Racial Disparities in Excess All-Cause Mortality During the Early COVID-19 Pandemic Varied Substantially Across States

SEHSD Working Paper 2021-01

Maria Polyakova
Victoria Udalova
Geoffrey Kocks
Katie Genadek
Keith Finlay
Amy Finkelstein*

Abstract

The impact of the COVID-19 pandemic has been starkly unequal across race and ethnicity. We examined the geographic variation in excess all-cause mortality by race to better understand the impact of the pandemic. We used individual level administrative data on U.S. population between January 2011 and April 2020 to estimate the geographic variation in excess all-cause mortality by race. All-cause mortality allows a better understanding of the overall impact of the pandemic than mortality attributable to COVID-19 directly. Nationwide, adjusted excess all-cause mortality was 6.8 per 10,000 for Black individuals, 4.3 for Hispanic individuals, 2.7 for Asian individuals, and 1.5 for White individuals. Nationwide averages mask substantial geographic variation. For example, despite similar excess White mortality, Michigan and Louisiana had markedly different excess Black mortality, as did Pennsylvania compared to Rhode Island. Wisconsin experienced no significant White excess mortality but had significant Black excess mortality. Further work understanding the causes of geographic variation in racial disparities – the relevant roles of social and environmental factors relative to co-morbidities, and of direct and indirect health effects of the pandemic – is crucial for effective policy making.

* Polyakova is an Assistant Professor in the Center for Health Policy at Stanford University School of Medicine and a Faculty Research Fellow at the National Bureau of Economic Research; Udalova is a Research Economist and the Enhancing Health Data (EHealth) program manager at the U.S. Census Bureau; Kocks is a graduate student in the Department of Economics at Massachusetts Institute of Technology; Genadek is a Research Economist at the U.S. Census Bureau and the University of Colorado - Boulder; Finlay is a Research Economist at the U.S. Census Bureau; Finkelstein is a Professor of Economics at Massachusetts Institute of Technology and a Research Associate at the National Bureau of Economic Research. Dr. Polyakova and Dr. Udalova contributed equally as co-first authors. We thank National Institute on Aging (Grant number (Finkelstein, R01-AG032449) and National Institute for Health Care Management (NIHCM) Foundation for financial support, and James Okun for outstanding research assistance. Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results were approved for release by the Disclosure Review Board of the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy. The published version of this paper can be obtained here: https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.02142.

Introduction

The COVID-19 pandemic in the United States has led to a sharp rise in all-cause mortality nationwide starting in March 2020 and continuing in subsequent months. Excess deaths stem from a combination of the direct effects of viral infections with SARS-CoV-2 novel coronavirus, as well as the indirect effects of wide-reaching societal changes associated with the pandemic.[1–5]

Many observers have emphasized large differences in the impact of the pandemic across demographic and socio-economic groups.[6–9] While some differences, such as the age gradient, have clear explanations based on biological pathways, the reasons for other gaps warrant further examination. This is particularly true of racial disparities in the pandemic's impact. Nationwide, there are stark differences in excess all-cause mortality by race, as well as in case fatality rates for COVID-19 infections.[10–14]

The reasons for these racial disparities are poorly delineated, making it hard to formulate evidence-based mitigation policies. One crucial question that policy-makers face in this context is whether racial disparities in all-cause mortality stem predominantly from disparities in the direct effect of the novel coronavirus infections, such as higher infection rates or higher case fatality rates, or, alternatively, if racial disparities in all-cause mortality are driven by the indirect effects of the pandemic, such as disparities in the effect of the pandemic on livelihoods and associated excess morbidity and mortality.

An important first input into this discussion is a measurement of the overall effect of the pandemic on different demographic groups, both nationally as well as across different geographies. While national and geographic variation in mortality associated with COVID-19 directly has been widely reported,[1-11,13] less evidence has been available on the pandemic's differential impact on all-cause mortality across demographic groups *and* geographies.

To fill in this gap, we drew on individual-level administrative data covering the near universe of the U.S. population from January 2011 through April 2020. This provided demographic information on age, race, sex, state of residence, and date of death (if any). We used these data to estimate excess all-cause mortality (hereafter: "excess mortality") separately for seven race groups during the first full month of the COVID-19 pandemic (April 2020). We report estimates for the whole nation and separately by state. We report unadjusted estimates, as well as estimates adjusted to a standardized population by demographics.

Methods

Institutional Review Board determination was obtained through Stanford University; this research was determined to not involve human subjects as defined in 45 CFR 46. The analysis used pre-existing deidentified data.

Data

We used the U.S. Census Bureau's version of the Social Security Administration's Numerical Identification (Census Numident) database covering the U.S. population and deaths from January 2011 to April 2020 (inclusive) to measure the all-cause monthly mortality rate. To our knowledge,

this represents one of the first studies to use the Census Numident to assess the ongoing mortality effects of the pandemic. The Census Numident covers all individuals with a Social Security Number (SSN) regardless of their geographic location. The dataset is cumulative, adding individuals as they receive SSNs upon birth or arrival to the US. Deceased individuals are not removed from the data. For each individual we observed a date of birth. For deceased individuals, we observed a date of death. The date of death is recorded regardless of whether the individual died inside or outside the United States. The version of the Census Numident available to us was released on August 27, 2020 and included deaths through May 2020. The last month of death records in each release of the Census Numident data tend to be incomplete due to the delays in the reporting of deaths. Our analysis thus included deaths only through April 2020. Sections 1.1 and 1.2 of the online Appendix describes our data sources in more detail.[15]

Death counts in the SSA Numident, the main source for the Census Numident, differ from another major source of U.S. vital statistics – those released by the National Center for Health Statistics (NCHS) of the Center for Disease Control and Prevention (CDC) – but provided three distinct advantages for our purposes. First, the Numident data provided an internally consistent numerator and denominator for measuring mortality, as it records not only the deceased, but also living individuals at any given moment in time; this denominator is not available in the CDC vital statistics measure. Second, we were able to link mortality records at the individual level to other demographic information about individuals, allowing us to estimate excess mortality by race and adjust these estimates to a standardized distribution of demographics by race, both nationally and by state, as discussed below. Third, we were able to use a self-reported record of race, potentially improving upon CDC vital statistics office data that use proxy race reports from funeral directors.[16]

As has previously been documented in the literature, estimates of death counts and characteristics of the deceased, such as race, differ between the SSA Numident and the CDC because of different underlying reporting mechanisms.[17] In addition, the CDC counts all deaths that occurred on U.S. territory regardless of nationality or immigration status, but does not cover deaths of U.S. persons outside of the U.S. Section 1.3 of the online Appendix discusses April 2020 difference in death counts by race and state between our baseline analytic dataset and CDC vital statistics.[15]

We measured individuals' sex and age (based on date of birth) in the Census Numident data. The U.S. Census Bureau's annual address database, Master Address File - Auxiliary Reference File (MAF-ARF), was used to attach a county and state of residence to each individual-month observation, when available.[15]

Self-reported race information was drawn primarily from the 2010 Decennial Census. When no record of race was available from 2010 Decennial Census, we used the race variable recorded in the Census Bureau's 2010 Modeled Race file. We analyzed the following seven race categories: Hispanic, non-Hispanic White (White), non-Hispanic Black (Black), non-Hispanic Asian (Asian), non-Hispanic American Indian and Alaska Native (American Indian and Alaska Native), non-Hispanic Native Hawaiian and Other Pacific Islander (Native Hawaiian and Other Pacific Islander), and non-Hispanic individuals of some other race/two or more races.

The monthly mortality rate for each demographic group of interest was defined as the ratio of the count of individuals whose death date fell within that month, divided by the count of individuals who were alive at the beginning of the month.

Statistical Analysis

Data were collapsed into counts of alive and deceased individuals by sex, age, race, county, month, and year. Monthly mortality rates from January 2011 through April 2020 were computed for each sex, age, race, county, and month-year combination.

Predicted and excess unadjusted all-cause mortality were computed based on a linear regression model. The outcome variable was the monthly mortality rate by demographic group. The right-hand side variables were a linear annual time trend to capture secular trends, indicator variables for each calendar month to non-parametrically capture seasonal variation, and indicator variables for January, February, March, and April 2020 to capture deviations from historical trend – if any – during the first four months of 2020. All slopes and intercepts were allowed to be race-specific. Section 2 of the Appendix shows the regression equation and provides more information about our specifications.[15]

Regression coefficients on the interaction between race indicators and the indicator for April 2020 directly measured the *difference* in excess mortality between racial groups. The *level* of excess mortality for each race (relative to a race-specific historical trend) was obtained from combining these regression coefficients with the other race-specific parameters of the regression model.

To adjust our estimates of excess mortality across races to a standardized distribution of demographics, we first estimated an augmented version of our baseline regression model that included differential intercepts and differential deviations from historical trends in January, February, March, and April 2020 for specific demographics: sex, individual state, and/or 5-year age group. We also estimate the demographically-augmented model separately by state. The

coefficients on the demographic variables were then used to obtain national estimates of adjusted excess mortality for each race for a standardized population that had the same distribution of sex, age (in 5-year age groups), and states of residence as the full baseline analytic dataset in April 2020. In other words, each race-specific estimate was adjusted to match the national distribution of age, sex, and state in April 2020. State-specific adjusted estimates of excess mortality were obtained for the same standardized distribution of sex and age.

All regressions were estimated using Stata version 16.1 (StataCorp) software on data collapsed by sex, age, race, county, month, and year and were weighted by the number of individuals who were alive at the beginning of the month in each data cell. 95% confidence intervals for the levels and gaps in excess all-cause mortality by race were computed using heteroskedasticity- robust standard errors and the delta method.

Limitations

The main limitation of our analysis was that information on the date of birth, date of death, race, sex, and geographic location was not available for some individuals residing in the U.S.

Race information was not available for any individuals born after 2010; our analysis therefore excluded children age 10 and younger. As mortality rate is very low in this population, omitting children is unlikely to substantially affect our main results.

Race data from a combination of administrative records, survey, and census data were used in our baseline analysis. Our results were similar when limiting our analysis only to individuals for whom we observe a 2010 Decennial Census race.[15]

Our baseline analysis was limited to individuals for whom a state record was observed and who resided in 50 U.S. States or the District of Columbia. Our unadjusted estimates of excess mortality were unaffected when including individuals who resided outside of 50 states or District of Columbia or for whom no geographic record was observed.[15] As the Census Numident database is based on the records of the Social Security Administration, mortality for individuals living in the U.S., but not captured in the SSA records (for example, individuals without Social Security Numbers), was not measured.

Finally, some individuals were missed in our analysis due to a lack of valid linkage keys. A unique individual level anonymous identifier common across all data sources was used to link the Census Numident to other data sources. The linkage keys were created using personally identifiable information and probabilistic record linkage. Prior literature shows that linkage key assignment may be non-random as immigrants, young people, and minorities are less likely to receive these keys.[18] Our analysis required individuals to have valid linkage keys.

Results

Baseline Analytic Dataset

The dataset included 27 billion person-month observations and 22.4 million deaths for individuals aged 11 to 99 years (inclusive) between January 1, 2011 and April 30, 2020. The mean age in the full dataset was 44 years, and 51% were female. The dataset was 67% White, 14% Hispanic, 12% Black, 5% Asian, 0.9% American Indian or Alaska Native, 0.2% Hawaiian or Pacific Islander, and 0.4% reporting two or more races or other race. The April 2020 data included 241.5 million

individuals aged 11 to 99 and 276,000 deaths. The distribution of age, sex, and geographic location in April 2020 was similar to the complete time range (Exhibit 1).

National all-cause excess mortality by race

Exhibit 2 shows raw monthly all-cause mortality for April of each year, separately for 3 race categories (results for Hawaiian and Pacific Islander individuals, American Indian and Alaskan individuals, Asian individuals, and individuals who reported having "Other and 2 or more races" can be found in Appendix Section 3.1).[15] It also superimposes the regression fit from the historical April mortality trend for each race. Two clear facts emerge. First, for all races, historical April mortality has followed a stable, close to linear, slightly upward trend over the last decade. Second, for all races, there is a pronounced upward deviation from this trend in April 2020. The difference between the observed and predicted values in April 2020 measures national race-specific excess all-cause mortality in April 2020.

These estimates are reported in Exhibit 3. Column (1) shows unadjusted excess mortality per 10,000 individuals by race; all racial groups exhibited excess mortality (p-value < 0.05 for two-sided null hypothesis of zero excess mortality). Hawaiian and Pacific Islander unadjusted excess mortality was the lowest observed of any group, at 1.3 excess deaths per 10,000 individuals, a 22% increase relative to the predicted rate for April 2020 of 5.9 deaths. Black unadjusted excess mortality was the highest observed, at 6.1 excess deaths per 10,000, a 79% increase relative to the predicted rate of 7.7 deaths. Unadjusted excess mortality was 2.1 among White individuals (21% increase from the predicted rate of 10.2 deaths), 2.7 among Hispanic individuals (64% increase), 2.9 among Asian individuals (64% increase), 1.9 among American Indian and Alaska Native individuals (22% increase), and 2.7 among those with other or two or more races (60% increase).

Differences relative to White unadjusted excess mortality were statistically significant at the 5% level for Hispanic, Black, and Asian individuals (Appendix Section 3.2).[15]

Adjusting for differences in age distributions is important for comparing all-cause mortality by race, as age distributions differ starkly across races (Appendix Section 3.3).[15] Adjusting for age exacerbates differences in excess mortality between White individuals and each other race (column 2). For example, the difference between Hispanic and White excess mortality rises from 0.62 deaths per 10,000 (unadjusted) to over 2 deaths once age-adjusted. Highest excess mortality is still observed among Black Americans, with 6.8 excess age-adjusted deaths per 10,000 individuals, a gap of 5.2 relative to White individuals. By contrast, adjusting for differences in the sex distribution between races does not meaningfully change our estimates of the levels or gaps in excess mortality for any race (Appendix Section 3.2).[15]

Adjusting for the state of residence has an important effect on the estimates of levels and differences in excess mortality for smaller racial groups (Appendix Section 3.2).[15] In particular, it substantially increases estimated excess mortality for American Indian and Alaska Native individuals, as well as for Hawaiian and Pacific Islander individuals. This, in turn, decreases the estimated gaps in excess mortality between these racial groups and White individuals.

Exhibit 3 column 3 reports results adjusting for sex, age, and state simultaneously. This allows us to compare excess mortality rates across races for a standardized distribution of sex, age, and state of residence. This adjusted excess mortality is lowest for White individuals at 1.5 excess deaths per 10,000 and highest for Black individuals at 6.8 excess deaths per 10,000. Relative to White individuals, adjusted excess mortality rates are higher by 2.7 deaths per 10,000 for Hispanic

individuals, 1.2 for Asian individuals, 2.4 for American Indian and Alaska Native individuals, and 2.7 for Hawaiian and Pacific Islander individuals.

Racial disparities by state

Age- and sex- adjusted excess mortality showed substantial geographic variation across states for each race. Appendix Section 3.4 reports point estimates and 95% confidence intervals of White, Black, and Hispanic excess mortality for each state.[15] The (unweighted) interquartile range across states was 0.09 to 1.4 excess deaths per 10,000 for White individuals; 1.2 to 4.8 excess deaths per 10,000 for Black individuals and 0.13 to 2.1 excess deaths per 10,000 for Hispanic individuals.

While Black and Hispanic individuals experienced higher adjusted excess mortality than White individuals in nearly every state, Exhibits 4 and 5 shows that these racial disparities differ substantially by state. We observed particularly staggering levels and racial differences in excess all-cause mortality in New York and New Jersey - the two states that were affected the most by the first wave of the pandemic. White non-Hispanic excess all-cause mortality was 7.2 per 10,000 in New York and 8.6 in New Jersey. Black individuals, however, experienced excess all-cause mortality that was 4.6 and 2.9 *times* higher in New York and New Jersey, respectively, amounting to 32.7 per 10,000 age-sex adjusted excess all-cause deaths among Black individuals in New York and 24.7 in New Jersey. Hispanic individuals in the two states fared only slightly better with 27.2 per 10,000 age-sex adjusted deaths in New York and 20.5 in New Jersey.

In general, the difference between Black excess mortality or Hispanic excess mortality and White excess mortality was larger in states where White excess mortality was higher. But even states with similar levels of adjusted White excess mortality exhibited substantial variation in the level

of adjusted Black or Hispanic excess mortality. Michigan for instance, had much higher adjusted Black excess mortality (18 excess deaths per 10,000) than Louisiana (10 per 10,000), even though the two states had comparable adjusted White excess mortality (3 per 10,000). Likewise, Pennsylvania had much higher adjusted Black excess mortality (11 excess deaths per 10,000) than Rhode Island (3 per 10,000), even though the two states had comparable adjusted White excess mortality (2 per 10,000). As another example, Hispanics experienced higher adjusted excess mortality in Pennsylvania than in Delaware even though White adjusted excess mortality was similar in both states. All reported comparisons are statistically significant at 5% confidence level (Appendix Section 3.4).[15]

In addition, in several states, Black and Hispanic Americans experienced an increase in mortality in April 2020, while White individuals did not. For, example, in Wisconsin, we estimate adjusted excess mortality for Black individuals to be 4.6 per 10,000 (95% CI: 2.9 to 6.3) and for Hispanic individuals to be 1.4 4 per 10,000 (95% CI: 0.46 to 2.33), while White adjusted excess mortality was a statistically insignificant 0.27 per 10,000 (95% CI: -0.04 to 0.6). In Kentucky, Alabama, South Carolina, California, and Washington, Black adjusted excess mortality was over 1.5 per 10,000 individuals respectively (p-value<0.05 for the null of zero adjusted excess mortality in each state) while White adjusted excess mortality was under 0.5 per 10,000 in each of those states.

Discussion

Excess all-cause mortality differed substantially across racial groups during the early spread of the novel Coronavirus. Our data allowed us to examine racial disparities in national excess all-cause mortality for seven different racial groups, to examine the role of demographic differences in these

racial disparities, and to examine differences in racial disparity by state. The results indicate pronounced differences in the overall impact of the pandemic across racial groups and in the extent of racial disparities across states.

All racial groups experienced substantial excess mortality in the first full month of the pandemic. When adjusted to a standardized distribution of sex, age, and state of residence, White and Asian Americans had the lowest excess mortality (at 1.5 and 2.7 excess deaths per 10,000), while Black Americans had the highest (at 6.8 excess deaths). Differences by race were more pronounced when adjusted for demographic differences, highlighting important differences in the age and geographic distributions by race. Crude excess mortality overestimates the mortality effect for White Americans, who are on average older, and underestimates it for other races who are on average younger. Our unadjusted national estimates of all-cause excess mortality are similar, but slightly higher, compared to those based on CDC data releases (2.75 per 10,000 in our data versus 2.4 per 10,000 based on CDC data);[1,2,19] compared to prior estimates of COVID-specific deaths, our estimates suggest that 35% of excess deaths during the first month of the pandemic are not directly attributable to the infections with the novel coronavirus. Our results are qualitatively consistent with findings of pronounced racial disparities in COVID-19-specific deaths nationwide.[6,7,9]

Racial disparities in adjusted excess all-cause mortality varied substantially across states. States that exhibited the highest levels of White excess mortality of 7 to 8 excess deaths per 10,000 and were generally the most affected by the first wave of the pandemic, New York and New Jersey, experienced staggering levels of Black non-Hispanic and Hispanic excess mortality ranging from 20 to more than 30 excess deaths per 10,000 individuals. Some states, such as Michigan and

Pennsylvania, had substantially higher Black excess mortality than other states with similar White excess mortality (Louisiana and Rhode Island, respectively). In addition, several states that experienced virtually no White excess mortality in April 2020 exhibited substantial Black excess mortality. For example, Wisconsin experienced significant Black excess mortality of 4.6 per 10,000 (95% CI: 2.9 to 6.3), while White excess mortality was statistically indistinguishable from zero (point estimate 0.27, 95% CI -0.04 to 0.6).

These findings are consistent with the growing literature that has documented the broader importance of geography for understanding the patterns of health disparities in the US.[20] While geographic difference in the overall *level* of excess mortality likely reflects the different timing of COVID-19's spread, the causes of the stark variation in racial disparities across states are less clear. The geographic patterns that we document for the first full month of the pandemic may point to an important role of the indirect effects of the pandemic on non-White individuals. In the states where we observe excess mortality among White individuals that is close to zero, it is possible that the community transmission of the virus was still limited, while the indirect effects of the pandemic, such as the impact on the economy that has been documented to be much more geographically spread early on,[2] was already affecting the non-White individuals.

As direct and indirect impacts of the pandemic would call for different mitigating policies---for instance, different vaccine distribution priority in the case of direct effects, and differential economic interventions in the case of indirect effects---understanding what drives the variation in all-cause mortality across states and racial groups is crucial for evidence-based policies. In addition, the geographic variation in the *extent* of racial disparities in the pandemic's early and

later impact may provide an opportunity to examine the relative roles of social and environmental factors (such as occupational and residential segregation) and underlying co-morbidities in contributing to these racial disparities; such evidence would have important implications both for policies aimed at mitigating racial disparities during the pandemic, and potentially more broadly beyond the pandemic. In this paper we demonstrate an enormous potential of linking administrative records and Census responses at the individual level to explore these future research questions.

An important limitation of our work is that we cannot pinpoint the drivers of the geographic variation in the racial disparities of the pandemic's impact. Several hypotheses about the underlying causes of racial disparities in COVID's mortality impact have been put forward, including racial differences in education and occupation,[21-24] neighborhood characteristics such as safety, food availability, and pollution,[22,24-26] risk of infection,[27] access to healthcare,[28] and comorbidities.[24,29] Our results suggest that factors driving racial disparities may well differ across states. Assembling additional empirical evidence on the sources of geographic variation in racial disparities during the time period we studied as well as later in the pandemic will likely help shed light not only on the underlying reasons for the unequal impact of the pandemic on mortality across racial groups in the US and best associated policy responses, but also on the drivers of racial health disparities more generally.

Endnotes

1. Rossen LM. Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity — United States, January 26–October 3, 2020. MMWR Morb Mortal Wkly Rep [Internet] 2020 [cited 2020 Oct 22];69. Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6942e2.htm

2. Polyakova M, Kocks G, Udalova V, Finkelstein A. Initial economic damage from the COVID-19 pandemic in the United States is more widespread across ages and geographies than initial mortality impacts. PNAS [Internet] 2020 [cited 2020 Oct 22];Available from: https://www.pnas.org/content/early/2020/10/19/2014279117

3. Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L. Excess Deaths From COVID-19 and Other Causes, March-April 2020. JAMA [Internet] 2020 [cited 2020 Aug 3];Available from: https://jamanetwork.com/journals/jama/fullarticle/2768086

4. Weinberger DM, Chen J, Cohen T, Crawford FW, Mostashari F, Olson D, et al. Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020. JAMA Intern Med [Internet] 2020 [cited 2020 Aug 3];Available from: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2767980

5. Katz J, Lu D, Sanger-Katz M. Tracking the Real Coronavirus Death Toll in the United States [Internet]. The New York Times. [cited 2020 Aug 3];Available from: https://www.nytimes.com/interactive/2020/05/05/us/coronavirus-death-toll-us.html

6. Bassett M, Chen JT, Krieger N. The unequal toll of COVID-19 mortality by age in the United States: Quantifying racial/ethnic disparities. HCPDS Working Paper [Internet] 2020;19(3). Available from: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1266/2020/06/20_Bassett-Chen-Krieger_COVID-19_plus_age_working-paper_0612_Vol-19_No-3_with-cover.pdf

7. Ford T, Reber S, Reeves RV. Race gaps in COVID-19 deaths are even bigger than they appear [Internet]. Brookings. 2020 [cited 2020 Aug 3];Available from: https://www.brookings.edu/blog/up-front/2020/06/16/race-gaps-in-covid-19-deaths-are-even-bigger-than-they-appear/

8. COVID-19 deaths analyzed by race and ethnicity [Internet]. APM Research Lab. 2020 [cited 2020 Oct 22];Available from: https://www.apmresearchlab.org/covid/deaths-by-race

9. Goldstein JR, Atherwood S. Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States [Internet]. Epidemiology; 2020 [cited 2020 Oct 22]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.05.21.20109116

10. Cromer SJ, Lakhani CM, Wexler DJ, Udler M, Patel CJ, Cromer D. Geospatial Analysis of Individual and Community-Level Socioeconomic Factors Impacting SARS-CoV-2 Prevalence and Outcomes. 2020;30.

11. McCarty TR, Hathorn KE, Redd WD, Rodriguez NJ, Zhou JC, Bazarbashi JN, et al. How Do Presenting Symptoms and Outcomes Differ by Race/Ethnicity Among Hospitalized Patients with COVID-19 Infection? Experience in Massachusetts. Clin Infect Dis [Internet] 2020 [cited 2020 Oct 22];Available from: https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa1245/5896009

12. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. New England Journal of Medicine 2020;382(26):2534–43.

13. Rentsch CT, Kidwai-Khan F, Tate JP, Park LS, King JT Jr., Skanderson M, et al. Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: A nationwide cohort study. PLOS Medicine 2020;17(9):e1003379.

14. Yehia BR, Winegar A, Fogel R, Fakih M, Ottenbacher A, Jesser C, 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.

15. Appendix is attached at the end of this paper.

16. Rothwell CJ, Madans JH, Atkinson D, Ni H. The Validity of Race and Hispanic-origin Reporting on Death Certificates in the United States: An Update. Vital and Health Statistics (172):29.

17. Barbieri M. Investigating the Difference in Mortality Estimates between the Social Security Administration Trustees' Report and the Human Mortality Database. SSRN Journal [Internet] 2018 [cited 2020 Oct 22];Available from: https://www.ssrn.com/abstract=3376712

18. The United States Census Bureau. 2010 Census Match Study Report. [Internet] The United States Census Bureau 2010 [cited 2020 Dec 15]; Available from: https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_247.pdf

19. Flagg A, Sharma D, Fenn L, Stobbe M. COVID-19's Toll on People of Color Is Worse Than We Knew [Internet]. The Marshall Project. 2020 [cited 2020 Oct 22];Available from: https://www.themarshallproject.org/2020/08/21/covid-19-s-toll-on-people-of-color-is-worse-than-we-knew

20. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, et al. The Association Between Income and Life Expectancy in the United States, 2001-2014. JAMA 2016;315(16):1750–66.

21. McLaren J. Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data. [Internet]. Cambridge, MA: National Bureau of Economic Research; 2020 [cited 2020 Oct 22]. Available from: http://www.nber.org/papers/w27407.pdf

22. Ray R. Why are Blacks dying at higher rates from COVID-19? [Internet]. Washington DC: Brookings Institution. 2020 [cited 2020 Oct 22];Available from: https://www.brookings.edu/blog/fixgov/2020/04/09/why-are-blacks-dying-at-higher-rates-from-covid-19/

23. Selden TM, Berdahl TA. COVID-19 And Racial/Ethnic Disparities In Health Risk, Employment, And Household Composition. Health Aff (Millwood) 2020;39(9):1624–32.

24. Logan TD, Hardy B. Racial economic inequality amid the COVID-19 crisis [Internet]. Washington DC: Brookings Institution. 2020 [cited 2020 Dec 10];Available from: https://www.brookings.edu/research/racial-economic-inequality-amid-the-covid-19-crisis/

25. Desmet K, Wacziarg R. Understanding Spatial Variation in COVID-19 across the United States. London: Centre for Economic Policy Research 2020; Available from: https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14842

26. Makridis C, Rothwell J. The real cost of political polarisation: Evidence from the COVID-19 pandemic. Centre for Economic Policy Research, COVID Economics 2020; 34 (3):50-88..

27. Zelner J, Trangucci R, Naraharisetti R, Cao A, Malosh R, Broen K, et al. Racial disparities in COVID-19 mortality are driven by unequal infection risks. [Internet]. Infectious Diseases (except HIV/AIDS); 2020 [cited 2020 Oct 22]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.09.10.20192369

28. Krishnan L, Ogunwole SM, Cooper LA. Historical Insights on Coronavirus Disease 2019 (COVID-19), the 1918 Influenza Pandemic, and Racial Disparities: Illuminating a Path Forward. Ann Intern Med 2020;173(6):474–81.

29. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. JAMA 2020;323(24):2466.

Exhibit 1. Summary Statistics

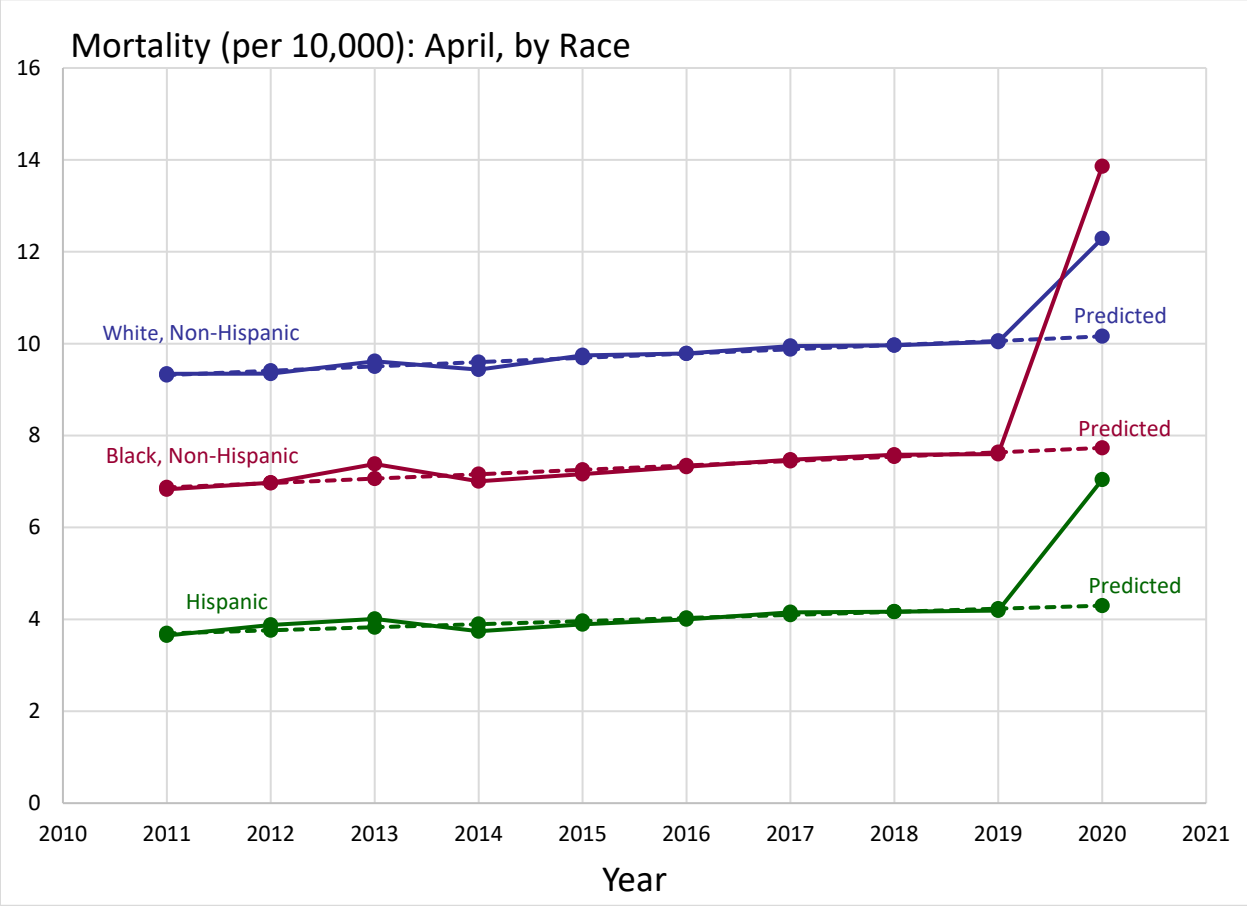
Variable	April 2020	January 2011 to April 2020
Counts		
Individual-Months (#)	241,500,000	26,680,000,000
Deaths (#)	276,000	22,400,000
Summary Statistics		
Average Age (Years)	45.14	44.28
Female (%)	51.17	51.23
Geography		
Northeast (%)	17.43	17.72
Midwest (%)	21.99	22.29
South (%)	37.67	37.34
West (%)	22.91	22.66
Race		
White, Non-Hispanic (%)	65.96	67.36
Hispanic (%)	14.70	13.67
Black, Non-Hispanic (%)	12.57	12.32
Asian, Non-Hispanic (%)	5.27	5.18
American Indian & Alaskan, Non-Hispanic (%)	0.19	0.19
Hawaiian & Pacific Islander, Non-Hispanic (%)	0.40	0.38

Source(s): Authors' calculations from Census Numident 2011-2010, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Notes: Table shows summary statistics for the baseline analytic dataset constructed from the U.S. Census Bureau version of the Social Security Administration's Numerical Identification System (Numident) database covering population and deaths from January 2011 to April 2020. The first

column shows summary statistics for individuals who were alive at the start of April 2020. The second column shows summary statistics for all individuals in our baseline analytic dataset who were alive as of January 1, 2011 and were tracked until April 2020 (inclusively). All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003.

Exhibit 2. Observed and Predicted April Mortality by Race, 2011-2020



Source(s): Authors' calculations from Census Numident 2011-2010, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Notes: Figures show observed and predicted all-cause mortality separately by race in the month of April for years 2011 to 2020. The race specific trends are estimated as discussed in the Statistical Analysis on the full dataset of 26,680,000,000 individual-months from January 2011 through April 2020. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003.

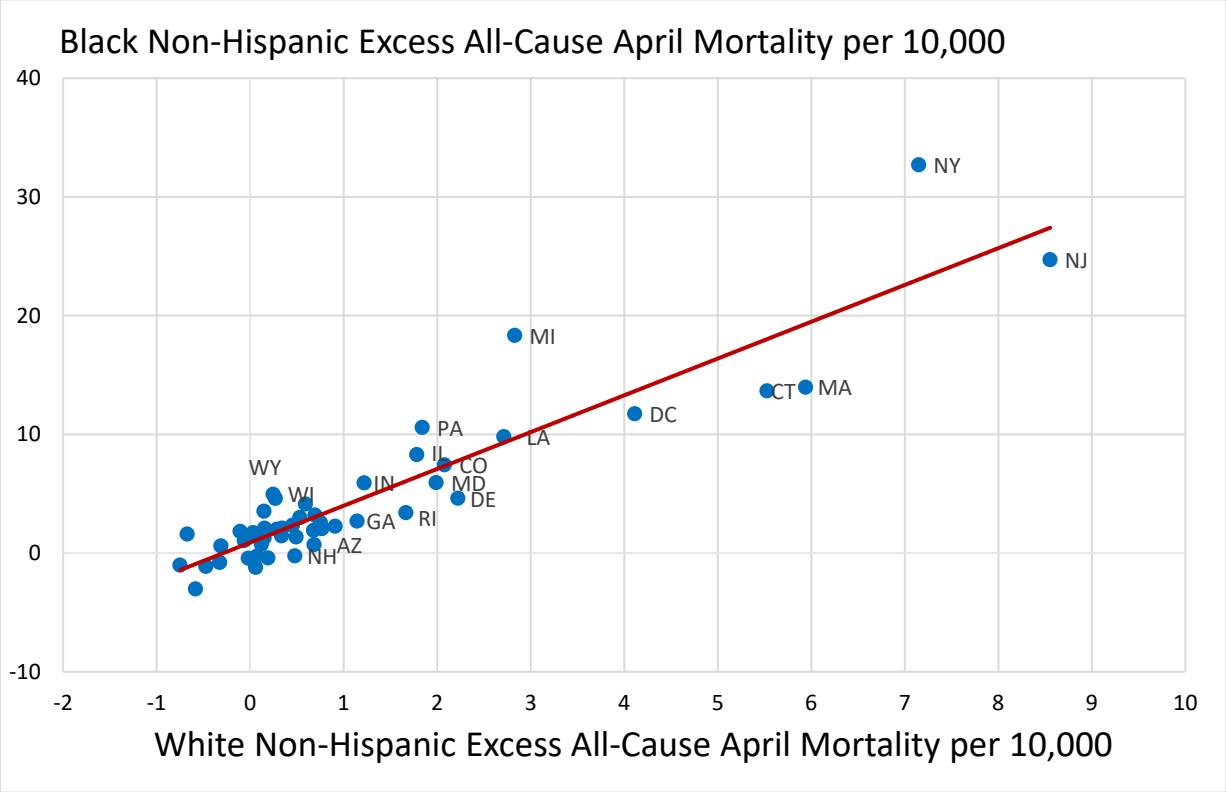
Exhibit 3. Model Estimates of All-Cause Excess Mortality in April 2020, by Race

	Excess Mortality (per 10,000): April 2020					
	Model 1		Model 2		Model 3	
	Unadjusted		Age Adjusted		Age, Sex, and State Adjusted	
Race/ethnicity	Rate	Standard Error	Rate	Standard Error	Rate	Standard Error
White, Non-Hispanic	2.13	0.08	1.64	0.04	1.54	0.04
Hispanic	2.74	0.13	3.75	0.10	4.26	0.10
Black, Non-Hispanic	6.13	0.17	6.85	0.13	6.77	0.13
Asian, Non-Hispanic	2.86	0.16	3.00	0.13	2.73	0.13
American Indian & Alaskan, Non-Hispanic	1.95	0.24	2.81	0.24	3.91	0.24
Hawaiian & Pacific Islander, Non-Hispanic	1.32	0.43	2.46	0.42	4.26	0.43
Other & Two or More, Non-Hispanic	2.74	0.30	4	0.30	3.42	0.30

Source(s): Authors' calculations from Census Numident 2011-2010, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Notes: Table shows regression-based estimates of levels of excess all-cause mortality by race in April 2020. Columns report the results of regression models with or without demographic adjustments as specified in column titles. Heteroskedasticity-robust standard errors are reported in parentheses. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003.

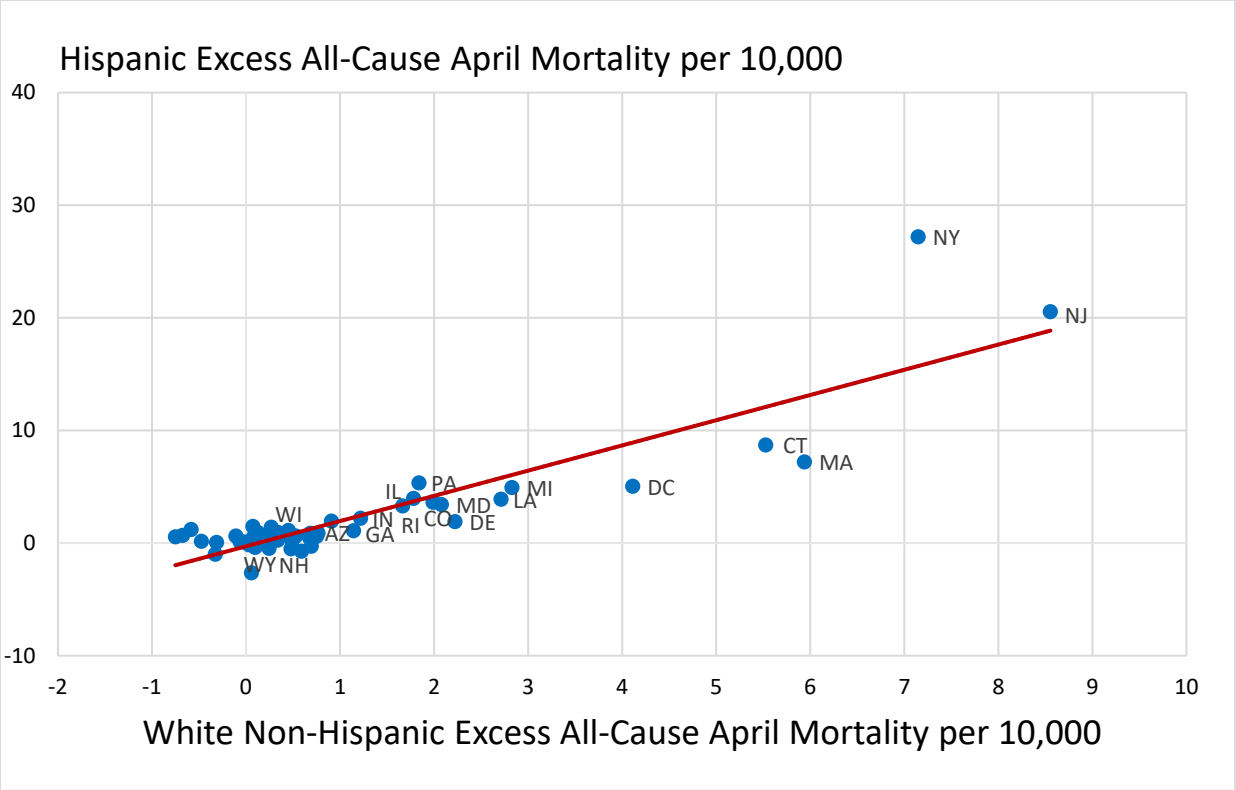
Exhibit 4. Association of Black Non-Hispanic and White Non-Hispanic Excess All-Cause Mortality Across States



Source(s): Authors' calculations from Census Numident 2011-2010, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Notes: Figure reports the association in the estimates of sex and age adjusted all-cause excess mortality in April 2020 by state among non-Hispanic Black individuals vs non-Hispanic White individuals. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis. Red line marks the line of best fit for the relationship between White and Black excess mortality by state. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003.

Exhibit 5. Association of Hispanic and White Non-Hispanic Excess All-Cause Mortality Across States



Source(s): Authors' calculations from Census Numident 2011-2010, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Notes: Panels report the association in the estimates of sex and age adjusted all-cause excess mortality in April 2020 by state among Hispanic vs non-Hispanic White individuals. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis. Red line marks the line of best fit for the relationship between White and Hispanic excess mortality by state. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003.

Online Appendix

Racial Disparities in Excess All-Cause Mortality During the Early COVID-19 Pandemic Varied Substantially Across States

1 Data

1.1 Data Sources

Our data source on mortality is the U.S. Census Bureau's version of the Social Security Administration's Numerical Identification file (Census Numident) (2020 quarter 2 updated vintage). The Census Numident covers all individuals with a Social Security Number (SSN) in the United States. For each individual we observe a date of birth and for deceased individuals we observe a date of death whether they died inside or outside the United States. The Census Numident data is one of only two sources of national mortality data (the other one is the National Death Index (NDI) maintained by the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control and Prevention (CDC)), and it has been used to measure all-cause mortality in previous research (Brown, Kowalski, & Lurie, 2020; Bailey et al., 2020; Miller et al., 2019; Chetty et al., 2016; Chetty et al., 2011). In general, the quality of the Census Numident data has been improving over time (United States Government Accountability Office, 2013; Office of the Inspector General, Social Security Administration, 2018a, 2018b, 2015).

We obtain race information from two main sources: 2010 Decennial Census and 2010 Modeled Race file produced by the Census Bureau based on the 2000 Decennial Census, the 2010 Decennial Census, the Census Numident, the Indian Health Services, and other administrative records. Even though the Census Numident contains the record of race directly, we do not use this measure due to limitations in the SSA's race record.[1]

We obtain geographic location information from 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF). These data allow us to attach annual residential location information to individual records. We assign location information from a given year to all months in that year. For year 2020, we use 2019 address information. Location information in MAF-ARF comes from a variety of Census survey and administrative records sources. For each individual, we obtain an annual address identifier called the Master Address File Identification Number (MAFID). Not every individual and not every address gets a MAFID, indicating either that the address is unknown or that the available address is low-quality.

[1]Specifically, the Numident has highly incomplete race and ethnicity coverage for individuals who were younger than 31 years old in 2020 and for those at older ages. Information is missing for younger individuals because, although beginning in 1989, SSA no longer received race or ethnicity information from SSN applications at birth. While race and ethnicity information can still be obtained from any SSN application submitted to SSA at any point after the birth, this 1989 policy change results in imperfect race and ethnicity coverage for those born in or after 1989. Information on older ages is missing because, although SSA has collected race data on the application for an SSN since the 1930s, paper forms were only digitized in the mid-1970s to create the Numident. At the time of the Numident creation, race information was missing for people who were receiving SSA benefits because the original paper SSN applications remained at field offices throughout the country to assist with determination of eligibility. As a result, the Numident lacks race/ethnicity information for individuals who began receiving SSA benefits on or prior to 1979. More information about SSA race or ethnicity collection history, see here (https://www.ssa.gov/policy/docs/ssb/v62n4/v62n4p9.pdf)

1.2 Baseline Analytic Dataset

We impose several restrictions to the Census Numident before proceeding with our statistical analyses.

Selecting individuals who were alive on January 1, 2011

We start by selecting individuals from the Census Numident who were alive as of January 1, 2011. This step is necessarybeacuse the full Census Numident database may contain individuals who are deceased, but have no recorded death date. Specifically, a well-known concern with the Numident file is that SSA only began to systematically capture death in 1962 and so death coverage may be incomplete prior to 1962. This means that some likely deceased individuals are observed in the Numident without a date of death. The restrictions we make here are imposed on the person-level Census Numident database.

We rely on three additional data sources linked to the Census Numident to determine if individuals are alive: 2010 Decennial Census, 2010 Medicare Enrollment Database (EDB), and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF). If an individual is observed in one or more of these additional datasets, this individual is likely to be alive. Each source has unique advantages over the other data sources but also its own limitations so by combining these multiple sources of data we are able to construct an accurate database of U.S. population alive at the start of our study period.

The Census Numident and other data sources are linked using a unique individual level anonymous identifier common across all data sources, called a Protected Identification Key (PIK). PIKs are created using personally identifiable information (PII) and probabilistic record linkage (Wagner and Lane, 2014). Individuals do not receive PIKs if either their PII are of low quality to assign a valid unique PIK or because they do not have a SSN. Prior literature shows that PIK assignment may be non-random as immigrants, young people, and minorities are less likely to receive a PIK (US Census Bureau, 2010). Our analysis requires individuals to have a valid PIK in at least one data source.

With these data linkages in place, we follow the following steps to select the list of individuals in the Census Numident who were alive on January 1, 2011:

1. **(Step 1)** Drop individuals if there is a record of year of death that is 2010 or earlier. These individuals are with certainty not alive on January 1, 2011. This step eliminates 103,500,000 observations (about 19 percent) from the original Census Numident . 413,200,000 individuals remain.

2. **(Step 2)** Drop individuals who have a missing year and month of birth; we allow for the exact birth day to be missing. After this step, 411,200,000 individuals remain.

3. **(Step 3)** Drop individuals who were younger than 11 before April 2020, or older than 99 in January 2011 or later. Age is computed based on the date of birth. We do not have race information for individuals born after 2010, so our analysis excludes children who are 10 or younger during our study period. We exclude individuals who are 100 years or older, as another way to account for the historic death undercount and data entry errors. 351,100,000 individuals remain after this step.

4. **(Step 4)** Drop individuals if we cannot confirm that they were alive on January 1, 2011 from the death record in Census Numident itself or through any other linked data sources. Specifically, we keep only those individuals who satisfy at least one of the following conditions:

   (a) died during January 2011-April 2020 as recorded in the Census Numident, or

   (b) alive in 2010 Decennial Census, or

(c) alive in 2010 Medicare Enrollment Database (EDB), or

(d) have a MAFID in MAF-ARF in both 2010 and 2011

Table S.1 shows the breakdown of 351,100,000 observations from Step 3 by conditions in (4b), (4c), and (4d) above:

Table S.1: Defining Individuals who are Alive as of Jan. 1, 2011

	Census Numident	MAFID in 2010 and 2011 in MAF-ARF	2010 EDB	2010 Decennial Census	Number of Observations	% of Individuals at Step 3[2]	Alive/Deceased
(1)	Yes	No	No	No	58,590,000	16.7%	Deceased
(2)	Yes	No	Yes	No	1,394,000	0.4%	Alive
(3)	Yes	No	No	Yes	29,910,000	8.5%	Alive
(4)	Yes	Yes	No	No	27,490,000	7.8%	Alive[3]
(5)	Yes	No	Yes	Yes	3,610,000	1.0%	Alive
(6)	Yes	Yes	Yes	No	3,380,000	1.0%	Alive
(7)	Yes	Yes	No	Yes	187,600,000	53.4%	Alive
(8)	Yes	Yes	Yes	Yes	39,190,000	11.2%	Alive

Notes: Table shows the number of observations available in the Census Numident and the three other data sources we use to construct our analytic dataset. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Decennial Census, 2010 Census Modeled Race File, 2010 Medicare Enrollment Database, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Overall, these person-level restrictions primarily exclude individuals whose deaths were not recorded due to incomplete death records prior to 1962, or people who only stayed in the U.S. for a short period of time and left, but still were issued an SSN (for example, foreign students).

After Step 4 we are left with 292,510,000 individual-level observations for individuals with SSN numbers who were alive on January 1, 2011 and older than 10 at some point during January 1, 2011 to April 30, 2020.

Defining race

A record of race was available for individual-level observations from **Step 4** from either 2010 Decennial Census (89% of observations) or from the 2010 Modeled Race File (11% of observations).

In Table S.2 we check the sensitivity of our unadjusted measure of excess mortality by race to the use of the race record from sources other than 2010 Decennial Census. In this check, we remove individuals from our dataset for whom no race was observed in 2010 Decennial Census. The estimates of crude and adjusted excess mortality for each race relative to excess mortality among White individuals are nearly identical to our baseline results in Panel B of Table 2.

[2]Due to the implementation of rouding rules to preserve data privacy, the sum of rows deviates slightly from the separately reported total at Step 3

[3]We note that our process of eliminating likely deceased, but not recorded as such, individuals from the Census Numident is not perfect. In our analysis we keep individuals who have a MAFID in MAF-ARF in both 2010 and 2011, and who do not have a record of death prior to January 1, 2011, but who are not found in either 2010 Decennial Census or 2010 Medicare Enrollment Database. It is possible that some of these individuals have died prior to 2010 but continue to have a valid address. We find that this group (group 4 in Table S.1) consists of many young individuals so many of these individuals were not eligible for Medicare in 2010 (thus would not be in the 2010 Medicare Enrollment Database) or were either missed by the 2010 Decennial Census entirely or simply did not have a valid PIK in that data source (for example, college age individuals can fall in this latter group). For these reasons, we keep individuals in this group. Future work may improve on this cleaning process by utilizing information from the tax returns.

Table S.2: Model Estimates of Relative Excess April All-Cause Mortality (per 10,000): Race Robustness

Race	(1) Unadjusted	(2) Age Adjusted	(3) Sex Adjusted	(4) State Adjusted	(5) Age, Sex, & State Adjusted
White, Non-Hispanic	0	0	0	0	0
	–	–	–	–	–
Hispanic	0.54	2.10	0.54	0.94	2.74
	(0.15)	(0.10)	(0.15)	(0.16)	(0.11)
Black, Non-Hispanic	3.93	5.10	3.94	3.90	5.18
	(0.19)	(0.14)	(0.19)	(0.19)	(0.14)
Asian, Non-Hispanic	0.66	1.45	0.67	0.32	1.36
	(0.18)	(0.13)	(0.18)	(0.19)	(0.13)
American Indian & Alaskan, Non-Hispanic	-0.24	1.16	-0.23	0.96	2.35
	(0.26)	(0.25)	(0.26)	(0.27)	(0.25)
Hawaiian & Pacific Islander, Non-Hispanic	-0.75	0.94	-0.75	0.87	2.88
	(0.45)	(0.44)	(0.45)	(0.47)	(0.45)
Other & Two or More	1.38	3.03	1.39	0.10	1.93
	(0.43)	(0.42)	(0.43)	(0.44)	(0.42)

Notes: Table shows our estimates of gaps in excess all-cause mortality for each race relative to Non-Hispanic Whites in April 2020. The analytic sample only includes individuals for whom race is observed in 2010 Decennial Census. Columns report the results of models with or without demographic adjustments as specified in column titles. Heteroskedasticity robust standard errors are in parentheses. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis and Appendix Section 2. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Geographic restrictions

We next convert the individual-level database of 292,510,000 individuals to individual-month year dataset for 112 months between January 2011 and April 2020. This expands the data to 32.8 billion person-month observations. We then attach information on each individual's residential address for each month between January, 2011 and April 2020. Address information is not available for individuals in the months after their death, or address information for some individuals for some months may be missing.

As the last step, we drop person-month observations without a valid address or with an address outside of 50 US states plus the District of Columbia.[4] After this step, 26.7 billion person-month observations remain and constitute our main analytic dataset.

In Table S.3 we check the sensitivity of our unadjusted measure of excess mortality by race to the address-based data restriction. In this check, we add back in person-month observations for those months when an individual was alive, but an address was either not observed or the address was outside of 50 US states plus the District of Columbia. Note that this only affects the denominator of our mortality measure, since this does not change how we treat individuals with observed data of death. The estimated excess mortality by race is nearly identical to our baseline results.

[4]At this point we also drop a small number of person-month observations during which individuals were younger than 11 for those individuals who turned 11 during our study period.

Table S.3: Model Estimates of Unadjusted Relative Excess April All-Cause Mortality (per 10,000): Geography Robustness

Race	(1) Baseline	(1) No geographic restriction
White, Non-Hispanic	0	0
	–	–
Hispanic	0.62	0.54
	(0.15)	(0.20)
Black, Non-Hispanic	4.00	4.07
	(0.19)	(0.25)
Asian, Non-Hispanic	0.73	0.72
	(0.18)	(0.21)
American Indian & Alaskan, Non-Hispanic	-0.18	-0.12
	(0.25)	(0.29)
Hawaiian & Pacific Islander, Non-Hispanic	-0.81	-0.80
	(0.44)	(0.46)
Other & Two or More	0.61	0.87
	(0.31)	(0.35)

Notes: Table shows our estimates of differences in excess all-cause mortality for each race relative to Non-Hispanic Whites in April 2020. The analytic sample includes individuals with missing addresses as well as those who live in US territories. Columns report the results of models with or without demographic adjustments as specified in column titles. Heteroskedasticity robust standard errors are in parentheses. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis and Appendix Section 2. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

1.3 Comparing Death Counts in Census Numident to CDC Vital Statistics

Table S.4 compares the count of deaths in our analytic dataset constructed from Census Numident data to the count of deaths for April 2020 reported in the National Center for Health Statistics' (NCHS) vital statistics data reported by the Center for Disease Control and Prevention (CDC), by state and for selected races. The differences in counts stem from three sources. First, SSA's Numident and CDC are generally known to arrive at different death counts for the United States (United States Government Accountability Office, 2013; Office of the Inspector General, Social Security Administration, 2018a, 2018b, 2015). Some differences may exist by race in the reporting quality of deaths, but these differences are likely too small to be relevant in our setting (Black et al. 2017). Second, our restrictions on Census Numident that require individuals to have a PIK, have an valid address, and a race record reduce the number of individuals alive and deceased in our analytic dataset relative to the raw Census Numident. Third, CDC does not report deaths for weeks, states, and races with few observations resulting in measurement error in counts. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF), CDC NCHS Vital Statistics Data April 2020.

Table S.4: (Census Numident Analytic Dataset Deaths)/(CDC Deaths) by State and Race: April 2020

	(1)	(2)	(3)	(4)
State	All Races	White, Non-Hispanic	Black, Non-Hispanic	Hispanic
AL	87.17%	87.11%	79.41%	.
AK	99.43%	90.15%	.	.
AZ	84.10%	85.50%	75.00%	74.74%
AR	83.63%	80.63%	75.32%	.
CA	87.57%	87.54%	80.88%	84.08%
CO	86.36%	89.54%	99.38%	77.14%
CT	84.79%	84.82%	78.35%	70.66%
DE	93.39%	87.42%	92.35%	.
DC	83.38%	63.12%	80.75%	.
FL	88.40%	87.63%	79.67%	84.40%
GA	84.25%	85.56%	79.91%	85.02%
HI	54.29%	66.04%	.	.
ID	81.62%	81.77%	.	79.01%
IL	88.59%	94.28%	81.64%	78.80%
IN	86.63%	85.40%	79.92%	104.27%
IA	88.65%	87.20%	69.86%	.
KS	93.26%	90.21%	92.92%	84.21%
KY	85.22%	84.95%	73.59%	.
LA	82.00%	83.02%	80.21%	86.96%
ME	87.36%	87.36%	.	.
MD	89.97%	91.68%	84.13%	70.56%
MA	84.97%	86.33%	69.71%	72.61%
MI	91.66%	92.29%	82.80%	81.16%
MN	93.55%	91.58%	85.23%	.
MS	84.39%	85.88%	81.70%	.
MO	85.72%	85.71%	81.21%	56.57%
MT	84.03%	77.56%	.	.
NE	92.20%	81.71%	94.17%	.
NV	83.70%	81.78%	80.46%	76.50%
NH	90.40%	90.40%	.	.
NJ	84.44%	87.11%	78.07%	73.80%
NM	81.84%	80.43%	.	73.48%
NY	84.27%	90.23%	83.78%	79.64%
NC	85.37%	86.77%	75.69%	57.19%
ND	80.91%	72.82%	.	.
OH	91.20%	90.04%	78.10%	74.95%
OK	86.12%	85.89%	75.03%	86.18%
OR	86.09%	84.55%	.	77.97%
PA	89.51%	91.09%	84.68%	85.53%
RI	88.48%	79.16%	.	.
SC	86.28%	88.26%	86.18%	56.00%
SD	87.77%	88.87%	.	.
TN	85.14%	83.99%	83.38%	60.34%
TX	83.90%	84.04%	81.14%	80.16%
UT	87.23%	80.36%	.	71.61%
VT	83.82%	83.82%	.	.
VA	89.16%	89.33%	80.24%	94.72%
WA	85.23%	86.20%	82.42%	75.23%
WV	81.87%	79.65%	84.51%	.
WI	91.39%	90.46%	77.52%	85.89%
WY	95.11%	81.52%	.	.

Notes: Table compares counts of deaths in the Census Numident Analytic dataset to counts of deaths in the CDC NCHS vital statistics data by state and race. The data is from our sample from the Social Security Administration's Numerical Identification System (Numident) database covering population and deaths from January 2011 to April 2020. The sample covers a total of 26,680,000,000 individual-months, where 241,500,000 of them are in April 2020. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF), CDC NCHS Vital Statistics Data April 2020

2 Details of Statistical Analysis

Our estimates of excess mortality by race, nationally and state by state, rely on the following regression specification:

$$y_{casrmt} = \gamma_r + \kappa_m + \phi t + \sum_{\mu=1}^{4} \theta_\mu 1_{(t=2020, m=\mu)} + \\ + \sum_{\rho=1}^{7}\sum_{\mu=1}^{12} \tau_{\rho,\mu} 1_{(r=\rho)} 1_{(m=\mu)} + \sum_{\rho=1}^{7} \beta_\rho t + \sum_{\rho=1}^{7}\sum_{\mu=1}^{4} \lambda_{\rho,\mu} 1_{(r=\rho)} 1_{(t=2020, m=\mu)} + \\ + f(a, s, state) + \epsilon_{casrmt}$$

where y_{casrmt} is mortality (defined as the number of individuals who died during a month divided by the number individuals alive at the beginning of the month) in county c among individuals of age a, sex s, race r in month m of year t. γ_r denotes fixed effects for each race category that capture time-invariant differences in mortality across race categories in the full sample. κ_m denotes fixed effects for calendar months that capture seasonality in mortality. ϕ is the coefficient on a linear time trend t in years. $\sum_{\mu=1}^{4} \theta_\mu 1_{(t=2020, m=\mu)}$ are fixed effects for January, February, March, and April 2020 that capture any deviations from month-specific historical trends in 2020. The next set of terms interact month fixed effects, linear time trend, and fixed effects for January to April 2020 with fixed effects for each race category. Thus, $\sum_{\rho=1}^{7}\sum_{\mu=1}^{12} \tau_{\rho,\mu} 1_{(r=\rho)} 1_{(m=\mu)}$ are race-specific month fixed effects that capture race-specific deviations from aggregate seasonality in the data. $\sum_{\rho=1}^{7} \beta_\rho t$ allow for the linear time trend in years to be race-specific. And $\sum_{\rho=1}^{7}\sum_{\mu=1}^{4} \lambda_{\rho,\mu} 1_{(r=\rho)} 1_{(t=2020, m=\mu)}$ allow for the deviations from the historical trend in January, February, March, and April 2020 to also be race specific. These latter set of coefficients $\lambda_{\rho,\mu}$ are our primary coefficients of interest. For estimation, all fixed effects are normalized accordingly - we omit one fixed effect as the baseline in each group of fixed effects. For all interaction terms with race fixed effects we normalize the parameter values to non-Hispanic Whites, so that, for example, each $\lambda_{\rho,\mu}$ captures the race-specific deviation from the historical trend for each race relative to non-Hispanic Whites. These are the estimates directly reported in Panel (B) of **Exhibit 3** in the main text. Estimates in Panel (A) of **Exhibit 3** are constructed by adding up all regression parameters separately for each race category. The regression is estimated with weights that are equal to the count of individuals of age a, sex s, race r who were alive in county c in month m of year t. Standard errors are heteroskedasticity robust (Huber-White standard errors).

Function $f(a, s, state)$ contains parts of the regression model that are added when performing adjustments for age, sex, and/or state. Corresponding to the 5 columns of Table 2 in the manuscript, we estimate 5 specifications in total (in each specification fixed effects are normalized to an omitted baseline):

1. **Unadjusted:** $f(a, s, state) = 0$, meaning we do not adjust for any characteristics and we report baseline estimates of excess mortality by race. These are also the estimates underlying the graphs in **Exhibit 2.**

2. **Age Bin Adjustment:** $f(a, s, state)$ consists of fixed effects for each of eighteen 5-year age group (e.g., 11-14, 15-19, 20-24, etc.) and an interaction between an indicator for each 5-year age group and each month from January 2020 to April 2020

3. **Sex Adjustment:** $f(a, s, state)$ consists of a fixed effect for being male, and an interaction between this indicator and each month from January 2020 to April 2020.

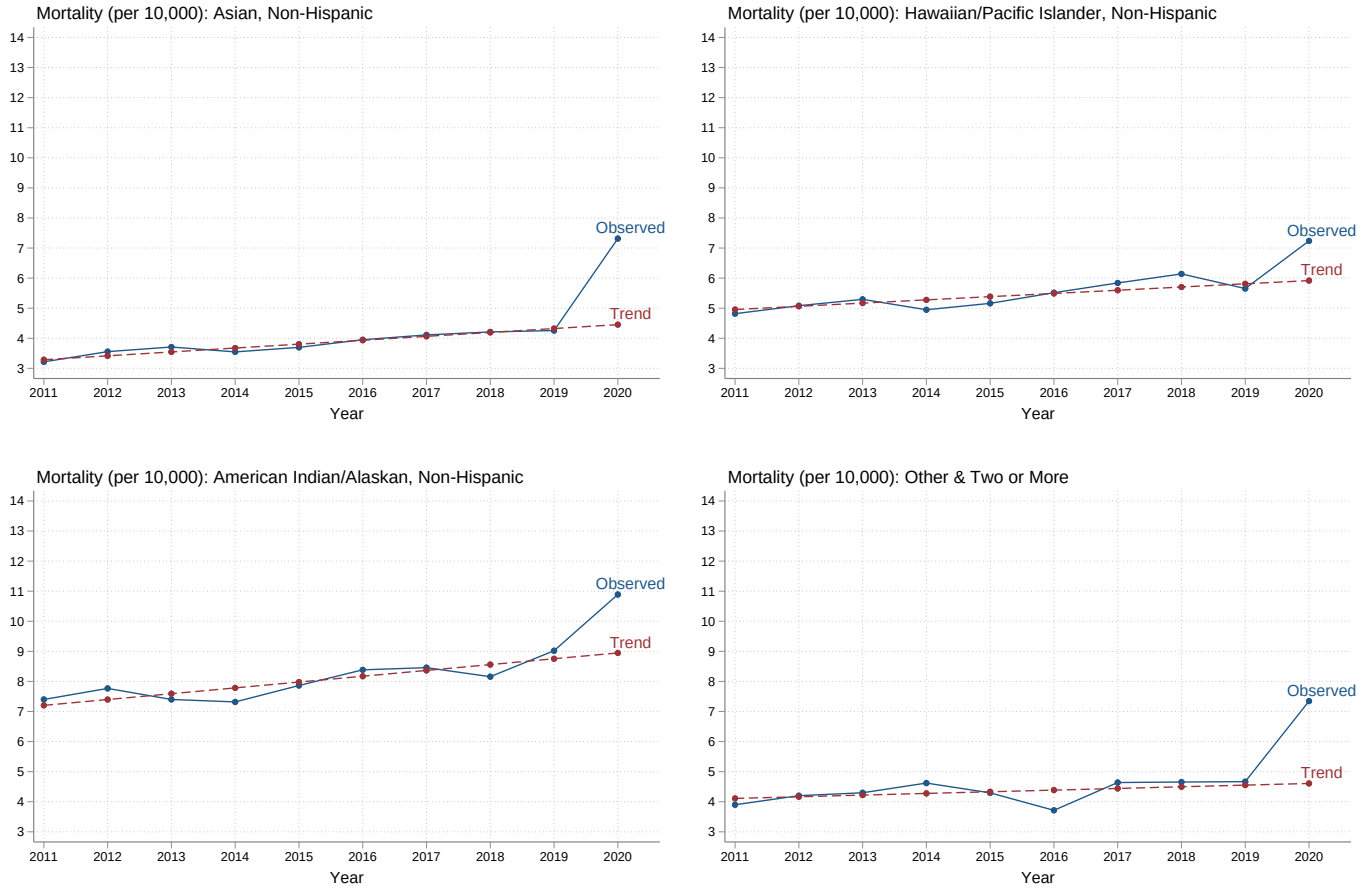
4. **State Adjustment:** $f(a, s, state)$ consists of a fixed effect for each state (and Washington, D.C.), and an interaction between an indicator for each state and each month from January 2020 to April 2020

5. **Age, Sex, and State Adjustment:** $f(a, s, state)$ consists of indicators for each 5-year age group, being male, and each state (including D.C.), as well as interactions between each of these indicators and each month from January 2020 to April 2020.

3 Appendix Exhibits

3.1 Observed and Predicted Mortality for Other Races

Figure S.1: Observed and Predicted April Mortality, 2011-2020



Notes: Figure reports a subpanel analogue of Exhibit 2 in the main manuscript for Hawaiian and Pcific Islander individuals, American Indian and Alaskan individuals and individuals who fall in the "other and two or more races" race category: observed and predicted (Trend) all-cause mortality (per 10,000) in the month of April for years 2011 to 2020. In April 2020. Race-specific trends are estimated as discussed in the Statistical Analysis and Appendix Section 2 with our unadjusted model on the full sample of 26,680,000,000 individual-months from January 2011 through April 2020. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

3.2 Estimates of Excess Mortality by Race

Table S.5: Excess Mortality (per 10,000): April 2020

Race	(1) Unadjusted	(2) Age Adjusted	(3) Sex Adjusted	(4) State Adjusted	(5) Age, Sex, & State Adjusted
White, Non-Hispanic	2.13	1.64	2.12	2.09	1.54
	(0.08)	(0.04)	(0.08)	(0.08)	(0.04)
Hispanic	2.74	3.75	2.74	3.09	4.26
	(0.13)	(0.10)	(0.13)	(0.13)	(0.10)
Black, Non-Hispanic	6.13	6.85	6.15	6.03	6.77
	(0.17)	(0.13)	(0.17)	(0.17)	(0.13)
Asian, Non-Hispanic	2.86	3.00	2.87	2.39	2.73
	(0.16)	(0.13)	(0.16)	(0.17)	(0.13)
American Indian & Alaskan, Non-Hispanic	1.95	2.81	1.95	3.12	3.91
	(0.24)	(0.24)	(0.24)	(0.25)	(0.24)
Hawaiian & Pacific Islander, Non-Hispanic	1.32	2.46	1.32	2.88	4.26
	(0.43)	(0.42)	(0.43)	(0.44)	(0.43)
Other & Two or More, Non-Hispanic	2.74	4.00	2.74	2.07	3.42
	(0.30)	(0.30)	(0.30)	(0.30)	(0.30)

Notes: This table shows estimates of excess all-cause mortalit by race in April 2020. Columns report the results of models with or without demographic adjustments as specified in column titles. Heteroskedasticity-robust standard errors are in parentheses. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis and Appendix Section 2. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Authors' calculations from Census Numident, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

Table S.6: Excess Mortality Relative to Non-Hispanic Whites (per 10,000): April 2020

Race	(1) Unadjusted	(2) Age Adjusted	(3) Sex Adjusted	(4) State Adjusted	(5) Age, Sex, & State Adjusted
White, Non-Hispanic	0	0	0	0	0
	–	–	–	–	–
Hispanic	0.62	2.11	0.62	1.00	2.72
	(0.15)	(0.10)	(0.15)	(0.16)	(0.11)
Black, Non-Hispanic	4.00	5.21	4.03	3.94	5.23
	(0.19)	(0.14)	(0.19)	(0.19)	(0.14)
Asian, Non-Hispanic	0.73	1.36	0.74	0.30	1.19
	(0.18)	(0.14)	(0.18)	(0.19)	(0.14)
American Indian & Alaskan, Non-Hispanic	-0.18	1.17	-0.17	1.04	2.37
	(0.25)	(0.24)	(0.25)	(0.26)	(0.24)
Hawaiian & Pacific Islander, Non-Hispanic	-0.81	0.83	-0.81	0.79	2.71
	(0.44)	(0.42)	(0.44)	(0.45)	(0.43)
Other & Two or More, Non-Hispanic	0.61	2.37	0.61	-0.02	1.88
	(0.31)	(0.30)	(0.31)	(0.32)	(0.30)

Notes: This table shows estimates of gaps in excess all-cause mortality for each race relative to Non-Hispanic White individuals in April 2020. Columns report the results of models with or without demographic adjustments as specified in column titles. Heteroskedasticity-robust standard errors are in parentheses. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis and Appendix Section 2. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Authors' calculations from Census Numident, 2010 Decennial Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

3.3 Distributions of Age and Sex by Race in Baseline Analytic Dataset

Table S.7: National and Race-Specific Distributions of Age and Sex

	All Races	Hispanic	White, Non-Hispanic	Black, Non-Hispanic	Asian, Non-Hispanic	American Indian & Alaskan, Non-Hispanic	Hawaiian & Pacific Islander, Non-Hispanic	Other & Two or more
Sex								
Female (%)	51.17	50.59	50.75	53.60	52.19	51.67	50.51	50.20
Age								
Average Age (Years)	45.14	38.55	47.40	41.80	44.73	41.58	39.68	37.06
Age Groups								
11 to 14 (%)	5.64	9.17	4.68	6.40	5.47	6.18	7.03	9.78
15 to 19 (%)	7.14	10.64	6.20	7.90	6.84	7.96	8.51	10.02
20 to 24 (%)	7.58	10.54	6.67	8.99	6.74	9.10	9.37	10.91
25 to 29 (%)	7.78	9.92	7.01	9.55	6.72	9.77	9.79	10.79
30 to 34 (%)	7.58	8.31	7.17	8.82	7.28	8.93	9.80	10.58
35 to 39 (%)	7.27	7.48	7.01	7.84	8.34	8.17	9.39	8.42
40 to 44 (%)	6.83	7.06	6.53	7.31	8.66	7.35	8.33	7.27
45 to 49 (%)	6.98	6.94	6.76	7.18	9.36	7.01	7.34	6.46
50 to 54 (%)	7.27	6.83	7.32	7.13	8.34	6.91	6.84	5.83
55 to 59 (%)	7.86	6.34	8.34	7.37	7.67	7.43	6.65	5.60
60 to 64 (%)	7.55	5.14	8.34	6.72	6.81	6.73	5.55	4.68
65 to 69 (%)	6.64	4.02	7.55	5.47	5.86	5.45	4.31	3.51
70 to 74 (%)	5.42	2.96	6.36	3.92	4.62	4.03	3.18	2.51
75 to 79 (%)	3.70	2.02	4.40	2.46	3.05	2.48	1.94	1.64
80 to 84 (%)	2.38	1.35	2.82	1.54	2.09	1.42	1.09	1.01
85 to 89 (%)	1.44	0.80	1.73	0.87	1.25	0.71	0.58	0.59
90 to 94 (%)	0.71	0.37	0.86	0.40	0.64	0.28	0.21	0.30
95 to 99 (%)	0.21	0.11	0.25	0.13	0.25	0.08	0.07	0.12

Notes: Table shows the distribution of age groups and sex overall and by race in April 2020 in our analytic dataset. When adjusting for sex and age, our regressions adjust to the national distributions of sex and age groups in this table. Age groups are 5 year bins starting at age 11. The first bin from ages 11 to 14 is only 4 years since our sample excludes individuals 10 and under. The distribution is based on our analytic sample from the Social Security Administration's Numerical Identification System (Numident) database. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF)

3.4 Estimates of Excess Mortality by State and Race for Selected Race Categories

Table S.8: Excess Mortality by State and Race in April 2020

State	White, Non-Hispanic		Black, Non-Hispanic		Hispanic	
	Excess Mortality	95% CI	Excess Mortality	95% CI	Excess Mortality	95% CI
NJ	8.55	(7.98, 9.13)	24.71	(22.89, 26.53)	20.53	(19.30, 21.76)
NY	7.15	(6.58, 7.72)	32.70	(30.71, 34.69)	27.18	(25.66, 28.70)
MA	5.94	(5.47, 6.41)	13.97	(12.30, 15.64)	7.19	(6.27, 8.11)
CT	5.53	(4.88, 6.17)	13.66	(11.70, 15.62)	8.70	(7.47, 9.92)
DC	4.11	(2.03, 6.19)	11.73	(9.45, 14.01)	5.04	(1.92, 8.16)
MI	2.83	(2.46, 3.19)	18.34	(15.99, 20.69)	4.92	(3.95, 5.88)
LA	2.71	(2.20, 3.23)	9.80	(8.81, 10.79)	3.88	(2.20, 5.56)
DE	2.22	(1.24, 3.21)	4.61	(2.77, 6.46)	1.89	(-0.01, 3.79)
CO	2.08	(1.66, 2.49)	7.43	(5.58, 9.28)	3.41	(2.68, 4.14)
MD	1.99	(1.53, 2.45)	5.93	(5.13, 6.73)	3.61	(2.75, 4.48)
PA	1.84	(1.56, 2.12)	10.58	(9.18, 11.98)	5.33	(4.51, 6.16)
IL	1.78	(1.46, 2.10)	8.29	(6.92, 9.66)	3.96	(3.35, 4.56)
RI	1.67	(0.79, 2.54)	3.40	(0.42, 6.38)	3.28	(1.49, 5.07)
IN	1.22	(0.86, 1.57)	5.91	(4.68, 7.15)	2.19	(1.26, 3.12)
GA	1.14	(0.80, 1.49)	2.69	(2.21, 3.18)	1.09	(0.51, 1.66)
VA	0.91	(0.59, 1.23)	2.25	(1.63, 2.87)	1.95	(1.30, 2.59)
FL	0.77	(0.53, 1.01)	2.05	(1.60, 2.50)	0.88	(0.48, 1.27)
OH	0.75	(0.50, 1.01)	2.56	(1.86, 3.26)	0.59	(-0.28, 1.45)
MS	0.69	(0.08, 1.31)	3.20	(2.30, 4.09)	-0.29	(-2.17, 1.58)
AZ	0.68	(0.24, 1.13)	0.71	(-0.51, 1.92)	0.87	(0.33, 1.40)
MN	0.68	(0.35, 1.00)	1.90	(0.87, 2.93)	0.45	(-0.42, 1.31)
MO	0.59	(0.23, 0.95)	4.13	(3.03, 5.23)	-0.73	(-1.72, 0.27)
VT	0.53	(-0.36, 1.42)	3.00	(-4.29, 10.28)	0.61	(-5.24, 6.46)
NC	0.49	(0.19, 0.80)	1.35	(0.81, 1.90)	0.19	(-0.38, 0.75)
NH	0.48	(-0.15, 1.11)	-0.24	(-3.85, 3.37)	-0.52	(-2.65, 1.61)
WA	0.45	(0.12, 0.78)	2.34	(1.09, 3.60)	1.12	(0.48, 1.75)
CA	0.34	(0.12, 0.57)	2.09	(1.44, 2.74)	0.92	(0.71, 1.14)
TX	0.34	(0.11, 0.56)	1.44	(1.00, 1.89)	0.25	(0.01, 0.48)
AL	0.29	(-0.16, 0.73)	2.00	(1.26, 2.74)	1.03	(-0.40, 2.46)
WI	0.27	(-0.04, 0.58)	4.60	(2.88, 6.33)	1.40	(0.46, 2.33)
WY	0.25	(-0.86, 1.35)	4.96	(-7.20, 17.11)	-0.48	(-3.36, 2.41)
ND	0.19	(-0.65, 1.03)	-0.42	(-2.26, 1.42)	0.47	(-2.76, 3.70)
SC	0.16	(-0.26, 0.57)	2.09	(1.35, 2.82)	-0.04	(-1.09, 1.02)
OR	0.15	(-0.24, 0.54)	1.28	(-1.04, 3.59)	0.27	(-0.45, 0.99)
KS	0.15	(-0.33, 0.63)	3.52	(1.52, 5.51)	0.42	(-0.60, 1.44)
KY	0.13	(-0.29, 0.55)	1.64	(0.39, 2.88)	0.92	(-0.58, 2.41)
IA	0.12	(-0.28, 0.53)	0.73	(-1.01, 2.48)	-0.08	(-1.25, 1.08)
TN	0.10	(-0.25, 0.45)	1.36	(0.59, 2.13)	-0.39	(-1.30, 0.52)
OK	0.08	(-0.41, 0.56)	1.46	(0.05, 2.87)	0.50	(-0.45, 1.45)
SD	0.07	(-0.71, 0.86)	-0.26	(-3.83, 3.32)	1.47	(-1.65, 4.59)
ME	0.06	(-0.55, 0.67)	-1.21	(-3.74, 1.33)	-2.65	(-5.75, 0.45)
WV	0.03	(-0.61, 0.67)	1.72	(-1.56, 5.00)	-0.17	(-3.77, 3.43)
UT	-0.02	(-0.55, 0.52)	-0.43	(-2.85, 1.98)	0.10	(-0.80, 1.00)
NV	-0.06	(-0.68, 0.56)	1.05	(-0.41, 2.51)	0.11	(-0.60, 0.82)
NE	-0.11	(-0.64, 0.43)	1.82	(-0.50, 4.14)	0.62	(-0.59, 1.83)
AR	-0.31	(-0.83, 0.21)	0.60	(-0.57, 1.76)	0.05	(-1.13, 1.22)
HI	-0.32	(-1.67, 1.02)	-0.79	(-3.97, 2.40)	-1.01	(-2.91, 0.89)
AK	-0.47	(-1.67, 0.73)	-1.13	(-4.54, 2.27)	0.15	(-2.44, 2.74)
MT	-0.58	(-1.31, 0.15)	-3.03	(-7.56, 1.49)	1.19	(-2.01, 4.40)
NM	-0.67	(-1.53, 0.18)	1.59	(-1.95, 5.13)	0.68	(-0.13, 1.48)
ID	-0.75	(-1.32, -0.18)	-1.02	(-4.75, 2.71)	0.55	(-0.81, 1.91)

Notes: Table report estimates of sex and age adjusted all-cause excess mortality in April 2020 among non-Hispanic Whites, non-Hispanic Blacks, and Hispanics for each US state. The details of the regression analysis used to construct the estimates are reported in the Statistical Analysis and Appendix Section 2. Each race- and state-specific estimate was adjusted to the national distribution of age and sex in April 2020. States are presented in descending order of non-Hispanic White excess mortality. The estimates are based on a sample of 26,680,000,000 individual-months from January 2011 through April 2020, where 241,500,000 individuals are alive at the start of April 2020. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-003. Source: Census Numident, 2010 Census, 2010 Census Modeled Race File, and 2010-2019 Master Address File - Auxiliary Reference File (MAF-ARF).

References

[1] United States Government Accountability Office, Social Security Death Data: Additional Action Needed to Address Data Errors and Federal Agency Access, https://www.gao.gov/products/GAO-14-46 (2013).

[2] Office of the Inspector General, Social Security Administration, Deceased Beneficiaries Who Had Different Dates of Death on the Social Security Administration's Numident and Payment Records, https://oig.ssa.gov/audits-and-investigations/audit-reports/A-09-17-50259. (2018a).

[3] Office of the Inspector General, Social Security Administration, Master Beneficiary Record Death Information That Did Not Appear on the Numident, https://oig.ssa.gov/audits-and-investigations/audit-reports/A-06-17-50190 (2018b).

[4] Office of the Inspector General, Social Security Administration, Title II Deceased Beneficiaries Who Do Not Have Death Information on the Numident, Audit Report (2012).

[5] Office of the Inspector General, Social Security Administration, Numberholders Age 112 or Older Who Did Not Have a Death Entry on the Numident, Audit Report (2015).

[6] D.A. Black, Y. Hsu, S.G. Sanders, L.S. Schofield, and L.J. Taylor, The Methuselah Effect: The Pernicious Impact of Unreported Deaths on Old Age Mortality Estimates, No. w23574. National Bureau of Economic Research (2017).

[7] The United States Census Bureau, 2010 Census Match Study Report, https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_247.pdf (2010).

[8] D.W. Brown, A. E. Kowalski, and I. Z. Lurie, Long-term impacts of childhood Medicaid expansions on outcomes in adulthood., *The Review of Economic Studies* 87, no. 2 (2020): 792-821.

[9] R. Chetty, M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler, The association between income and life expectancy in the United States, 2001-2014." *JAMA* 315, no. 16 (2016): 1750-1766.

[10] B. J. Martha, H. W. Hoynes, M. Rossin-Slater, and R. Walker, Is the social safety net a long-term investment? Large-scale evidence from the food stamps program, No. w26942. National Bureau of Economic Research (2020).

[11] M. Sarah, S. Altekruse, N. Johnson, and L. R. Wherry, Medicaid and mortality: new evidence from linked survey and administrative data, No. w26081. National Bureau of Economic Research (2019).

[12] W. Deborah and M. Lane, The person identification validation system (PVS): applying the Center for Administrative Records Research and Applications'(CARRA) record linkage software, No. 2014-01. Center for Economic Studies, US Census Bureau (2014)