Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States

Joshua R. Goldstein, Ph.D.
University of California, Berkeley, Department of Demography
josh.goldstein@berkeley.edu

Serge Atherwood, Ph.D.
University of Texas at San Antonio, Department of Demography
serge.atherwood@utsa.edu

May 19, 2020

Sources of support: NIH grants R01AG058940, 5P30AG012839, and P2CHD073964.

Abstract

Different estimation methods produce diverging accounts of racial/ethnic disparities in COVID-19 mortality in the United States. The CDC's decision to present the racial/ethnic distribution of COVID-19 deaths at the state level alongside the weighted racial/ethnic distribution of the counties within each state reporting those deaths—in effect, a geographic adjustment—makes it seem that Whites have the highest death rates. Age adjustment procedures used by others, including the New York City Department of Health and Mental Hygiene, lead to the opposite conclusion that Blacks and Hispanics are dying from COVID-19 at higher rates than Whites. In this paper, we use indirect standardization methods to adjust per capita death rates for both age and geography simultaneously, avoiding the one-sided adjustment procedures currently in use. Using CDC data, we find age-and-place-adjusted COVID-19 death rates are 80% higher for Blacks and over 50% higher for Hispanics, relative to Whites, on a national level. State-specific estimates show wide variation in mortality disparities. Comparison with nonepidemic mortality reveals potential roles for preexisting health disparities and differential rates of infection and care.

Introduction

Racial and ethnic disparities in COVID-19 mortality in the United States are attracting intense attention but estimates of their extent differ widely [e.g., 1–2]. With only limited data, the challenge for researchers is to distinguish racial/ethnic direct disparities in infection risk and lethality of COVID-19 from background compositional effects. In this paper, we make new use of demographic methods to address this challenge, providing age-and-place-standardized estimates of racial/ethnic COVID-19 mortality disparities both at the national level and for selected states.

Age is recognized as one of the strongest predictors of COVID-19 mortality [3–4]. An accurate assessment of racial/ethnic disparities in per capita death rates thus needs to account for differences in age structure across groups, as is done for other causes of death [e.g., 5–6]. Per capita rates that do not adjust for age will tend to find Whites with higher mortality rates, as a result of their older age structure relative to non-White groups.

Especially in the early stages of a viral outbreak, there may also be large spatial variation in the prevalence of infection. The early centers of coronavirus infection in the United States have been large cities such as New York, Chicago, Los Angeles, and Seattle—metropolitan regions in which African Americans, Hispanics, and Asians are disproportionately concentrated. Adjustments for geography, the CDC notes, "ensure that the population estimates and percentages of COVID-19 deaths represent comparable geographic areas, in order to provide information about whether certain racial and ethnic subgroups are experiencing a disproportionate burden of COVID-19 mortality" [1].

Place adjustment and age adjustment tend to work in opposite directions and either approach in isolation is potentially misleading. This divergence in results can be seen in Fig. 1 for New York City, the primary center of COVID-19 infection in the United States. Geographic

(A) CDC approach, adjusting only for place

Jurisdiction of Residence	Indicator	Non-Hispanic White	Non-Hispanic Black or African American	Non-Hispanic American Indian or Alaska Native[2]	Non-Hispanic Asian[3]	Hispanic or Latino	Other[4]
New York City	Distribution of COVID-19 deaths (%)	28.2	26.8	0.0	9.3	28.2	7.5
	Weighted distribution of population (%)	30.5	23.3	0.2	15.5	28.5	1.8

(B) NYHealth approach, adjusting only for age

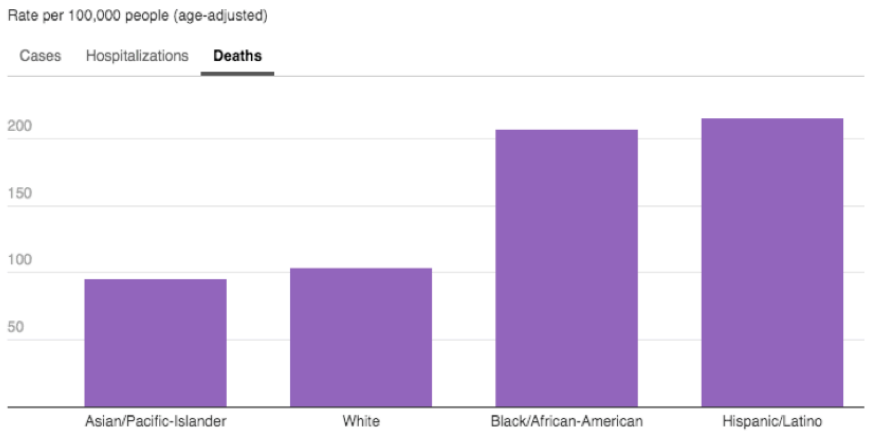

Figure 1. Conflicting indicators of COVID-19 mortality by race/ethnicity in New York City. Mortality adjusted for geographic distribution is represented by panel A, taken from the CDC webpage for provisional COVID-19 death counts by race/ethnicity on May 18, 2020 [1]. Mortality adjusted for age distribution is represented by panel B, taken from the New York City DOHMH COVID-19 webpage [2].

adjustment without age adjustment, as practiced by the Centers for Disease Control and Prevention (CDC) [1, 7],[1] is shown in panel A. This approach makes racial disparities seem small, with the share of deaths by race being only slightly higher for Blacks than Whites, relative to the geographically adjusted populations. Hispanics appear to have much lower mortality than Whites, representing fewer than 10% of the deaths but more than 15% of the population. On the other hand, age adjustment, as practiced by the New York City Department of Health and Mental Hygiene (DOHMH), is shown in panel B. This approach tells a completely different story, with both Blacks and Hispanics having mortality rates about twice as high as Whites.

It is possible to resolve such conflicting estimates by adjusting for both place and age at the same time. In this paper, we apply such a method to produce standardized estimates of COVID-19 mortality that are better suited for measuring the magnitude of racial disparities and their causes. Using this approach, we estimate COVID-19 mortality risks for non-Hispanic African Americans ("Blacks"), non-Hispanic Whites ("Whites"), non-Hispanic Asians ("Asians"), and Hispanics of any race ("Hispanics") nationwide and for selected states.

The procedures we use are designed for limited data and can be implemented widely for estimates at the state and national levels. The absence of detailed death counts that simultaneously report county, age, *and* race/ethnicity of deaths make it impossible to compare mortality disparities directly, but the reporting of the marginal distributions (the number of deaths by age, race, *or* county) allow the use of indirect standardization [8] to adjust per capita mortality rates for age and geography. These methods allow us to measure disparities across racial and ethnic groups, while making sure that we are effectively comparing people of the same age living in the same county.

Methods

The indirect standardization approach is to construct a counterfactual standard schedule of COVID-19 mortality $\widehat{M}_{ij}^{s}$ that varies by age i and county j but is the same for all racial/ethnic groups. A comparison of the observed counts of deaths by race/ethnicity to those predicted by

[1] The CDC's method is to reweight the exposed population to create "weighted population distributions... [to] ensure that the population estimates and percentages of COVID-19 deaths represent comparable geographic areas" [1]. Reweighting has a similar effect to the indirect standardization method we use (see Methods section).

the standard schedule reveals the severity of COVID-19 mortality separately from the effects of age and place.

Indirect standardization for both age and place is done by constructing a set of age-specific mortality rates for COVID-19 at the national level. These are then adjusted by level to produce the observed counts of death in each county, to produce the set of standard rates $\widehat{M}_{ij}^{s}$. Full details are given in the Appendix; R code for replication is available at https://github.com/josh-goldstein-git/race_covid.

Data

We use provisional COVID-19 death count data by age, race/ethnicity, and county available online from the CDC [1]. These data summarize the cumulative number of COVID-19 deaths entered into the National Vital Statistics System by the health departments of each state, which in turn compile death records from medical examiners and other authorities at the county level. Because of the coordination required to compile these counts and delays in reporting during the ongoing epidemic, the CDC numbers are explicitly preliminary. In compliance with federal confidentiality requirements, counts are suppressed for counties in which there have been fewer than 10 deaths from the epidemic. As of May 11, CDC reported provisional COVID-19 mortality counts for 322 counties, representing about two-thirds of the national population. CDC's counts include deaths for which the cause of death is given as COVID-19 as well as deaths for which the coronavirus is identified as the probable cause. Our method does not require counts to be complete, but it does depend on the reliability of the distributions by age, race/ethnicity, and place as given by the CDC.[2] We use U.S. Census Bureau estimates of population counts by county, race/ethnicity, and age. The most recent detailed population estimates available are from 2017.

[2] Potential reasons that the reported distributions by race, county, and age could vary from the actual distributions include differential timing of reports from hospitals, institutions such as prisons, old age homes, and nursing homes, and at-home deaths. We do not know of research on this issue.

Results

Unadjusted per capita death rates (relative to Whites) are shown at the top of Figure 2 under "Crude Rates". Taken at face value, this suggests African Americans across the U.S. as a whole have more than 50% greater risk while Hispanics and Asians have about 25% lower risk relative to Whites. However, as previously discussed, these disparities in unadjusted crude death rates are confounded by differences in age structure and the geographically localized nature of the early stages of the epidemic. Accounting for the county-level geography of COVID-19 mortality reduces disparities for Blacks to about half of the level of the per capita rates (56% to 24%), as can be seen in the "Place Standardized" portion of the figure. Place standardization also reduces the relative death rate for Hispanics slightly and Asians substantially. Age-standardization is shown in the next section, giving dramatically different results, with Blacks having a mortality risk relative to Whites of 2.7, Hispanics 1.9, and Asians 1.25. Disparities increase when the younger age structure of each of these groups, relative to Whites, is considered.

Finally, Figure 2 shows the age-and-place standardized relative rates in the lowest portion. We find that COVID-19 mortality of African Americans is 80% higher than Whites, higher than the crude rate and the place-standardized rate, but lower than the age-adjusted rate. For Hispanics, mortality is more than 50% higher than Whites. Asians are near parity.

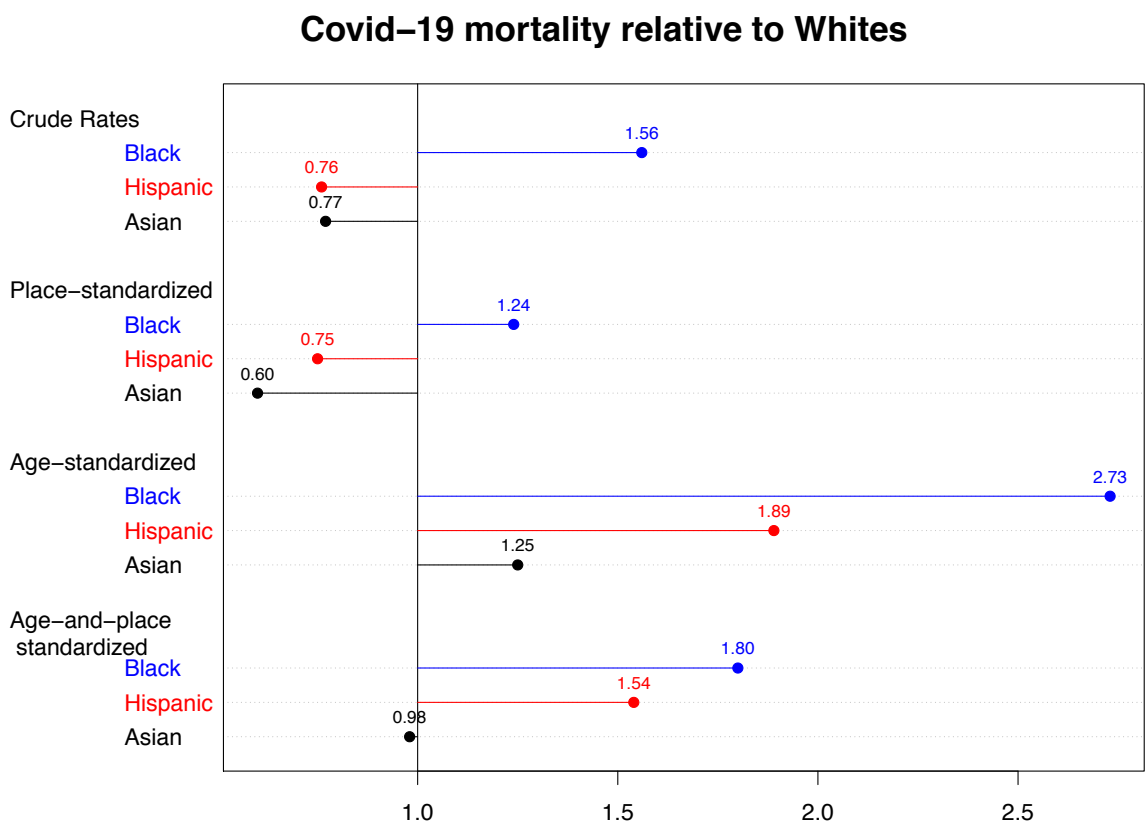

Figure 2. COVID-19 mortality risks by race/ethnicity, relative to Whites, by method. Crude rates are unadjusted per capita, based on the total population for all reported counties. Place-standardized ratios use indirect standardization to adjust for the geographic composition of each group, by county. Age-standardized ratios use indirect standardization to adjust for the age structure of each group. Age-and-place standardization adjusts for both factors simultaneously. Based on CDC mortality data current to May 13, 2020, including ascertained and probable COVID-19 deaths for counties with 10 or more deaths and 2017 U.S. Census Bureau estimates of county-level demography. Blacks, Asians, and Whites are non-Hispanic, and Hispanics are of any race.

It is possible to carry out similar calculations at the state level, aggregating counties within each state. We show the age-place adjusted relative rates of death from COVID-19 for Blacks and Hispanics in Figure 3 for each state with available data. The solid line in the figure at 1.0 shows the relative rate for Whites (the reference group). The dashed lines show the approximate national all-cause mortality level of Blacks and Hispanics, respectively, relative to Whites. For African Americans, the South (MS, LA, AL, SC) and Midwest (WI, IL, MI, IA, KS) have high excess mortality, but the Northeast (NY, NJ, MA) has smaller disparities. For

Hispanics, the West (NV, AZ, WA, CO, CA) has high excess mortality, while Florida is notably lower.

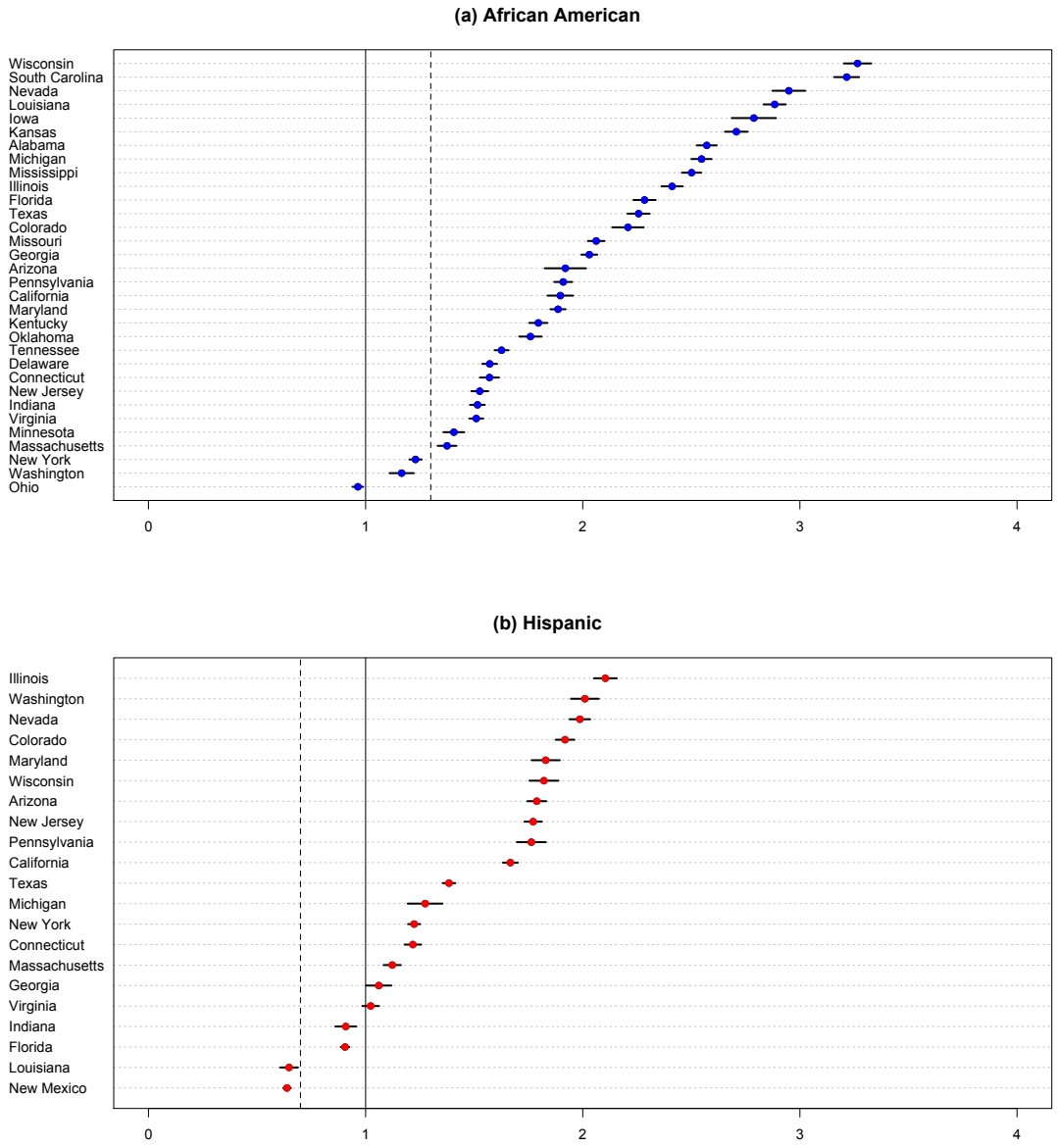

Figure 3. COVID-19 mortality relative to Whites, adjusted for age and place, by state. The solid vertical line at 1.0 is the COVID-19 relative rate for Whites. The dashed vertical lines are the approximate national all-cause mortality levels relative to Whites for (a) Blacks/African Americans and (b) Hispanics. States are included with sufficient numbers of cases for estimation. Uncertainty intervals represent approximate 95% confidence intervals. Estimates are based on CDC provisional data updated to May 13, 2020.

Discussion

An accurate accounting of racial and ethnic mortality disparities from COVID-19 requires adjusting for the age structure of sub-populations and also the geography of the epidemic. Indirect standardization for both age and place confirms the excess mortality of African Americans seen in the crude rates and reveals the excess mortality of Hispanics that is hidden by the crude rates.

We can compare the racial/ethnic risks from coronavirus-specific mortality with the risks from all-cause mortality in nonepidemic times in order to reveal the role of underlying health conditions inasmuch as these are responsible for previously identified patterns of mortality by race/ethnicity [e.g., 9–11]. U.S. life expectancy estimates imply, relative to Whites, about 30% higher all-cause mortality across all ages for Blacks and about 25% lower all-cause mortality across all ages for Hispanics. This means, for Blacks, that less than half of the disparity in COVID-19 mortality (i.e., 0.3 of the 0.8) is potentially attributable to "normal" mortality disparities. For Hispanics, who typically have lower mortality, the coronavirus epidemic appears to have had an even more dramatic effect, completely reversing the so-called "Hispanic paradox" [12–13].

The finding that mortality disparities during the COVID-19 epidemic by race/ethnicity are larger than in nonepidemic times suggests factors other than underlying health disparities are responsible. Principal among these is the differential risk of infection related to exposure at work, in transportation, and at home. Differential access to healthcare may also play a role.

Mortality adjustments that take into account the geographic distribution of the epidemic should be considered lower bounds on the true level of inequality. Standardizing for county of residence removes some of the relationship between race/ethnicity and place from analysis, although it leaves a role for within-county differences, such as neighborhood variation within counties. Standardization is typically used to remove the effects of "'extraneous' influences" [14, p. 24]. While it is clear that we should be comparing individuals of like ages, it is not so clear that place of residence is a purely "extraneous" influence, particularly given the long history of racial discrimination in housing [13]. Some of the geographic variation that place standardization accounts for seems appropriate, since the centers of the early stages of the epidemic were "destination cities" like New York and New Orleans which happened to have large shares of

minority residences. However, the magnitude of the outbreak in other large minority cities like Milwaukee, Chicago, and Detroit suggest that variation by place is not just a distortion to be controlled for, but is an integral part of the pattern of racial inequality.[3] As the epidemic spreads, differences due to the initial geographic seeding of the virus in the population may become less important, it may be less advisable to standardize by geography, or it may become necessary to present both the age-adjusted and the age-place-adjusted measures.

Finally, our methods reveal state-level variation in disparities which may lead to an a better understanding of the causes of racial/ethnic disparities in COVID-19 death rates. We find that the highest excess mortality for Blacks is the Deep South and the Upper Midwest, while disparities are smallest in the Northeast and West. Understanding the reasons for this pattern is an important topic for future research.

This paper offers a method for more refined measurement of racial and ethnic disparities in COVID-19 mortality using limited data. Indirect standardization is a powerful tool that can be widely adopted for the reporting of mortality that adjusts for both age and place. We believe this approach is superior to the one-sided adjustment procedures currently in use by state and local agencies as well as at the federal level at the CDC.

References

[1] Provisional death counts for coronavirus disease (COVID-19): Weekly updates by select demographic and geographic characteristics; Table 2a. https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/#Race_Hispanic. Accessed May 16, 2020.

[2] New York City Department of Health and Mental Hygiene. https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-05142020-1.pdf. Accessed May 16, 2020.

[3] Dowd JB, Andriano L, Rotondi V, Brazel DM, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *PNAS*. 2020; 117(18): 9696–9698. doi: 10.1073/pnas.2004911117. https://www.pnas.org/content/117/18/9696.

[3] As mentioned above, controlling for county-level differences does not entirely hide this history. Part of the inequality we are finding in mortality rates is due to spatial differences within counties. For example , Cook County includes both the poorest and richest parts of Chicago. Wayne County includes both the poorest parts of Detroit and some richer suburbs. However, further research is needed on the role that the choice of spatial unit has on measurements of racial disparities.

[4] Du R-H, Liang L-R, Yang C-Q, Wang W, Cao T-Z, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: A prospective cohort study. *Eur Respir J.* 2020; in press. https://erj.ersjournals.com/content/erj/early/2020/04/01/13993003.00524-2020.full.pdf.

[5] Dee DL, Bensyl DM, Gindler J, Truman BI, Allen BGA, D'Mello T, et al. Racial and ethnic disparities in hospitalizations and deaths associated with 2009 pandemic influenza A (H1N1) virus infections in the United States. *Annals of Epidemiology*. 2011; 21(8): 623–630. https://doi.org/10.1016/j.annepidem.2011.03.002.

[6] Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A HINI virus circulation: A modelling study. *The Lancet: Infectious Diseases*. 2012; 12(9): 687–695. https://doi.org/10.1016/S1473-3099(12)70121-4.

[7] Provisional death counts for coronavirus disease (COVID-19); Technical notes. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nvss/vsrr/covid19/tech_notes.htm. Accessed May 16, 2020.

[8] Wachter KW. *Essential Demographic Methods.* Cambridge, MA: Harvard University Press; 2014.

[9] McQuillan GM, Kruszon-Moran D, Kottiri BJ, Curtin LR, Lucas JW, Kington, RS. Racial and ethnic differences in the seroprevalence of 6 infectious diseases in the United States: Data from NHANES III, 1988–1994. *Am J Public Health*. 2004; 94(11): 1952–1958.

[10] Zajacova A, Dowd JB, Aiello AE. Socioeconomic and race/ethnic patterns in persistent infection burden among U.S. adults. *J Gerontol A Biol Sci Med Sci*. 2009; 64A(2): 272–279. doi:10.1093/gerona/gln012.

[11] Daw J. Contribution of four comorbid conditions to racial/ethnic disparities in mortality risk. *Am J Prev Med*. 2017; 52(1S1): S95–S102. http://dx.doi.org/10.1016/j.amepre.2016.07.036.

[12] Dominguez K, Penman-Aguilar A, Chang M-H, Moonesinghe R, Castellanos T, et al. Vital signs: Leading causes of death, prevalence of diseases and risk factors, and use of health services among Hispanics—2009–2013. *MMWR Morb Mortal Wkly Rep*. 2015 May 8; 64(17): 469–478.

[13] Massey DS, Denton NA. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA: Harvard University Press; 1998.

[14] Preston SH, Heuveline, P, Guillot, M. (2001) *Demography: measuring and modeling poopulation processes.* Oxford, UK: Blackwell Publishers.

APPENDIX

Standardization of COVID-19 mortality disparities by race/ethnicity

We estimated Standardized Mortality Ratios (SMR) for each race/ethnicity group (race) by aggregated spatial unit (e.g., a state or the country as a whole). Our estimates take into account, for each small area spatial unit (e.g., counties within a state), the age composition by race and the average severity of the epidemic across all races.

The SMR we calculate is

$$SMR_k = \frac{D_k(J)}{D_k^s(J)} = \frac{D_k(J)}{\sum_i \sum_{j \in J} M_{ij}^s \, N_{ijk}} \ ,$$

where

$D_k(J)$ are the observed counts of COVID-19 deaths of race k in aggregated spatial unit J,

M_{ij}^s is our standard schedule of COVID-19 per capita mortality rates by age group i and small-area spatial unit j, and

N_{ijk} is the observed census count of the population in age group i, in small area spatial unit j, of race k.

We report the SMRs of each group k, relative to that of Whites (W), denoting these $R_k(J)$, such that

$$R_k(J) = \frac{SMR_k(J)}{SMR_W(J)} \ ,$$

We construct the standard schedule M_{ij}^s by assuming independence between small area α_j severity and the age schedule of COVID-19 mortality M_i.

First, we calculate age-specific per capita mortality M_i for age group i for the United States as a whole,

$$M_i = \frac{D_i}{N_i} \ ,$$

where the D_i are based on provisional COVID-19 death counts from the Centers for Disease Control and Prevention and the N_i on U.S. Census Bureau population estimates.

Then we compute age standardized indices α_j of the intensity of epidemic mortality by small area j based on our observed counts D_j. Since, by independence, the per capita rates are the product of age schedule and small area intensity,

$$D_j = \sum_i (M_i \alpha_j) N_{ij} \,,$$

rearranging gives us an estimate of the age standardized intensity for each small area,

$$\hat{\alpha}_j = D_j / \sum_i M_i N_{ij} \,.$$

We then, assuming independence, construct our standardized schedule as

$$\widehat{M}_{ij}^{s} = \hat{\alpha}_j M_i \,.$$

Standardized Mortality Ratios by race give us an estimate of relative mortality, assuming that, within aggregated spatial units, mortality that is specific to age, place, and race is the product of independent terms for age M_i, small area spatial unit α_j, and race θ_k, such that $M_{ijk} = M_i \alpha_j \theta_k$. Under this assumption, observed deaths of race k in aggregated spatial unit J is given by

$$D_k(J) = \sum_i \sum_{j \in J} M_{ijk} N_{ijk} = \theta_k \sum_i \sum_{j \in J} M_i \alpha_j N_{ijk} \,.$$

Substituting this into our definition defined in the first equation of this appendix shows that $SMR_k(J) = \theta_k$.