rate differential. This decrease can be attributed to equalizing household density and income between white and Hispanic individuals.

On the other hand, if Black individuals had the same correlation between various factors and case and mortality rates as white individuals, but keeping the levels of characteristics across race as observed, the difference between Black and white individuals would fall by 33% for case rates and would even become negative for mortality rates. This is largely due to the fact that the correlation between population density (and to a smaller extent comorbidities) and case/mortality rates is stronger for Black individuals than for white individuals. Additionally, for Hispanic individuals, having equal correlations between characteristics and case/mortality rates as white individuals would explain away the entire case and mortality rate differential observed between Hispanic and white individuals. This result comes from the fact that population density and household density are highly correlated with case rates and mortality rates among Hispanic individuals.

Few papers have examined the role of a comprehensive set of socio-economic, environmental, and health-related factors in the increased COVID-19 burden among non-white individuals. Several papers explore the relationship between demographic, socio-economic, and health characteristics and COVID-19 cases and deaths and find strong correlations between racial composition, public transportation, and density-related measures and COVID-19 burden across all levels of geography (**?????**). However, these studies do not explicitly focus on the differential COVID-19 case and mortality rates by race. This gap comes from the unavailability of data sources that include both an extensive set of socio-economic and health-related factors. Several studies using individual-

level data from a variety of states and hospital systems have documented racial disparities in COVID-19 infections, hospitalizations, and deaths (**?????**). Some of these studies show that race is uncorrelated with COVID-19 illness severity after controlling for socio-economic and clinical factors; however, most socio-economic characteristics are limited to only education and household income. Our paper addresses this gap by: (a) analyzing a larger set of socio-economic and environmental characteristics, and (b) focusing on separating the role of socio-economic vs. health factors in both COVID-19 spread and COVID-19 illness severity.

The paper proceeds as follows. Section 2 describes the data used in this paper in detail. Section 3 documents racial disparities in COVID-19 nationwide and over time. It also describes differences in demographic, socio-economic, environmental and health characteristics by race. Section 4 explores correlates of COVID-19 case and mortality rates for each racial group. Section 5 presents the results from the Oaxaca-Blinder decomposition for Black and Hispanic individuals relative to white individuals. Finally, Section 6 concludes.

2 Data

2.1 COVID-19 cases and deaths

We obtain the most comprehensive data set on individual-level COVID-19 cases and deaths available in the U.S., collected by the Center for Disease Control (CDC) and the National Center for Health Statistics.[4] The data contains all COVID-19 cases reported to the CDC since the beginning of the pandemic (January 2020) until April 17, 2022, amounting to a total of 71.4 million cases. Each case is dated and reported alongside the state and county of residence of the individual, as well as age, sex, whether the case was hospitalized, whether the case ended up in the ICU, and whether the individual associated with the case report subsequently died from COVID-19.[5] Each case is also associated with race information: white, Black, Asian, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, Hispanic, and other races.[6] After excluding COVID-19 cases with missing race data, our data contains 46 million cases, all of which are included in the descriptive statistics at the national level reported in Section 3.1.

We use the cumulative COVID-19 cases per 100,000 individuals, henceforth referred to as the "case

[4] COVID-19 was added to the Nationally Notifiable Condition List on April 5, 2020 and was classified as "immediately notifiable, urgent (within 24 hours)". However, COVID-19 case surveillance data are collected by jurisdictions and shared voluntarily with CDC.

[5] Geography and demographics (including race) are suppressed for low frequency (<5) records.

[6] Figure A1 plots the share of cumulative COVID-19 cases reported to the CDC up to April 17, 2022 that have missing race information. As this figure shows, North Dakota did not report race information for its COVID-19 cases. Washington, Texas, New York, and Connecticut also have a particularly high rate of missing race data. As **?** discuss, if measurement error in cases and deaths due to testing and reporting is random, then the estimates in this paper will have higher variance but will remain unbiased. However, if there are systematic differences in testing and reporting across racial groups, all analyses in this paper would combine the effect of the COVID-19 incidence, testing, and reporting into one effect. Conditional on reporting race data, there is little reason to suspect measurement error in reporting of race information in the CDC data since all COVID-19 case report forms used the same race categories, so we would expect unbiased but higher variance estimates.

rate" and the cumulative COVID-19 deaths per 100,000 individuals, henceforth referred to as the "mortality rate", as the main outcomes of interest. We use population counts by race from the 2010 Decennial Census (10% sample) to calculate the case and mortality rates nationwide.

For all regression analyses, cases and deaths were first aggregated to 10-year age by race by county categories. To calculate cases and deaths per 100,000 individuals at the county level, we again use population counts by race, 10-year age categories, and county from the 2010 10% Decennial Census. In the CDC data, 2.5% of cases have missing state or county identifiers which we exclude from the analysis. Furthermore, because the Census data suppresses county identifiers for small counties (population <100,000), around 41% of reported COVID-19 cases were excluded (36% of cases during the first wave of the pandemic), yielding a final sample of 27.2 million cases for all regression analyses, of which 1.2 million were during the first six months of the pandemic.

2.2 Socio-economic, health-related, and environmental characteristics

We use the 2016-2019 American Community Survey (ACS) data to obtain the percent of the population that is 65 years of age or older, percent unemployed, percent not in labor force, percent living in group quarters or multigenerational households (2+ generations), percent living in a household with 2+ families, percent uninsured, percent college graduates, percent using public transportation to travel to work, and average household income. These statistics were aggregated at the 10-year age categories by race by county level.

Our second main data source focuses on health-related factors, in particular comorbidities associated with more severe COVID-19 disease as documented by the CDC.[7] We use age-race-county-specific prevalence rate of asthma, cancer, cardiovascular disease (heart attack, angina, coronary heart disease, stroke), diabetes, obesity (BMI $\geq$ 30), smoking, and other chronic illnesses leading to immune suppression from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS).[8]

Land area size from the 2018 Census Gazetteer Files, combined with county-level population data from the 2010 Census, was used to calculate population density for counties. As a measure of county-level pollution, we use the average satellite-measured PM2.5 levels across counties as measured in **?** to capture pollution levels across counties for 2012-2019.

We use mobile device tracking data from SafeGraph through May 31, 2020 to calculate the median time spent at home during January-May as a proxy measure of stay-at-home order adherence. The SafeGraph data reports the daily discrete distribution of time spent at home at the census block-level for devices whose home has been determined to be in that census block.[9] We aggregate this

[7] For a complete list of these comorbidities, see CDC's Underlying Medical Conditions Associated with Higher Risk for Severe COVID-19 – Information for Healthcare Professionals. We have excluded comorbidities that cannot be identified in the BRFSS data.

[8] Later years of the BRFSS do not consistently report counties.

[9] SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if

daily distribution to the median time spent at home in the period prior to state-wide shutdowns and widespread media coverage (in January and February, which we call the "pre-shutdown period") and after (in March-May, "post-shutdown period") within each county using Pareto interpolation. We do not use Safegraph data beyond the early months of the pandemic as social distancing became replaced by masking as the main way of preventing COVID-19 spread in the later part of the pandemic. To our knowledge, comprehensive county-level masking data for the U.S. is unavailable.

As a proxy measure for health care quality and access, we use the average annual county-level age-adjusted Prevention Quality Indicator (PQI) rate per 100,000 individuals for Medicare beneficiaries for 2014-2018, obtained through the Centers for Medicare and Medicaid Services.[10]

3 Summary statistics

3.1 Nationwide cases and deaths

Figure 1 reports *age-adjusted* COVID-19 case and mortality rates per 100,000 individuals documented between January 2020 and April 2022 for white, Black, Hispanic, Asian, American Indian/Alaskan Native (AIAN), Native Hawaiian/Pacific Islander (NHPI), and other racial groups.[11] Nationwide, all non-white racial groups except Asian have higher case rates than white individuals. The case rates among AIAN and NHPI individuals are the highest across all racial groups and are 1.7 and 2 times higher than the case rate for white individuals, respectively. The case rate for Hispanic individuals is 1.8 times higher than for white individuals (22,994 vs. 12,626 cases per 100,000) and the case rate for Black individuals (14,898 cases per 100,000) is 1.2 times higher than the case rate for white. White individuals, however, have a similar case rate as Asian individuals.

These racial disparities also persist in the mortality rate, as shown in the second chart in Figure 1. The mortality rates among AIAN and NHPI individuals are 1.5 and 1.3 times higher than the mortality rate among white individuals and are the highest mortality rates across all racial groups. The mortality rates among white, Black, and Hispanic individuals are roughly similar. Asian individuals are the least likely to die from COVID-19, with mortality rates approximately 30% lower than that of white individuals.[12]

Figure 2 and Figure 3 show how these disparities have evolved throughout the duration of the

fewer than five devices visited an establishment in a month from a given census block group. The data was generated via a series of GPS pings from anonymous mobile devices. A device's home census block was determined as the common nighttime location of the mobile device over a 6-week period.

[10] The PQI is calculated using inpatient data and captures conditions for which hospitalization could be prevented via appropriate outpatient care and conditions that could be less severe if treated early and appropriately. As a result, higher PQI values correspond to worse health care quality and access.

[11] Age adjustment was performed using the 2010 Census population by age and race as the standard population. Since age is the main risk factor for severe COVID-19, and since white individuals tend to be older than non-white individuals, not adjusting for age generally results in smaller disparities in COVID-19 burden across racial groups.

[12] Appendix Figure A2 shows that similar, if not starker, disparities persist when looking at COVID-19 related hospitalizations and ICU rates per 100,000 reported to the CDC. However, hospitalizations and ICU stays may not be as reliably reported as cases and deaths; see Appendix Figure A3 and Figure A4.

pandemic. Across all waves of the pandemic, nonwhite individuals were more likely to contract COVID-19, in particular Hispanic, AIAN, and NHPI individuals. Death rates were also higher among nonwhite individuals – Black individuals were almost twice as likely to die from COVID-19 in the first wave of the pandemic as compared to white. AIAN and NHPI individuals had some of the highest observed mortality rates during the subsequent COVID-19 waves.

Focusing specifically on the first wave of the pandemic, Figure 4 shows that the racial disparities in the beginning of the pandemic are even starker: Black and Hispanic individuals were 2.9 and 3.9 times more likely to contract COVID-19 relative to white individuals.[13] They were also 2.2 and 1.3 times more likely to die from COVID-19 relative to white individuals. AIAN individuals were 2.9 times more likely to contract COVID-19 and 1.6 times more likely to die from COVID-19 relative to white individuals. While NHPI individuals were 3.2 times more likely to have COVID-19 compared to white individuals, they were less likely to die from COVID-19. Interestingly, Asian individuals had a higher COVID-19 case and mortality rate relative to white individuals in the first wave of the pandemic. Because COVID-19 policy and prevention responses across counties and at the individual level are endogenous to the spread of COVID-19, the remainder of the paper focuses on the relationship between socio-economic, environmental, and health-related factors and COVID-19 cases and deaths during the first wave of the pandemic, with a particular focus on what factors and characteristics were associated with the *initial* racial disparity in COVID-19 burden.

3.2 Socio-economic and health-related factors

Table 1 presents summary statistics of race-specific socio-economic, health-related, and environmental characteristics across counties in our sample. We choose characteristics that may be correlated with COVID-19 transmission or mortality as documented by the literature or by epidemiological modeling (e.g., (**?**), (**?**)) and group these characteristics into the following broad categories: age-related (percent elderly), housing-related (percent residing in group quarters, which includes nursing homes, percent residing in households with two or more families, and percent residing in multigenerational households), density-related (population density and percent using public transportation), income-related (unemployment rate, percent not in labor force, uninsured rate, household income, and percent working in service industry), knowledge-related (percent with college degree), environmental factors and social distancing (pollution and percent change in median time spent at home between Jan-Feb and March-May of 2020), and, lastly, health (health care quality, measured by PQI rate per 100,000, comorbidities associated with severe COVID-19 disease). Population density, pollution, the PQI rate, and the median time spent at home do not vary by race because this data was unavailable at the county level.

As is widely known, non-white racial groups are less likely to be elderly, which aligns with our earlier note that adjusting COVID-19 case and mortality rates by age exacerbates the differences in COVID-19 burden by race. Non-white racial groups are more likely to live in denser households,

[13] We define the first wave of the pandemic as January 2020 to June 2020.

more likely to use public transportation, more likely to be uninsured, more likely to have lower income and less education (except for Asians). Non-white individuals are also more likely to be unemployed and work in the service industry. Comorbidity rates are highest among Black and AIAN individuals, while all other racial groups have relatively similar comorbidity rates. Rates of diabetes, obesity, and smoking are particularly high among Black and AIAN individuals.

These differences in the levels of various characteristics across racial groups may contribute to differences in COVID-19 case and mortality rates by race even if the size of the correlation between these characteristics and COVID-19 burden is the same across racial groups. This motivates the first part of the Oaxaca-Blinder decomposition, which will determine how much these differences in *levels* of various characteristics correlated with COVID-19 burden explain COVID-19 race differentials *holding the correlation between characteristics and case/mortality rates constant.*

4 Relationship between socio-economic and health characteristics and COVID-19 burden by race

Even if racial groups had the same socio-economic, health, and environmental characteristics, we can observe racial disparities in COVID-19 case and mortality rates if the impact of these characteristics on COVID-19 case and mortality rates varies by race. To examine whether this may be the case, Table 2 and Table 3 present the relationships between COVID-19 case rates and mortality rates, respectively, and race-specific characteristics across the health, socio-economic, and environmental domains. In other words, the results in Table 2 and Table 3 represent OLS regression coefficients from separately regressing county-level COVID-19 case and mortality rates for white, Asian, Black, Hispanic, AIAN, NHPI individuals on county-level characteristics for the corresponding racial group. Counties with large AIAN and NHPI populations are much smaller in size and thus are frequently excluded from the data as discussed in Section 2; the estimates for these racial groups are as a result noisier. Table 2 and Table 3 use data before June 2020 and include fixed effects for 10-year age categories. Density, pollution, and the percent change in median time spent at home between January-February and March-May are not available separately by race and are thus included as an overall measure for each county. Lastly, because the full set of characteristics discussed in Section 3 are likely to be correlated, we minimize the number of included covariates by combining the percent of individuals in group quarters, in households with 2+ families, and in multigenerational households into a single measure of percent of individuals living in dense households. We also include a covariate for the presence of any comorbidity associated with increased COVID-19 risk instead of including all comorbidities separately.

4.1 COVID-19 case rates

Table 2 indicates that residing in denser households is positively associated with higher case rates among all non-white racial groups (controlling for all other factors), but is particularly large for

Hispanic individuals, even though noisy and marginally significant. Our estimates imply that counties with a 10 percentage point increase in the percent of Hispanic individuals living in dense households are associated with an 11 percent higher average Hispanic COVID-19 case rate. The correlation between household size and case rate is negative for white individuals; however, when breaking down this measure into percent of individuals in group quarters vs. other types of dense households, the number of white individuals living in group quarters such as nursing homes is strongly positively associated with the white COVID-19 case rate at the county-level.

Average household income is generally negatively associated with the case rate across all racial groups, but is statistically significant for white, Hispanic, and NHPI individuals. Education is particularly negatively correlated with the COVID-19 case rate among white and Black individuals – a 10 percentage point increase in the percent of Black individuals with a college degree is associated with a 10% decrease in the average Black COVID-19 case rate (around 5% for white individuals), controlling for other factors. As documented in **?**, knowledge about COVID-19 prevention differs by demographic groups, but our results complement these findings by suggesting education gradients in COVID-19 prevention conditional on race.

Use of public transportation and population density are strongly correlated with case rates across almost all races, indicating that density-related measures are another important factor in the initial spread of COVID-19. Public transportation use is particularly strongly correlated with the Hispanic case rate – an additional 10 percentage point of Hispanic individuals using public transportation to travel to work is associated with a 13% increase in the average Hispanic COVID-19 case rate. The correlation between population density and case rates is among the largest across all characteristics and all racial groups. Pollution is another significant predictor of COVID-19 spread among Asian and Hispanic individuals, but not white, Black, or AIAN individuals. Lastly, the percent change in the time spent at home during March 2020-June 2020 relative to January-February 2020 is positively associated with the case rate across almost all racial groups with the exception of NHPI, possibly reflecting endogenous response at the individual level. Health-related factors such as comorbidities and health care quality are uncorrelated with the COVID-19 case rate across all racial groups. [14]

4.2 COVID-19 mortality rates

Table 3 shows the results from the same OLS models as for case rates discussed above, but using mortality rates instead. We find that similar factors correlate with COVID-19 mortality rates as

[14] Under the assumption that within counties racial groups interact with each other, characteristics of a given racial group may correlate with COVID-19 case and death rates of other racial groups, even when controlling for the characteristics of the other racial groups. Policy interventions to curb the COVID-19 spread in certain racial groups may magnify in impact if behavior or characteristics of targeted racial groups spills over to other groups. For instance, Table A1 and Table A2 show the OLS regression results from regressing the COVID-19 case and death rate for the listed racial group on the characteristics of white individuals, controlling for own-group characteristics. These results suggest that non-white individuals living in counties where white individuals are richer and more educated have lower COVID-19 case rates, even controlling for non-white income and education levels. However, regressions of this sort may capture area factors that we are not controlling for in our regressions that affect all racial groups equally and are correlated with income and education of white individuals.

for case rates, suggesting that the mortality rate measure might be picking up increased COVID-19 spread. Neither health care quality nor comorbidities are correlated with mortality rates, with the exception of Black individuals, where counties with a 10 percentage point higher share of Black individuals with an underlying comorbidity are associated with a 13% higher COVID-19 mortality rate among Black individuals, although only marginally significant.

To address the fact that the mortality rate may be capturing overall COVID-19 burden rather than severity of disease, we estimate the same models as in Table 3 but instead use the mortality rate conditional on COVID-19 infection, i.e., the percent of individuals that die conditional on being infected with COVID-19. These results, shown in Table 4, indicate that counties with lower health care quality are particularly positively correlated with higher mortality rates among white individuals conditional on infection. Lower health care quality is also positively associated with higher mortality rates (conditional on infection) for all other racial groups, but these results are not statistically significant. Importantly, counties where Black individuals had a higher rate of comorbidities had a higher conditional Black mortality rate. An additional 10 percentage points of Black individuals with an underlying comorbidity are associated with a 3.75% increase in the average percent of Black individuals dying from COVID-19 if infected.

5 Oaxaca-Blinder decomposition

Given the differences in county-level characteristics and coefficients on those characteristics by race, the Oaxaca-Blinder decomposition introduced in **?** and **?** allows for a systematic examination of the extent to which observed differences in COVID-19 case rates between white and non-white racial groups stem from the differences in levels of observed characteristics by race discussed in Section 3 vs. the differences in the strength of the correlation between these characteristics and COVID-19 burden across racial groups demonstrated in Section 4. In Table 5 and Table 6, we decompose the average differences in COVID-19 case and mortality rates between Black and white and Hispanic and white individuals into three components: one due to differences in average characteristics (or levels of characteristics documented in Section 3) by race ("differences in endowments"), differences due to the "impact" of characteristics on case rates across racial groups ("differences in coefficients"), as estimated in Section 4, and differences due to the interaction between the levels and coefficients for each characteristic ("differences in interactions"). We exclude AIAN and NHPI individuals since sample sizes for these racial groups are small.

In Table 5 and Table 6, the columns labeled "endowments" show the change in the average case or death rate differential by race if levels of the non-white racial group were equal to the levels for whites, either of individual socio-economic, health-related, or environmental factors or across all factors. The columns labeled "coefficients" show the change in the case or mortality rate differential between non-white and white if the correlations between individual or aggregate factors for the non-white racial group were equal to those of whites. And lastly, the columns labeled "interactions"

show the change in the average case or mortality rate differential by race from changing both the levels and the coefficients. Differences in the constants comprise the "unexplained" component of the decomposition – the part that is unrelated to any of the factors included in the models. Note that all models also include 10-year age category fixed effects and controls for missing values for included covariates, which may also contribute to explaining part of the observed case rate differentials by race but are not of interest for this analysis. In a similar vein, we do not discuss the impact of the interactions in explaining the case rate differential in as much detail.

5.1 COVID-19 case rates

As shown Table 5, if Black individuals had the same observed characteristics as white individuals, but different coefficients, we would expect the Black-white case rate difference to *decrease* by 152 cases per 100,000 relative to the observed case rate difference of 624 additional cases per 100,000 among Black individuals, which represents a 25% decrease relative to the average case rate difference between Black and white individuals reported at the top of Table 5. This increase largely stems from equalizing the education and income levels (and to some extent household density) between Black and white individuals, which are particularly strongly correlated with case rates among Black individuals.

If, however, observed characteristics between Black and white individuals remained at levels reported in Section 3, but instead Black individuals had the same magnitude of correlations between characteristics and case rates as white individuals, the difference in case rates between Black and white individuals would fall by 208 cases per 100,000, which is a 33% reduction in the case rate differential by race. These reductions largely stem from the fact that population density, social distancing, comorbidity rates, and household density are more strongly correlated with case rates for Black individuals than white individuals. Because the correlation between many factors and the case rate for white individuals are higher than those for Black individuals, equalizing the coefficients across the two racial groups would be associated with an approximately fivefold increase in the Black-white case rate difference, as indicated by the difference in coefficients.

The second panel of Table 5 shows that if Hispanic individuals had the same levels of socio-economic, environmental, and health-related characteristics as white individuals documented in Section 3, but holding the race-specific correlations between these factors and case rates constant, the difference in the COVID-19 case rate between Hispanic and white individuals would fall by 259 cases per 100,000, which is a 23% reduction in the case rate differential of 1,112 cases per 100,000. This reduction stems from equalizing the household density and income levels of white and Hispanic individuals, both of which are highly correlated with case rates among Hispanic individuals. On the other hand, if we allow for Hispanic and white individuals to have different endowments of the characteristics considered in this decomposition, but instead assume that Hispanic individuals had the same correlations between these factors and case rates as white individuals, the case rate differential between Hispanic and white individuals would fall by 1,645 cases per 100,000, leading

to higher case rates among white individuals. As for Black individuals, this is largely due to the stronger correlation between population density, social distancing, pollution, and household density for Hispanic individuals relative to white individuals.

5.2 COVID-19 mortality rates

Table 6 performs the Oaxaca-Blinder decomposition for the COVID-19 mortality rate instead of the case rate. If Black individuals had the same levels of socio-economic, health-related, and environmental characteristics as white individuals, the average mortality rate difference between Black and white individuals would fall by 55 deaths per 100,000, representing a 47% decrease in the average difference in mortality rates between Black and white individuals. Similar to cases, this is achieved by equalizing education, income, and household density between Black and white individuals, as well as comorbidity rates, but to a smaller extent. The second panel of Table 6 shows that if Hispanic individuals had the same endowments as white individuals reported in Section 3, the average mortality rate difference between Hispanic and white individuals would decrease by 22 deaths per 100,000 (a 26% decrease relative to the average mortality rate differential).

On the other hand, Table 6 shows that if Black and Hispanic individuals had the same correlations between characteristics and mortality rates as white individuals, but holding the levels of these characteristics constant, then the mortality rate differential would fall by twice as much as the observed mortality rate differential, leading to higher mortality rates among whites relative to Black and Hispanic individuals. As for cases, this is largely driven by the stronger correlation between population density, social distancing, and household density and mortality rates for non-white individuals relative to white individuals.

6 Conclusion

At the onset of the COVID-19 pandemic, non-white individuals were 3-4 times more likely to contract COVID-19 and 1.5-2 times more likely to die from COVID-19. While our findings suggest that population density, household income, and pollution are important factors for the spread of COVID-19 across all racial groups, we also find that Hispanic case rates are particularly high in counties where Hispanic individuals were more likely to commute to work using public transportation and more likely to live in denser or multifamily households. Furthermore, counties where Black individuals are more educated also have a lower Black case rate, and this correlation is much weaker for all other racial groups. We find little evidence that health-related factors are correlated with the initial COVID-19 case and mortality rate disparities; the notable exception is the fact that the rate of underlying comorbidities among Black individuals is highly correlated with the Black mortality rate from COVID-19. If Black and Hispanic individuals had the same socio-economic characteristics as white individuals, observed case and mortality rate differentials by race would fall by 25-35%.

An important caveat of this paper is that our findings are not causal and use data aggregated at the county level. Additionally, we do not account for endogenous individual and county- or state-level response to the COVID-19 pandemic. However, our findings suggest that even though we find little evidence that health-related factors were associated with an increased COVID-19 spread among non-white individuals, direct interventions targeting the higher rates of diabetes, obesity, and smoking among Black individuals, which we show are correlated with more severe COVID-19 illness and mortality in Black individuals conditional on infection, may decrease susceptibility to infectious disease in future generations. Importantly, our results imply that policies beyond the health care and public health domains may play a huge role in the disparate impact of pandemic spread by race. For instance, housing policies targeted to Hispanic individuals that decrease the rate of multi-family or multi-generational households (e.g., by increasing housing affordability) may also be associated with lower risk of disease spread among this racial group. Similarly, educational and informational policies on infectious disease prevention might play a much larger role for improving and maintaining public health for Black individuals than for any other racial group.

References

Milena Almagro, Joshua Coven, Arpit Gupta, and Angelo Orane-Hutchinson. Racial Disparities in Frontline Workers and Housing Crowding During COVID-19: Evidence from Geolocation Data. SSRN Scholarly Paper ID 3695249, Social Science Research Network, Rochester, NY, September 2020. URL https://papers.ssrn.com/abstract=3695249.

Marcella Alsan, Stefanie Stantcheva, David Yang, and David Cutler. Disparities in Coronavirus 2019 Reported Incidence, Knowledge, and Behavior Among US Adults. *JAMA Network Open*, 3(6):e2012403, June 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.12403. URL https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2767261.

Marcella Alsan, Amitabh Chandra, and Kosali Simon. The Great Unequalizer: Initial Health Effects of COVID-19 in the United States. *Journal of Economic Perspectives*, 35(3):25–46, August 2021. ISSN 0895-3309. doi: 10.1257/jep.35.3.25. URL https://pubs.aeaweb.org/doi/10.1257/jep.35.3.25.

Joseph A Benitez, Charles J Courtemanche, and Aaron Yelowitz. Racial and Ethnic Disparities in COVID-19: Evidence from Six Large Cities. Working Paper 27592, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27592. Series: Working Paper Series.

Alan S. Blinder. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436, 1973. ISSN 0022166X. doi: 10.2307/144855. URL https://www.jstor.org/stable/144855?origin=crossref.

George J. Borjas. Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. Technical Report w26952, National Bureau of Economic Research, April 2020. URL https://www.nber.org/papers/w26952.

Klaus Desmet and Romain Wacziarg. Understanding Spatial Variation in COVID-19 across the United States. Technical Report w27329, National Bureau of Economic Research, June 2020. URL https://www-nber-org.ezp-prod1.hul.harvard.edu/papers/w27329.

Zhanlian Feng, Mary L. Fennell, Denise A. Tyler, Melissa Clark, and Vincent Mor. The Care Span: Growth of racial and ethnic minorities in US nursing homes driven by demographics and possible disparities in options. *Health Affairs (Project Hope)*, 30(7):1358–1365, July 2011. ISSN 1544-5208. doi: 10.1377/hlthaff.2011.0126.

Tian Gu, Jasmine A. Mack, Maxwell Salvatore, Swaraaj Prabhu Sankar, Thomas S. Valley, Karandeep Singh, Brahmajee K. Nallamothu, Sachin Kheterpal, Lynda Lisabeth, Lars G. Fritsche, and Bhramar Mukherjee. Characteristics Associated With Racial/Ethnic Disparities

in COVID-19 Outcomes in an Academic Health Care System. *JAMA Network Open*, 3(10): e2025197, October 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.25197. URL https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2771935.

Christopher R. Knittel and Bora Ozaltun. What Does and Does Not Correlate with COVID-19 Death Rates. Technical Report w27391, National Bureau of Economic Research, June 2020. URL https://www.nber.org/papers/w27391.

John McLaren. Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data. Technical Report w27407, National Bureau of Economic Research, June 2020. URL https://www.nber.org/papers/w27407.

Ronald Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709, 1973. ISSN 0020-6598. doi: 10.2307/2525981. URL http://www.jstor.org/stable/2525981. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Gbenga Ogedegbe, Joseph Ravenell, Samrachana Adhikari, Mark Butler, Tiffany Cook, Fritz Francois, Eduardo Iturrate, Girardin Jean-Louis, Simon A. Jones, Deborah Onakomaiya, Christopher M. Petrilli, Claudia Pulgarin, Seann Regan, Harmony Reynolds, Azizi Seixas, Frank Michael Volpicelli, and Leora Idit Horwitz. Assessment of Racial/Ethnic Disparities in Hospitalization and Mortality in Patients With COVID-19 in New York City. *JAMA Network Open*, 3(12): e2026881, December 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.26881. URL https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2773538.

Eboni G. Price-Haywood, Jeffrey Burton, Daniel Fort, and Leonardo Seoane. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *New England Journal of Medicine*, 382(26):2534–2543, June 2020. ISSN 0028-4793. doi: 10.1056/NEJMsa2011686. URL https://doi.org/10.1056/NEJMsa2011686. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMsa2011686.

Tiana N. Rogers, Charles R. Rogers, Elizabeth VanSant-Webb, Lily Y. Gu, Bin Yan, and Fares Qeadan. Racial Disparities in COVID-19 Mortality Among Essential Workers in the United States. *World Medical & Health Policy*, August 2020. ISSN 2153-2028. doi: 10.1002/wmh3.358. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7436547/.

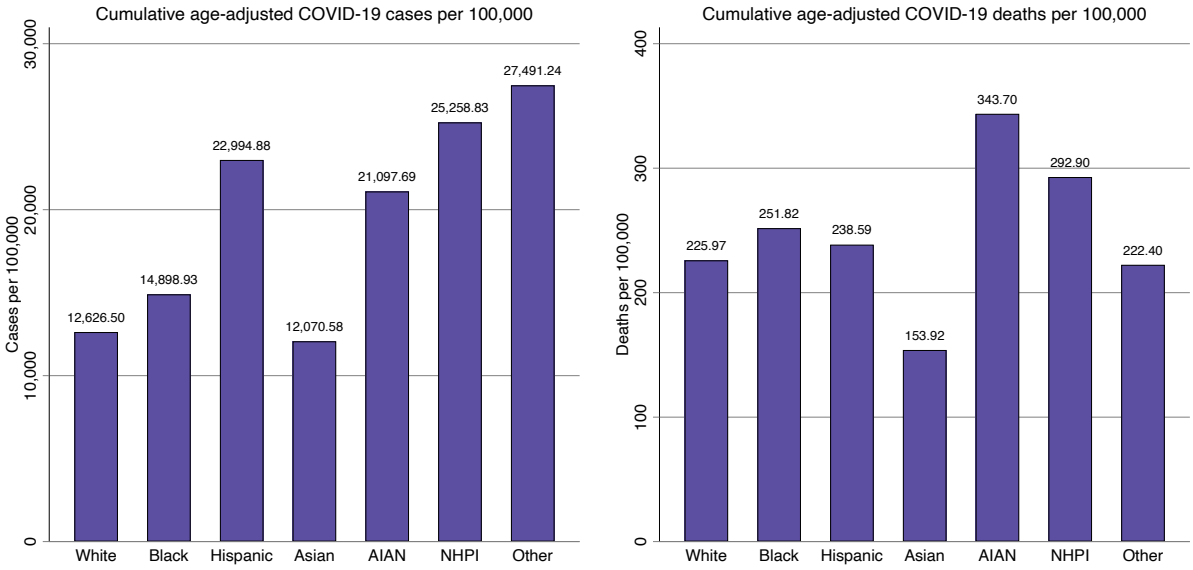
Aaron van Donkelaar, Randall V. Martin, Chi Li, and Richard T. Burnett. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology*, 53(5):2595–2611, March 2019. ISSN 0013-936X, 1520-5851. doi: 10.1021/acs.est.8b06392. URL https://pubs.acs.org/doi/10.1021/acs.est.8b06392.

Xiao Wu, Rachel C. Nethery, Benjamin M. Sabath, Danielle Braun, and Francesca Dominici. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *medRxiv*, April 2020. doi: 10.1101/2020.04.05.20054502. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7277007/.

Baligh R. Yehia, Angela Winegar, Richard Fogel, Mohamad Fakih, Allison Ottenbacher, Christine Jesser, Angelo Bufalino, Ren-Huai Huang, and Joseph Cacchione. Association of Race With Mortality Among Patients Hospitalized With Coronavirus Disease 2019 (COVID-19) at 92 US Hospitals. *JAMA Network Open*, 3(8):e2018039, August 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.18039. URL https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2769387.

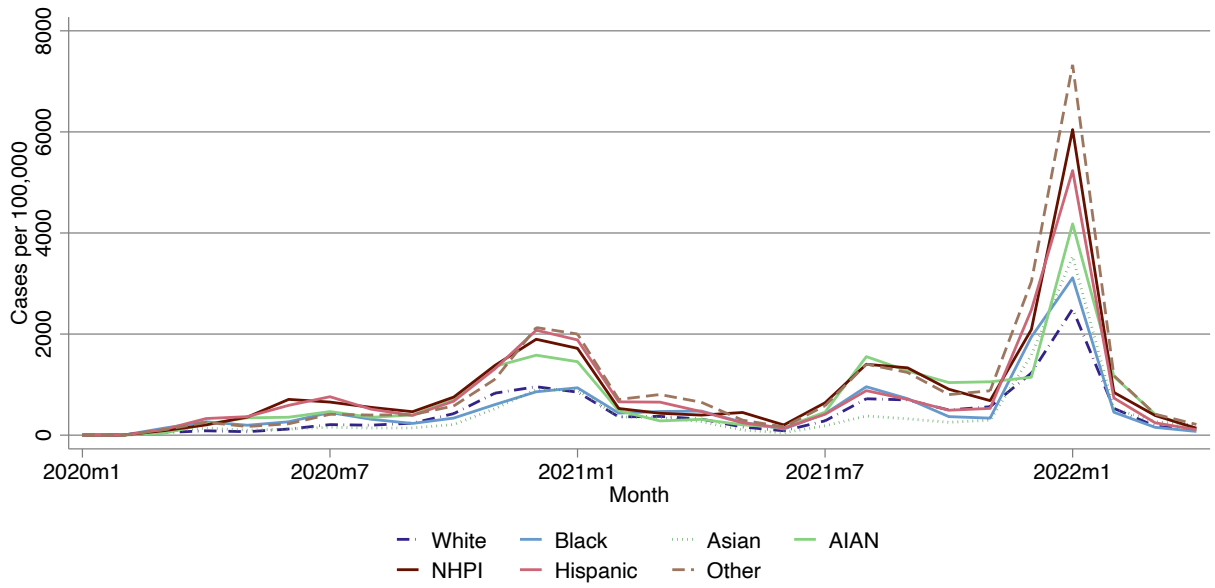
7 Tables and Figures

Figure 1: National cumulative COVID-19 case and mortality rates per 100,000 by race



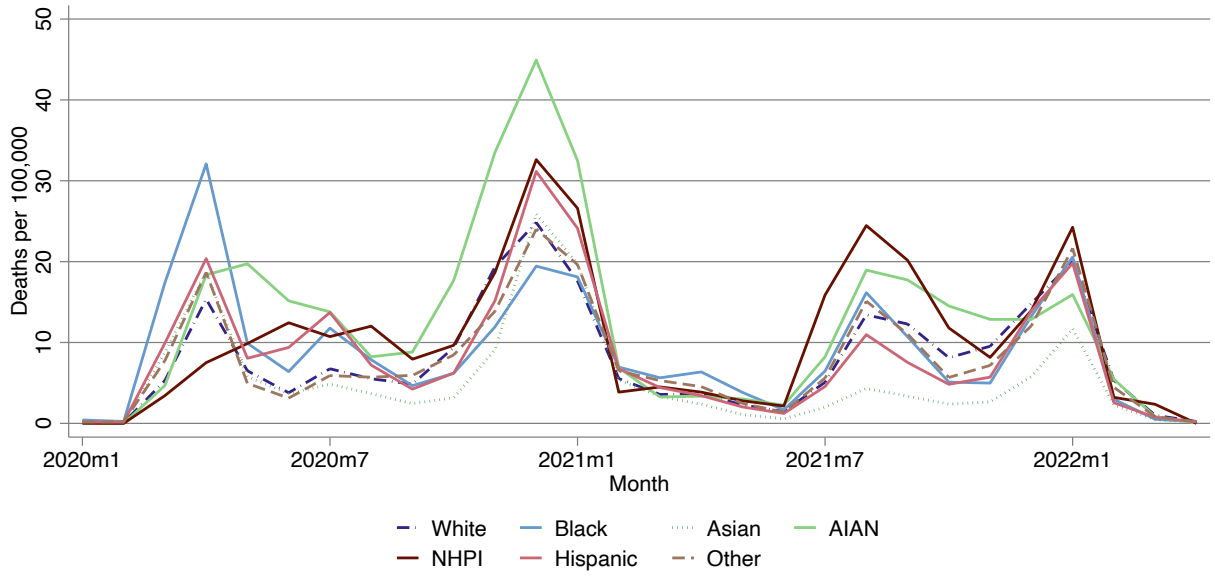
Note. This figure shows cumulative age-adjusted COVID-19 cases and deaths per 100,000 individuals reported up to April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Age adjustment was performed using the 2010 Decennial Census population. Back to Section 3.1.

Figure 2: National age-adjusted COVID-19 case rates per 100,000 by race over time



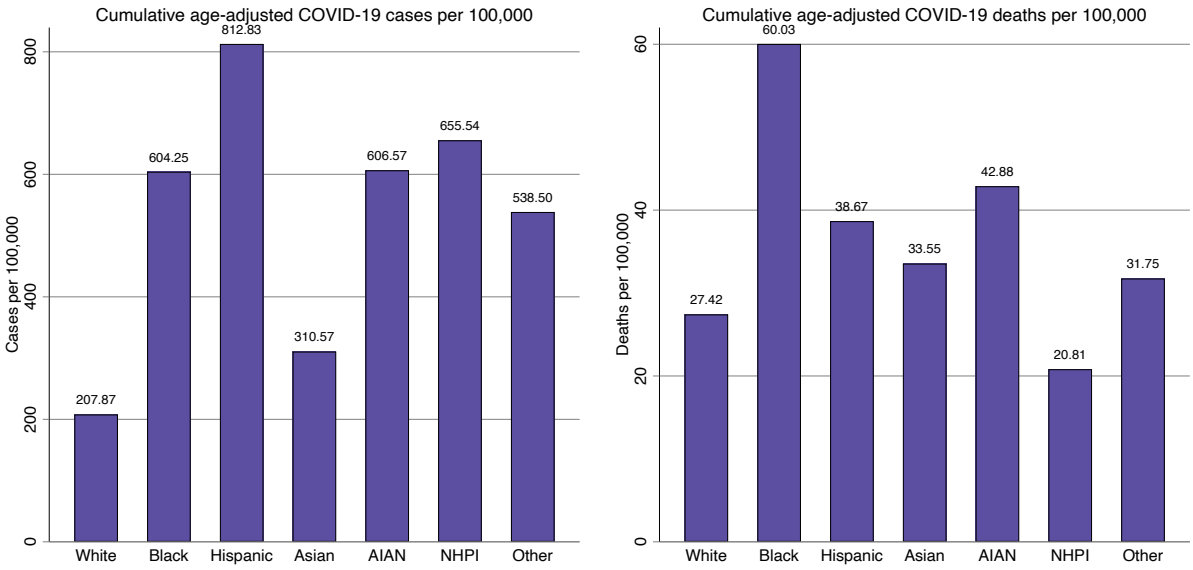
Note. his figure shows monthly age-adjusted COVID-19 cases per 100,000 individuals reported up to April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Age adjustment was performed using the 2010 Decennial Census population. Back to Section 3.1.

Figure 3: National age-adjusted COVID-19 mortality rates per 100,000 by race over time



Note. This figure shows monthly age-adjusted COVID-19 deaths per 100,000 individuals reported up to April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Age adjustment was performed using the 2010 Decennial Census population. Back to Section 3.1.

Figure 4: National cumulative COVID-19 case and mortality rates per 100,000 by race



Note. The map plots cumulative COVID-19 cases and deaths per 100,000 individuals reported as of April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Population counts by race were obtained from the 2016-2019 American Community Survey. Back to Section 3.1.

Table 1: Summary statistics across counties by race

	White		Asian		Black		Hispanic		AIAN		NHPI	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Characteristics that vary by race												
Socio-economic												
Age 65+, %	27.3	41.4	24.2	39.8	25.1	40.4	25.2	40.4	18.8	36.3	12.6	31.1
Lives in dense housing, %	14.7	8.0	23.1	17.6	24.0	16.2	22.1	12.9	24.3	28.5	32.6	32.9
In group quarters, %	2.8	4.2	2.8	9.1	7.5	12.8	3.0	6.8	5.7	16.3	4.6	16.3
2+ families in household, %	6.7	6.3	6.7	10.3	7.5	9.2	8.3	8.2	9.8	20.0	12.8	24.4
Multigenerational household (2+ gen), %	6.0	3.1	14.5	14.2	10.4	10.2	12.0	9.9	10.7	20.3	18.5	26.6
Unemployed, %	2.6	2.2	2.2	4.2	4.6	6.9	3.2	4.2	4.1	12.2	2.9	10.5
Not in labor force, %	45.5	30.7	47.7	32.2	49.3	30.6	45.7	30.4	48.7	36.2	36.2	36.8
Works in service industry, %	33.8	21.9	36.6	27.4	32.4	23.5	30.8	22.3	33.5	32.6	37.6	35.9
Average household income, $	107,025	37,811	116,086	48,393	73,052	36,738	78,344	28,997	83,623	57,010	105,556	77,727
Uninsured, %	5.4	5.4	7.2	9.8	8.7	11.0	15.1	14.8	12.2	21.8	11.6	23.6
College graduates, %	27.0	19.0	37.5	28.3	16.2	15.4	14.4	14.7	16.4	24.8	17.3	28.1
Use public transit, %	2.7	6.8	5.0	11.4	7.2	13.1	4.5	10.8	6.5	19.1	6.6	19.4
Health-related												
Any comorbidity, %	55.9	16.9	33.2	36.1	64.4	30.4	55.6	32.8	67.9	39.8	51.7	43.3
Asthma, %	14.2	12.1	8.1	20.7	16.2	22.8	14.5	23.2	18.3	32.2	13.9	30.2
Cancer, %	16.1	16.3	4.5	17.1	6.4	15.2	7.5	18.6	12.8	29.0	5.5	20.0
Cardiovascular disease, %	10.5	12.1	6.0	19.5	11.0	20.2	9.9	21.4	15.2	30.4	5.7	20.4
Diabetes, %	9.8	9.1	11.1	24.5	17.7	24.3	15.0	24.4	16.2	30.7	13.9	30.7
Obese, %	23.2	13.6	8.8	21.6	35.0	29.3	27.5	28.7	31.7	38.5	27.8	39.1
Current smoker	38.8	27.8	34.2	41.2	49.4	37.9	39.7	38.0	50.2	44.1	41.3	44.0
Other chronic disease	12.8	14.9	6.9	20.4	13.4	24.7	13.1	24.8	22.5	39.3	15.6	34.0
Characteristics that do not vary by race												
Population density (log)	6.2	1.3	6.2	1.3	6.2	1.3	6.2	1.3	6.2	1.3	6.3	1.4
Median time spent at home (Mar 2020-May 2020), hrs	12.4	1.8	12.4	1.8	12.4	1.8	12.4	1.8	12.4	1.8	12.8	1.7
Median time spent at home (Jan 2020-Feb 2020), hrs	10.6	1.3	10.6	1.3	10.6	1.2	10.6	1.2	10.6	1.2	10.8	1.2
% change in median time spent at home	16.7	8.8	16.8	8.9	16.7	8.9	16.8	8.8	16.8	8.8	18.2	8.5
Average PM 2.5	7.6	1.2	7.6	1.2	7.6	1.2	7.6	1.2	7.5	1.2	7.5	1.3
Average PQI rate	4,269	1,058	4,250	1,051	4,288	1,057	4,266	1,060	4,232	1,087	3,923	1,120

Note. Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 3.2.

Table 2: Regressions of county-level cumulative COVID-19 cases per 100,000 by race on race-specific characteristics

	(1) White b/se	(2) Asian b/se	(3) Black b/se	(4) Hispanic b/se	(5) AIAN b/se	(6) NHPI b/se
Characteristics that vary by race						
% in dense household	-3.8305***	2.0491	4.0565	14.8991**	2.1278*	6.6221
	(0.9399)	(1.7822)	(6.1003)	(7.0933)	(1.2165)	(6.7170)
Avg. hh income	-0.0017***	-0.0015*	-0.0021*	-0.0069***	-0.0000	-0.0038**
	(0.0002)	(0.0009)	(0.0011)	(0.0024)	(0.0007)	(0.0019)
% using public transit	11.4820***	4.7483	-1.0099	17.0793**	1.9383	-3.4751
	(1.9279)	(4.2248)	(3.1412)	(7.6259)	(1.3152)	(10.8372)
% college graduates	-1.0564**	0.4766	-8.7027**	1.1885	0.4602	-4.6751
	(0.5327)	(2.0646)	(4.2354)	(7.1427)	(1.9072)	(8.0611)
Comorbidities						
% with any comorbidity	-0.2493	-1.4038	2.7560	-1.8243	-0.6629	-14.4640
	(0.3039)	(1.1855)	(2.4915)	(2.2870)	(0.7716)	(11.2994)
Characteristics that do not vary by race						
Avg. % change in time spent at home	4.6471***	15.4275***	24.3035***	16.1030***	2.2959	-40.3280***
	(0.8160)	(5.1799)	(8.7322)	(6.0882)	(3.4384)	(12.7682)
Population density (log)	58.3885***	11.1536	136.2395***	156.9672**	13.8910	185.4459**
	(8.6956)	(48.2267)	(46.7492)	(60.9726)	(24.6221)	(85.3714)
Avg. PQI rate	0.0032	-0.0264	0.0233	0.0062	-0.0220	-0.1353
	(0.0054)	(0.0469)	(0.0506)	(0.0777)	(0.0210)	(0.1268)
Avg. PM 2.5	-9.3628*	127.3362***	-110.4963	131.8469**	-67.0725**	166.7627
	(5.5927)	(43.0527)	(68.0643)	(56.1364)	(27.5767)	(128.6960)
10-year age FE	Yes	Yes	Yes	Yes	Yes	Yes
Missing: housing vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: employment vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: transit vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: education vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: comorbidity vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: other vars	Yes	Yes	Yes	Yes	Yes	Yes
N	3,816	3,639	3,705	3,747	3,011	1,368
R-squared	0.3326	0.0246	0.0615	0.0770	0.0262	0.0506
Dependent var. mean	219	511	843	1,331	266	770

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note. Cases represent the cumulative COVID-19 cases per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 4.1.

Table 3: Regressions of county-level cumulative COVID-19 deaths per 100,000 by race on race-specific characteristics

	(1) White b/se	(2) Asian b/se	(3) Black b/se	(4) Hispanic b/se	(5) AIAN b/se	(6) NHPI b/se
Characteristics that vary by race						
% in dense household	-0.2351	0.8678*	2.7108	1.6191	0.9410**	0.0389
	(0.3811)	(0.4628)	(3.7637)	(1.1393)	(0.3828)	(0.1606)
Avg. hh income	-0.0007***	0.0001	-0.0005	-0.0013***	0.0001	-0.0002*
	(0.0001)	(0.0002)	(0.0005)	(0.0004)	(0.0002)	(0.0001)
% using public transit	2.8804***	3.2826**	-1.3383	4.5448**	-0.4649*	-0.6348**
	(1.0372)	(1.6679)	(1.4492)	(2.0932)	(0.2560)	(0.2692)
% college graduates	-0.7157***	0.5470	-3.5002	1.9622	0.2342	-0.4085**
	(0.2342)	(0.4099)	(2.7125)	(2.0850)	(0.4228)	(0.1779)
Comorbidities						
% with any comorbidity	-0.0870	0.0294	2.1284*	0.3094	-0.2206	-0.2785
	(0.1414)	(0.2637)	(1.1857)	(0.4878)	(0.2669)	(0.5535)
Characteristics that do not vary by race						
Avg. % change in time spent at home	1.7376***	2.3957***	9.9982*	5.6576***	-0.1542	-1.0421
	(0.3763)	(0.8256)	(5.3706)	(1.1921)	(0.7354)	(0.7296)
Population density (log)	28.4934***	25.7133***	82.2192***	43.3244***	3.3571	10.0066
	(4.4349)	(6.9361)	(23.6833)	(12.1070)	(5.4286)	(8.1848)
Avg. PQI rate	0.0001	0.0060	0.0224	0.0100	0.0029	-0.0035
	(0.0025)	(0.0070)	(0.0261)	(0.0095)	(0.0054)	(0.0063)
Avg. PM 2.5	-9.1581***	-5.9512	-54.1331	-13.9766	-9.6941*	-1.8203
	(2.2968)	(8.0933)	(34.2647)	(11.1905)	(5.0727)	(13.7443)
10-year age FE	Yes	Yes	Yes	Yes	Yes	Yes
Missing: housing vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: employment vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: transit vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: education vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: comorbidity vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: other vars	Yes	Yes	Yes	Yes	Yes	Yes
N	3,816	3,639	3,705	3,747	3,011	1,368
R-squared	0.3121	0.1521	0.0463	0.1440	0.0319	0.0477
Dependent var. mean	43	64	160	129	27	30

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note. Deaths represent the cumulative COVID-19 deaths per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 4.2.

Table 4: Regressions of county-level cumulative COVID-19 deaths conditional on infection by race on race-specific characteristics

	(1) White b/se	(2) Asian b/se	(3) Black b/se	(4) Hispanic b/se	(5) AIAN b/se	(6) NHPI b/se
Characteristics that vary by race						
% in dense household	0.0312	0.0069	0.0312	-0.0199	0.0929	0.0052
	(0.0243)	(0.0366)	(0.0319)	(0.0327)	(0.0583)	(0.0436)
Avg. hh income	-0.0000	0.0000	-0.0000	-0.0000	0.0000	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
% using public transit	0.0766**	0.0985**	0.0233	0.0767***	-0.0688	-0.0160
	(0.0317)	(0.0403)	(0.0185)	(0.0285)	(0.0471)	(0.0457)
% college graduates	-0.0082	0.0088	-0.0049	-0.0500	-0.0425	0.0087
	(0.0146)	(0.0263)	(0.0408)	(0.0355)	(0.0478)	(0.0349)
Comorbidities						
% with any comorbidity	0.0126	0.0035	0.0267***	-0.0035	-0.0068	0.0076
	(0.0086)	(0.0127)	(0.0094)	(0.0096)	(0.0262)	(0.0211)
Characteristics that do not vary by race						
Avg. % change in time spent at home	0.0488**	0.0214	0.0717**	0.0989***	-0.0437	-0.0924
	(0.0200)	(0.0483)	(0.0309)	(0.0278)	(0.0996)	(0.1093)
Population density (log)	0.8440***	0.3248	0.7630***	0.2941	0.3514	-1.0760
	(0.2126)	(0.3650)	(0.2410)	(0.2090)	(0.6390)	(1.0591)
Avg. PQI rate	0.0004***	0.0006	-0.0000	0.0003	0.0005	0.0004
	(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0008)	(0.0009)
Avg. PM 2.5	-0.2644	-0.8888***	-0.5386**	-0.5449***	-0.0738	1.4381
	(0.1646)	(0.3368)	(0.2632)	(0.1982)	(0.7357)	(0.8817)
10-year age FE	Yes	Yes	Yes	Yes	Yes	Yes
Missing: housing vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: employment vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: transit vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: education vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: comorbidity vars	Yes	Yes	Yes	Yes	Yes	Yes
Missing: other vars	Yes	Yes	Yes	Yes	Yes	Yes
N	3,202	1,766	2,619	2,823	543	291
R-squared	0.6463	0.4918	0.4720	0.4387	0.2530	0.3706
Dependent var. mean	8	8	8	6	7	4

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note. Conditional deaths represent the cumulative COVID-19 deaths divided by the cumulative COVID-19 cases (multiplied by 100) reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 4.2.

Table 5: Oaxaca-Blinder decomposition: COVID-19 cases per 100,000

	Black vs. white			Hispanic vs. white		
	Endowments	Coefficients	Interactions	Endowments	Coefficients	Interactions
Case rate difference		624			1112	
Total explained by included covariates	-152	-208	93	-259	-1645	219
% in dense household	-29	-147	56	-85	-321	106
Avg. hh income	-58	27	13	-156	318	118
% using public transit	3	57	-33	-17	-16	6
% college graduates	-74	96	65	12	-25	-22
% with any comorbidity	7	-123	-8	-14	57	12
Avg. % change in time spent at home	0	-291	0	-1	-170	0
Population density (log)	-4	-427	2	0	-538	0
Avg. PQI rate	0	-76	0	0	-11	0
Avg. PM 2.5	2	675	-2	2	-938	-2
Total unexplained (constant)		-609			1521	

Note. Cases represent the cumulative COVID-19 cases per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 5.

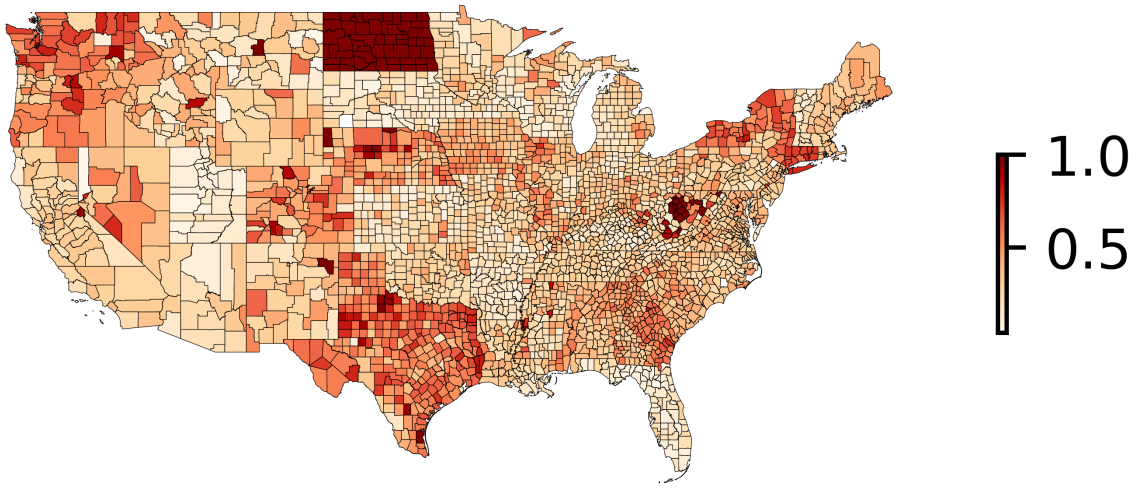
Table 6: Oaxaca-Blinder decomposition: COVID-19 deaths per 100,000

	Black vs. white			Hispanic vs. white		
	Endowments	Coefficients	Interactions	Endowments	Coefficients	Interactions
Mortality rate difference		117			86	
Total explained by included covariates	-55	-303	23	-22	-188	-3
% in dense household	-19	-55	21	-9	-32	11
Avg. hh income	-14	-11	-5	-30	37	14
% using public transit	4	19	-11	-4	-5	2
% college graduates	-30	35	24	19	-30	-26
% with any comorbidity	6	-91	-6	2	-14	-3
Avg. % change in time spent at home	0	-122	0	0	-58	0
Population density (log)	-2	-294	1	0	-81	0
Avg. PQI rate	0	-85	0	0	-38	0
Avg. PM 2.5	1	300	-1	0	32	0
Total unexplained (constant)		335			185	

Note. Deaths represent the cumulative COVID-19 deaths per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 5.

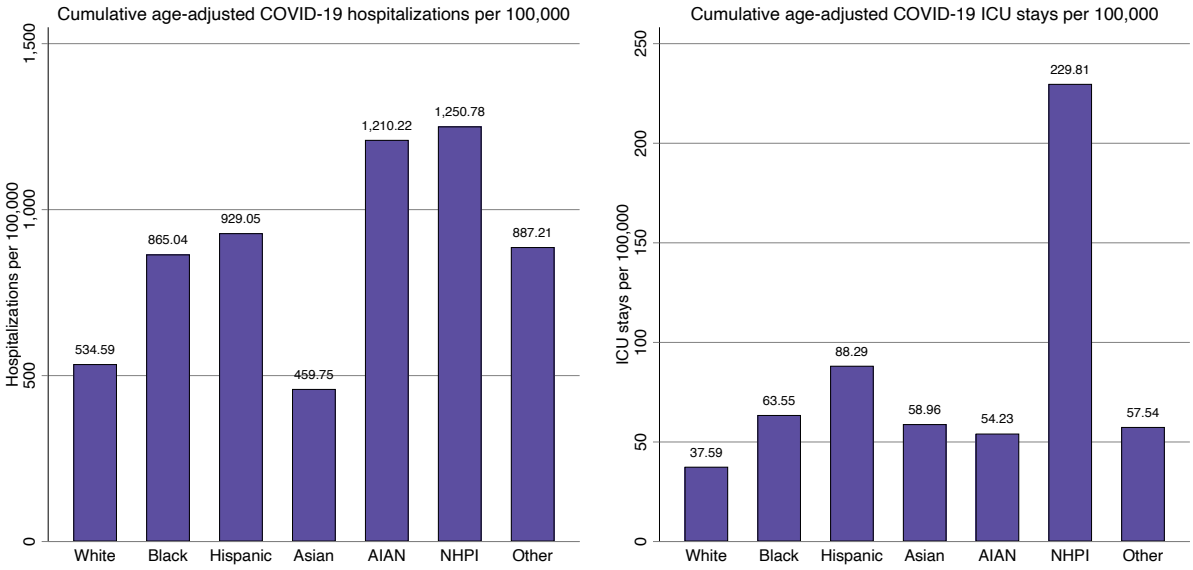
Appendix

Appendix Figure A1: Share of COVID-19 cases with missing race information



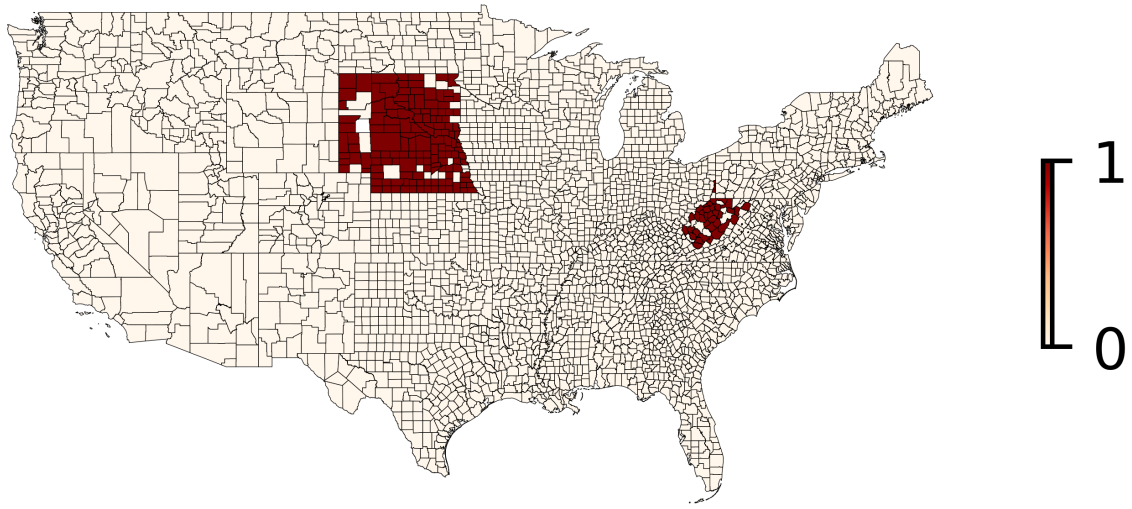
Note. The map plots the share of cumulative COVID-19 cases with missing race data by county, reported as of April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Back to Section 2.1.

Appendix Figure A2: National cumulative COVID-19 hospitalization and ICU rates per 100,000 by race



Note. This figure shows cumulative age-adjusted COVID-19 hospitalizations and ICU stays per 100,000 individuals reported as of April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Age adjustment was performed using the 2010 Decennial Census population. Back to Section 3.1.

Appendix Figure A3: Counties with missing hospitalizations data



Note. The map plots the counties reporting at least one hospitalization associated with COVID-19 as of April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Back to Section 3.1.

Appendix Table A1: Regressions of county-level cumulative COVID-19 cases per 100,000 by race on characteristics of other racial groups, controlling for own-race characteristics

	(1) Asian b/se	(2) Black b/se	(3) Hispanic b/se	(4) AIAN b/se	(5) NHPI b/se
Characteristics that vary by race					
% in dense household for whites	-2.834	-32.867***	-50.112***	-2.061	-55.922**
	(12.304)	(6.629)	(12.151)	(3.941)	(23.479)
Avg. hh income for whites	-0.008***	-0.004*	-0.010***	-0.003**	-0.021***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.006)
% using public transit for whites	-5.497	44.175**	-7.747	6.219	31.395
	(5.719)	(22.336)	(22.419)	(6.271)	(24.974)
% college graduates for whites	-3.025	-4.403	-3.328	6.495**	13.199
	(3.542)	(6.495)	(8.376)	(2.651)	(13.624)
Comorbidities					
% with any comorbidity for whites	-2.420	-1.710	-3.436	-2.189**	3.491
	(2.401)	(1.812)	(2.322)	(0.853)	(7.856)
Characteristics that do not vary by race					
Avg. % change in time spent at home	15.405***	14.647***	15.565**	1.660	-18.257
	(5.345)	(4.805)	(7.442)	(3.814)	(14.697)
Population density (log)	228.041***	134.326**	405.938***	25.387	349.679**
	(30.401)	(54.803)	(56.566)	(37.846)	(140.282)
Avg. PQI rate	-0.045	-0.030	-0.213**	-0.008	-0.438**
	(0.057)	(0.045)	(0.091)	(0.027)	(0.199)
Avg. PM 2.5	46.005	-89.282*	124.279**	-58.940*	181.580
	(29.370)	(52.716)	(56.126)	(32.302)	(156.743)
Own-race characteristics	Yes	Yes	Yes	Yes	Yes
10-year age FE	Yes	Yes	Yes	Yes	Yes
Missing: housing vars	Yes	Yes	Yes	Yes	Yes
Missing: employment vars	Yes	Yes	Yes	Yes	Yes
Missing: transit vars	Yes	Yes	Yes	Yes	Yes
Missing: education vars	Yes	Yes	Yes	Yes	Yes
Missing: comorbidity vars	Yes	Yes	Yes	Yes	Yes
Missing: other vars	Yes	Yes	Yes	Yes	Yes
N	2,496	2,542	2,568	2,006	844
R-squared	0.0337	0.0786	0.0762	0.0247	0.0297
Dependent var. mean	548	1,022	1,646	302	1,149

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note. Cases represent the cumulative COVID-19 cases per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 4.1.

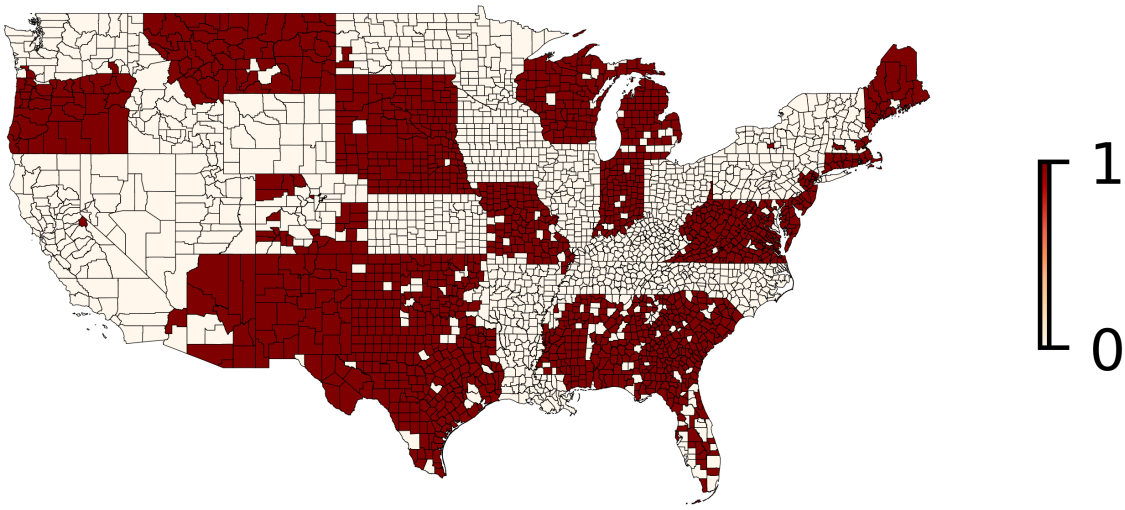
Appendix Table A2: Regressions of county-level cumulative COVID-19 deaths per 100,000 by race on characteristics of other racial groups, controlling for own-race characteristics

	(1) Asian b/se	(2) Black b/se	(3) Hispanic b/se	(4) AIAN b/se	(5) NHPI b/se
Characteristics that vary by race					
% in dense household for whites	3.102*	-0.714	0.337	-0.941	-0.447
	(1.736)	(3.007)	(2.051)	(0.771)	(1.264)
Avg. hh income for whites	-0.001***	0.000	-0.001**	-0.001*	-0.000
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
% using public transit for whites	-1.719	16.076	7.808	-1.209	-2.392*
	(2.482)	(10.923)	(7.189)	(1.287)	(1.279)
% college graduates for whites	-3.183***	-5.459**	-4.757**	2.514**	0.819
	(0.861)	(2.586)	(2.368)	(0.999)	(1.138)
Comorbidities					
% with any comorbidity for whites	-0.343	1.016	-0.482	-0.459*	-0.045
	(0.373)	(1.095)	(0.492)	(0.267)	(0.383)
Characteristics that do not vary by race					
Avg. % change in time spent at home	3.108***	5.459*	6.173***	0.189	-1.344
	(1.020)	(2.898)	(1.426)	(0.981)	(1.354)
Population density (log)	66.548***	58.746***	91.385***	7.824	19.543
	(9.435)	(22.512)	(15.030)	(10.772)	(12.517)
Avg. PQI rate	-0.011	-0.026	-0.022*	0.009	-0.007
	(0.009)	(0.022)	(0.013)	(0.008)	(0.011)
Avg. PM 2.5	-8.635	-15.588	-14.586	-11.156*	-6.423
	(8.792)	(15.543)	(12.408)	(6.650)	(17.955)
Own-race characteristics	Yes	Yes	Yes	Yes	Yes
10-year age FE	Yes	Yes	Yes	Yes	Yes
Missing: housing vars	Yes	Yes	Yes	Yes	Yes
Missing: employment vars	Yes	Yes	Yes	Yes	Yes
Missing: transit vars	Yes	Yes	Yes	Yes	Yes
Missing: education vars	Yes	Yes	Yes	Yes	Yes
Missing: comorbidity vars	Yes	Yes	Yes	Yes	Yes
Missing: other vars	Yes	Yes	Yes	Yes	Yes
N	2,496	2,542	2,568	2,006	844
R-squared	0.2008	0.0592	0.1976	0.0344	0.0621
Dependent var. mean	84	201	164	36	48

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note. Deaths represent the cumulative COVID-19 deaths per 100,000 individuals reported up to June 1, 2020 to the National Vital Statistics and the Centers for Disease Control. Population counts by 10-year age categories and race were obtained from the 2010 Decennial Census (10%). Socio-economic characteristics were obtained from the 2016-2019 American Community Survey (ACS). Health-related characteristics come from the 2010-2012 Behavioral Risk Factor Surveillance System (BRFSS). Data on time at home was obtained from SafeGraph as of May 31, 2020. Pollution data comes from **?** for 2012-2019. Age-adjusted Prevention Quality Indicator (PQI) data was obtained through Centers for Medicare and Medicaid Services for 2014-2018. Land area size from the 2018 Census Gazetteer Files was combined with county-level population data from the 2010 Census to calculate population density for counties. Back to Section 4.2.

Appendix Figure A4: Counties with missing ICU data



Note. The map plots the counties reporting at least one ICU stay associated with COVID-19 as of April 17, 2022 to the National Vital Statistics and the Centers for Disease Control. Back to Section 3.1.