medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Unraveling US National COVID-19 Racial/Ethnic Disparities using County Level Data Among 328 Million Americans

Short Title: US county COVID-19 health disparities

Daniel Li,[1,2] Sheila M. Gaynor,[1] Corbin Quick,[1] Jarvis T. Chen,[3] Briana J.K. Stephenson,[1] Brent A. Coull,[1,4] Xihong Lin[1,5,6*]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston MA.

[2]Ohio State University College of Medicine, Columbus OH.

[3]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston MA.

[4]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston MA.

[5]Department of Statistics, Harvard University, Cambridge MA.

[6]Broad Institute of MIT and Harvard, Cambridge, MA.

*Correspondence to: Xihong Lin, xlin@hsph.harvard.edu.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Racial and ethnic disparities in COVID-19 outcomes reflect the unequal burden experienced by vulnerable communities in the United States (US). Proposed explanations include socioeconomic factors that influence how people live, work, and play, and pre-existing comorbidities. It is important to assess the extent to which observed US COVID-19 racial and ethnic disparities can be explained by these factors. We study 9.8 million confirmed cases and 234,000 confirmed deaths from 2,990 US counties (3,142 total) that make up 99.8% of the total US population (327.6 out of 328.2 million people) through 11/8/20. We found national COVID-19 racial health disparities in US are partially explained by various social determinants of health and pre-existing comorbidities that have been previously proposed. However, significant unexplained racial and ethnic health disparities still persist at the US county level after adjusting for these variables. There is a pressing need to develop strategies to address not only the social determinants but also other factors, such as testing access, personal protection equipment access and exposures, as well as tailored intervention and resource allocation for vulnerable groups, in order to combat COVID-19 and reduce racial health disparities.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Main Text

Mounting evidence suggests that the COVID-19 pandemic has had a disproportionate racial and ethnic impact across communities in the US. For example, Black/African Americans and Hispanic/Latinos have been observed to have more COVID-19 cases, hospitalizations, and deaths than non-Hispanic Whites[1–4]. Proposed explanations draw on a social determinants of health framework, considering how structural patterns of privilege and disadvantage affect how communities live (e.g. in high density housing), work (e.g. in "essential" jobs), travel (e.g. on public transportation), and access healthcare, as well as people's underlying susceptibility to serious illness or death (e.g. prevalence of comorbidities such as hypertension or diabetes)[5–8]. However, little work has been done to systematically investigate the extent of racial health disparities in COVID-19 using large representative national samples.

There is a pressing need to examine how demographic (particularly racial and ethnic composition), socioeconomic, and health factors jointly contribute to disparities in COVID-19 outcomes. Several US studies have investigated racial and ethnic disparities. However these studies either looked at only a single health care institution or system[9,10], used national voluntary surveys that are subject to selection bias[11], or small scale ethnicity specific opinion surveys.[12] In the absence of large scale nationally representative individual-level data in the US to investigate this matter, analysis of ecologically aggregated county-level data across the US can provide valuable insights.

In this study, we investigated COVID-19-related health disparities using county-level data on demographic, socioeconomic, and health factors. We assessed the extent to which observed health disparities by race/ethnicity and by racial segregation in

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

county-level COVID-19 case and death rates persist after adjusting for county-level socioeconomic and health factors and additional unexplained state-to-state variation. We then calculated model-based county cumulative case and death rates to identify vulnerable US counties with the greatest total COVID-19 burden to date.

Brief Materials and Methods

We obtained demographic, socioeconomic and comorbidity data from a COVID-19 GitHub repository that drew from the US Department of Agriculture, Area Health Resources Files, County Health Rankings and Roadmaps, Centers for Disease Control and Prevention, and Kaiser News Health[13]. We obtained COVID-19 county cases and deaths from 1/22/20-11/8/20 from USA Facts[14] and additional demographic data from the US Census Bureau[15]. Numbers of COVID-19 tests by state were obtained from The COVID Tracking Project[16]. Daily percentage time home mobility metrics from 3/8/20-9/30/20 were obtained from SafeGraph[17]. Additional details about data sources are available in the supplement materials.

We used a Poisson mixed model to model county cumulative case and death counts. We included fixed effects for each state to account for state-to-state variation not explained by variables in the model (such as state testing rates)**,** a random effect for each county to account for overdispersion, and an offset term for log county population (2019 US Census estimates). Univariate and multivariate regression with county-level race/ethnicity percentages were performed. We used the fitted models to calculate estimated county cumulative case and death rates to identify counties with elevated rates across the study period. **Table S2** lists the covariates used in modeling.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We performed several additional analyses to gain further insight into the findings and investigate sensitivity of the findings to model assumptions. We examined correlations between county race percentage and socioeconomic and health variables to understand differences in univariate and multiple regression results. We performed a sensitivity analysis investigating potential non-linear associations between county race/ethnicity percentage and cases and deaths by categorizing county race/ethnicity percentage into quartiles. We assessed adding county-level diabetes and kidney disease rates to improve modeling. To gain additional insight into our results, we investigated how the mobility measure using SafeGraph county resident percentage time at home[17] was associated with demographic, socioeconomic, and health factors. We studied the effects of monthly average percentage time at home between March and August on cumulative case and death rates. Lastly, we performed exploratory Case Fatality Rate (CFR) and Infection Fatality Rate (IFR) analyses to characterize associations among infected cases. Additional details about statistical analyses are in the supplement.

Results

As of 11/8/20, there were 9,827,966 total confirmed COVID-19 cases and 234,538 total COVID-19 deaths among all US counties. The primary analyses included 2,990 US counties (of 3,142 total US counties) which had complete information for all covariates (**Fig. S1**). Per 2019 US Census estimates, these counties comprised 327,553,510 Americans (99.8% of 328,239,523 total). There were 9,808,959 confirmed COVID-19 cases and 234,357 deaths among these 2,990 counties. Case and death

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

rates in the subsequent discussions refer to cumulative case and death rates. **Fig. 1** shows the case (**A**) and death (**B**) rates per 100,000 people for all counties. **Fig.1** shows heatmaps for race/ethnicity percentage (**C**) and residential racial segregation (0 complete integration to 100 complete segregation, **D**). Counties in the Southeast tended to have a larger proportion of Black/African American individuals, and counties in the Southwest tended to have a larger proportion of Hispanic/Latino individuals. Counties with a large percentage of American Indian individuals, such as the Navajo Nation in Arizona, New Mexico, and Utah, also tended to have a greater degree of residential racial White/non-White segregation. **Fig. S2** shows heatmaps for all other covariates.

Relative risks (RRs) for the multiplicative increase in county COVID-19 cumulative case and death rates were calculated for a one standard deviation increase of a county-level variable. Our univariate regressions showed county-level Black/African American, Hispanic/Latino, Native American percentages, Asian percentages, and Native Hawaiian/Pacific Islander percentages were highly significantly and strongly associated with higher case and/or death rates (**Table 1**). These univariate results are consistent with the literature.[18–22]

Adjusting for all non-race demographic, socioeconomic, and county-level disease rates, higher county Black/African American percentage was still associated with higher county COVID-19 case (RR: 1.05, 95% CI: 1.02-1.08) and death rates (RR: 1.13, 95% CI: 1.07-1.19); higher Hispanic/Latino percentage with higher county case rates (RR: 1.09, 95% CI: 1.06-1.12) and death rates (RR: 1.08, 95% CI: 1.02-1.14); higher American Indian/Native Alaskan percentage with higher county death rates (RR: 1.10, 95% CI: 1.05-1.15); and higher Native Hawaiian/Pacific Islander percentage with higher

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

county case rates (RR: 1.03, 95% CI: 1.01-1.05); (**Fig. 2a**). The associations of county American Indian, Asian, and Native Hawaiian percentage and case/death rates that became insignificant did so after controlling for other race/ethnicity percentages or percentage with no health insurance (**Table S3-S5**). Greater White/non-White segregation was associated with higher county case (RR: 1.05, 95% CI: 1.03-1.07) and death rates (RR: 1.07, 95% CI: 1.03-1.10).

After adjusting for demographic variables and prevalence of comorbid conditions, county average household size, percentage in poverty, and percentage of individuals with no high school diploma were associated with higher county case and/or death rates. Adjusting for demographic and socioeconomic variables, the percentages of individuals with heart failure, hypertension, and stroke were associated with higher county case and/or death rates, and county-level asthma and chronic obstructive pulmonary disease (COPD) rate were associated with lower county case and/or death rates.

After adjusting for demographic and socioeconomic variables and county-level comorbidity disease rates, there was a stronger association between American Indian/Native Alaskan percentage death rates among counties in the top quartile of White/non-White segregation (RR: 1.13, 95% CI: 1.07-1.20) compared to counties with less White/non-White segregation (RR: 1.06, 95% CI: 1.00-1.12). The difference in death rate associations was statistically significant (effect modification: 1.07, 95% CI: 1.00-1.14) (**Fig. 2B**).

Estimated cumulative case and death rates for each US county were calculated from Poisson mixed models (**Figs. 3A and 3B**). Of the counties presented in **Figs. 3C**

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and 3D, during the study period, those comprising the Navajo Nation, Miami, and New York City had the highest estimated case rates. Counties comprising New York City, the Navajo Nation, and Detroit had the largest estimated death rates. Observed and model-based rates had good agreement, suggesting adequacy of model fit (**Fig. S3**). Results for every US county with full covariates are provided online.

We calculated the correlation of demographic, socio-economic and existing medical conditions to better understand the racial disparity findings and differences between multiple regression and univariate regression results. We found different county race/ethnicity percentages were correlated with different socioeconomic variables and disease rates (**Fig. S4**). County Black/African American percentage was positively associated with county poverty (R^2 = 0.39; 95% CI: 0.36, 0.42), no high school diploma (R^2 = 0.33; 95% CI: 0.29, 0.36), and hypertension (R^2 = 0.56; 95% CI: 0.54, 0.58). County Hispanic/Latino percentage was positively associated with county average household size (R^2 = 0.42; 95% CI: 0.39, 0.42) and county percentage of individuals with no health insurance (R^2 = 0.45; 95% CI: 0.42, 0.48). County American Indian/Native Alaskan percentage was positively associated with county percentage of individuals with no health insurance (R^2 = 0.31; 95% CI: 0.28, 0.34). County Asian percentage was negatively associated with county smoking (R^2 = 0-0.40; 95% CI: -0.43, 0.37) and COPD (R^2 = -0.36; 95% CI: -0.39, -0.33).

We investigated potential non-linear associations between county race/ethnicity percentage and cases/deaths. Controlling for demographic and socioeconomic variables and county disease rates, we found counties with Black/African American percentages in the top quartile had greater case and death rates than those in the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

bottom quartile (**Table S6**). Similarly, counties with Hispanic/Latino percentages in the top quartile had greater case and death rates than those in the bottom quartile.

To gain additional insight into observed health disparities, we used mobility data to study whether time spent at home due to stay-at-home orders and social distancing were associated with demographic and socioeconomic variables. Controlling for all other county level variables, adjusted effects for the change in county percentage time spent at home were calculated for a one standard deviation increase in a county-level variable (**Fig. S5**). Counties with greater percentage of residents ages 20-29 years (-1.39; 95% CI: -1.68, -1.10) and greater no health insurance (-0.63, 95% CI: -1.11, -0.15) were associated with spending less time at home. Higher county population density, metro > 1 million people, metro/near metro 20,000 to 1 million people (relative to nonmetro <20,000 people), and higher average household size were associated with county residents spending more time at home. Including average percentage time at home as a covariate also did not change adjusted cumulative case and death relative risks (**Fig. S6**).

We performed exploratory Case Fatality Rate (CFR) analyses and Infection Fatality Rate (IFR) analyses (**Fig. S7**). Assuming the ascertainment rates of reported cases vary by state and counties modeled using fixed and random effects respectively, the CFR and IFR regression analyses produced identical results (see **Methods** for further discussion). These results had similar directions to the primary death rate results but there were fewer significant associations. These analyses are likely subject to bias due to several factors, such as differential underestimation of the total number of cases (including asymptomatic and mildly symptomatic cases) by race and ethnicity[23,24],

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

selection bias of subjects who have been tested (e.g. symptomatic subjects and vulnerable subjects were more likely to be tested), insufficient testing capacity in many areas, and variable testing rates between and within states (**Methods**).

Discussion

Our ecological study used US county-level data to investigate US health disparities by studying the joint effects of demographic, socioeconomic, and health variables on COVID-19 cumulative case and death rates as of 11/8/20 using representative national data. We found racial and ethnic health disparities among Black/African American, Hispanic/Latino, American Indian/Native Alaskan, and Native Hawaiian/Pacific Islander communities persisted after adjusting for county-level socioeconomic and prevalence of comorbid conditions. We also observed that increased racial residential segregation increased COVID-19 case and death rates, with different impacts across racial and ethnic communities. Lastly, our estimated COVID-19 case and death rates account for potential instability in observed rates from counties with small populations or few confirmed cases/deaths and can assist in identifying counties with the greatest total COVID-19 burden.

Univariate associations between race/ethnic composition and COVID-19 outcomes were considerably stronger prior to adjustment for socioeconomic factors and comorbidities, indicating socioeconomic factors and comorbidities may partially explain observed racial and ethnic disparities. Several unstudied factors may additionally contribute residual racial and ethnic disparities, including differences in neighborhood testing rates, percentage of healthcare workers, percentage of essential workers,

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

exposure to infected individuals within households and in communities, Personal Protection Equipment (PPE) access, use of public transportation, access, quality, and utilization rates of available healthcare facilities/resources, access to living resources (such as a lack of access to clean water in many households of American Indian/Alaskan Native communities), and health literacy. Since many of these measures were either not available or quantifiable at the county-level, we were unable to control for them in analyses.

Counties with increased racial residential segregation experienced increased COVID-19 case and death rates. While Black and American Indian communities have historically been segregated into counties, Hispanics communities are more prone to micro-segregation within counties[25–27]. Counties and states may appear to be less segregated on a larger level, but Hispanic/non-Hispanic communities may still remain racially segregated at the neighborhood level[28,29]. In addition, after adjusting for socioeconomic variables and comorbidities, counties with more racial segregation had stronger associations between Native American percentage and COVID-19 case and death rates. This suggests stronger COVID-19 disparities were experienced on Native American territories and reservations, which have more racial segregation, compared to Native American living in other communities.

Our county-level results on racial and ethnic disparities also reinforce and expand findings reported from existing individual-level studies. Single institution studies in the US have also found that Black COVID-19 patients were more likely to be hospitalized, enter the intensive care unit, and die[9,30]. United Kingdom Biobank and electronic health record studies looking at individual-level data have also found that

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Black and Asian individuals have an increased risk of COVID-19 hospitalization and deaths after adjusting for covariates[31–33].

Our county-level results in the other domains are consistent with those from several smaller scale individual-level studies. We found counties with a greater proportion of individuals with ages 60+ years tended to have increased death rates. Individual-level studies also reported that older patients were more likely to develop severe COVID-19 symptoms and have greater mortality rates[34]. We found county average household size was associated with increased case and death rates. Household size is known to affect COVID-19 contact and transmission rates[35]. We found county level rates of heart failure, hypertension, and stroke were associated with case and/or death rates. These pre-existing health conditions are important biological and clinical risk factors for COVID-19 disease severity and mortality[34,36]. We found county COPD percentage was negatively associated with COVID-19 case and death rates. Some studies have observed a lower than expected prevalence of COPD in COVID-19 patients, but this association is still being investigated as other studies have reported COPD is a significant risk factor for COVID-19 infection at the individual level[37–39]. Lastly, we found counties with more individuals ages 20-29 tended to have residents spend less time at home. As early as early February 2020, there have been discussions about young adults being more likely to go out and socialize despite social distancing guidelines[40]; this issue received more media attention in June and July.

Our SafeGraph mobility metric analyses suggested residents of counties with a higher percentage of people with no high school diplomas or health insurance tended to spend less time at home. These communities may have a higher percentage of

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

essential workers who are unable to work from home, may be more likely to take public transportation, and may be more susceptible to contracting COVID-19[41]. Such areas may require additional attention and interventions.

We observed different county-level socioeconomic associations across various races/ethnicities. This emphasizes that demographic, socioeconomic and immigration complexities faced by various populations are likely to differ substantially between vulnerable communities, especially those with high proportions of under-represented minorities[42]. To address this challenge, public health interventions, medical care services, and outreach efforts need to be tailored to the unique challenges and needs of each community.

Because we control for fixed state effects, adjusted race/ethnicity composition associations adjust for unmeasured state-specific factors (e.g. differences in testing capacity or response procedures). Within states, counties with larger non-white populations may tend to be more socioeconomically underserved, and the numbers of reported cases are likely to underestimate the total number of infected cases[23,24]. If under-reporting is driven by race/ethnicity, our reported adjusted associations likely understate the true extent of racial/ethnic disparities. To address racial health disparity on COVID-19 in US, county-level race/ethnicity specific case and death count data are needed for research purposes.

Our findings are based on ecological associations at the county level, and analysis at this level is subject to several limitations (**Methods**). Associations observed at an aggregated level may be in the same direction, different direction, or not exist at the individual level[43]. As with all observational studies, associational findings do not

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

imply causality. It is of interest to in the future conduct studies on COVID-19 disparities using individual-level data with additional information on household and community exposures to COVID-19 cases, occupation and work conditions, housing conditions, public transportation usage, basic living resources, and COVID-19 treatments. Despite these limitations, our US county-level ecological study identified elevated risks of COVID-19 cases and deaths in areas with substantial non-White populations after adjusting for socioeconomic and disease prevalences.

Multi-faceted efforts are needed to combat the pandemic by addressing these COVID-19 health disparity issues. Increased resources, such as testing priority and accessible points of care, should be allocated to counties with more racial/ethnic minority populations or residential racial segregation, as well as those counties with more crowded housing, more elderly residents, less education infrastructure, greater prevalences of hypertension, and less living resources, such as a lack of clean water. Intervention measures can include policies requiring face coverings, guaranteeing workers can take paid sick leave, providing personal protective equipment to essential workers, and ensuring prioritized and robust testing, tracing, and isolation infrastructure. Outreach efforts can include transportation assistance, social and community support, and increased accessibility and affordability of health care.

Data and code availability

All materials and code for analysis are available on https://github.com/lin-lab.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgments: Resource support was provided by the Cannon cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. This work was funded by NIH grant T32-CA933739 (DL), HD092580 (JTC and BAC), and R35-CA203654-04 (XL).

Author Contributions: DL, SMG, CQ, and XL contributed to the conception and design of the work and to the acquisition and analysis of the data. All authors contributed to interpreting the data. DL drafted the work. All authors substantively revised the work, approved the submitted work, and agreed to be personally accountable for their own contributions.

Competing Interest Declaration: The authors declare no competing interests.

Additional Information: Correspondence and requests for materials should be addressed to Xihong Lin.

References

1. Garg, S. *et al*. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 458–464 (2020).
2. Laurencin, C. T. & McClinton, A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. *J. Racial and Ethnic Health Disparities* (2020) doi:10.1007/s40615-020-00756-0.
3. Haynes, N., Cooper, L. A. & Albert, M. A. At the Heart of the Matter: Unmasking and Addressing COVID-19's Toll on Diverse Populations. *Circulation* (2020) doi:10.1161/CIRCULATIONAHA.120.048126.
4. Bassett, M. T., Chen, J. T. & Krieger, N. The unequal toll of COVID-19 mortality by age in the United States: Quantifying racial/ethnic disparities. *Harvard Center for Population and Development Studies Working Paper Series* **19**, (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. Bibbins-Domingo, K. This Time Must Be Different: Disparities During the COVID-19 Pandemic. *Ann Intern Med* (2020) doi:10.7326/M20-2247.

6. Dorn, A. van, Cooney, R. E. & Sabin, M. L. COVID-19 exacerbating inequalities in the US. *The Lancet* **395**, 1243–1244 (2020).

7. Yancy, C. W. COVID-19 and African Americans. *JAMA* (2020) doi:https://doi.org/10.1001/jama.2020.6548.

8. Havranek, E. P. *et al.* Social Determinants of Risk and Outcomes for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation* **132**, 873–898 (2015).

9. Price-Haywood, E. G., Burton, J., Fort, D. & Seoane, L. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N Engl J Med* NEJMsa2011686 (2020) doi:10.1056/NEJMsa2011686.

10. Azar, K. M. J. *et al.* Disparities In Outcomes Among COVID-19 Patients In A Large Health Care System In California: Study estimates the COVID-19 infection fatality rate at the US county level. *Health Affairs* **39**, 1253–1262 (2020).

11. Allen, W. E. *et al.* Population-scale Longitudinal Mapping of COVID-19 Symptoms, Behavior, and Testing Identifies Contributors to Continued Disease Spread in the United States. *Nature Human Behavior* (2020) doi:10.1101/2020.06.09.20126813.

12. *2020 COVID-19 Poll of African American Voters*. https://www.africanamericanresearch.us/survey-results.

13. Altieri, N. *et al.* Curating a COVID-19 data repository and forecasting county-level death counts in the United States. *arXiv* (2020).

14. About USA Facts. *USA Facts* https://usafacts.org/about-usafacts/.

15. County Population by Characteristics: 2010-2018. *United States Census Bureau* https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html (2020).

16. Data API. *The COVID Tracking Project* https://covidtracking.com/api.

17. Shelter in Place Index: The Impact of Coronavirus on Human Movement. *SafeGraph* https://www.safegraph.com/dashboard/covid19-shelter-in-place?s=US&d=05-20-2020&t=counties&m=index (2020).

18. Gross, C. P. *et al.* Racial and Ethnic Disparities in Population-Level Covid-19 Mortality. *J GEN INTERN MED* **35**, 3097–3099 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

19. Block, R., Berg, A., Lennon, R. P., Miller, E. L. & Nunez-Smith, M. African American Adherence to COVID-19 Public Health Recommendations. *HLRP: Health Literacy Research and Practice* **4**, e166–e170 (2020).
20. Rodriguez-Diaz, C. E. *et al.* Risk for COVID-19 infection and death among Latinos in the United States: examining heterogeneity in transmission dynamics. *Annals of Epidemiology* (2020) doi:10.1016/j.annepidem.2020.07.007.
21. Moore, J. T. *et al.* Disparities in Incidence of COVID-19 Among Underrepresented Racial/Ethnic Groups in Counties Identified as Hotspots During June 5–18, 2020 — 22 States, February–June 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1122–1126 (2020).
22. Chang, R. C., Penaia, C. & Thomas, K. Count Native Hawaiian And Pacific Islanders In COVID-19 Data—It's An OMB Mandate. *Health Affairs* (2020) doi:10.1377/hblog20200825.671245.
23. Hao, X. *et al.* Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* (2020) doi:10.1038/s41586-020-2554-8.
24. Havers, F. P. *et al.* Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern Med* (2020) doi:10.1001/jamainternmed.2020.4130.
25. Bailey, Z. D. *et al.* Structural racism and health inequities in the USA: evidence and interventions. *The Lancet* **389**, 1453–1463 (2017).
26. Williams, D. R. & Collins, C. Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. *Public Health Reports* **116**, 404–16 (2001).
27. White, K., Haas, J. S. & Williams, D. R. Elucidating the Role of Place in Health Care Disparities: The Example of Racial/Ethnic Residential Segregation. *Health Serv Res* **47**, 1278–1299 (2012).
28. Lichter, D. T., Parisi, D. & Taquino, M. C. Emerging Patterns of Hispanic Residential Segregation: Lessons from Rural and Small-Town America: Emerging Patterns of Hispanic Residential Segregation. *Rural Sociology* **81**, 483–518 (2016).
29. Documét, P. I. & Sharma, R. K. Latinos' Health Care Access: Financial and Cultural Barriers. *Journal of Immigrant Health* **6**, 5–13 (2004).
30. Gu, T. *et al.* COVID-19 outcomes, risk factors and associations by race: a comprehensive analysis using electronic health records data in Michigan Medicine. *medRxiv* (2020) doi:https://doi.org/10.1101/2020.06.16.20133140.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

31. The OpenSAFELY Collaborative *et al*. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *medRxiv* (2020) doi:10.1101/2020.05.06.20092999.

32. Patel, A. P., Paranjpe, M. D., Kathiresan, N. P., Rivas, M. A. & Khera, A. V. Race, Socioeconomic Deprivation, and Hospitalization for COVID-19 in English participants of a National Biobank. *medRxiv* (2020) doi:10.1101/2020.04.27.20082107.

33. Raisi-Estabragh, Z. *et al*. Greater risk of severe COVID-19 in Black, Asian and Minority Ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25(OH)-vitamin D status: study of 1326 cases from the UK Biobank. *Journal of Public Health* (2020) doi:10.1093/pubmed/fdaa095.

34. Coronavirus Disease 2019 (COVID-19): People Who Are at Higher Risk for Severe Illness. *Centers for Disease Control and Prevention* https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html (2020).

35. Jing, Q.-L. *et al*. Household Secondary Attack Rate of COVID-19 and Associated Determinants. *medRxiv* (2020) doi:10.1101/2020.04.11.20056010.

36. Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239 (2020).

37. Halpin, D. M. G., Faner, R., Sibila, O., Badia, J. R. & Agusti, A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection? *The Lancet Respiratory Medicine* **8**, 436–438 (2020).

38. Onder, G., Rezza, G. & Brusaferro, S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA* (2020) doi:10.1001/jama.2020.4683.

39. Lippi, G. & Henry, B. M. Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). *Respiratory Medicine* **167**, 105941 (2020).

40. Lewis, M. *et al*. COVID-19 Outbreak Among College Students After a Spring Break Trip to Mexico — Austin, Texas, March 26–April 5, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 830–835 (2020).

41. Del Real, J. A. In an immigrant community battling coronavirus, ‘essential’ means ‘vulnerable’. *The Washington Post* 5/9/2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

42. Chowkwanyun, M. & Reed, A. L. Racial Health Disparities and Covid-19 — Caution and Context. *N Engl J Med* NEJMp2012910 (2020) doi:10.1056/NEJMp2012910.

43. Wakefield, J. & Salway, R. A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **164**, 119–137 (2001).

44. Area Health Resources Files. *Health Resources & Services Administration* https://data.hrsa.gov/topics/health-workforce/ahrf (2019).

45. About Us. *County Health Rankigs & Roadmaps* https://www.countyhealthrankings.org/about-us.

46. Coronavirus Locations: COVID-19 Map by County and State. *USA Facts* https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.

47. Residential segregation - black/white*. *County Health Rankigs & Roadmaps* https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors/family-social-support/residential-segregation-blackwhite.

48. Rural-Urban Continuum Codes. *United States Department of Agriculture* https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/.

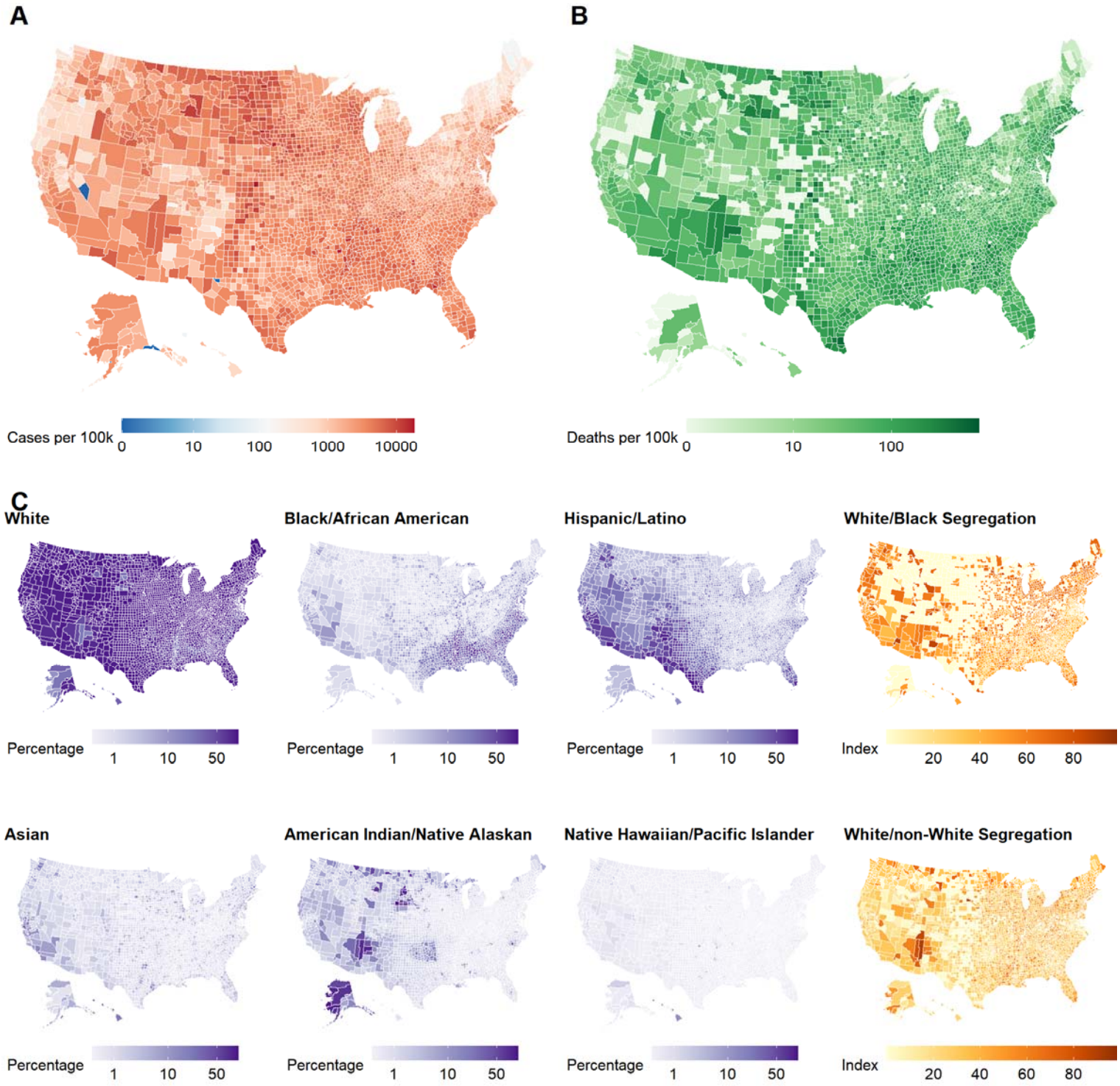
49. Freedman, D. A. Ecological Inference and the Ecological Fallacy. *International Encyclopedia of the Social & Behavioral Sciences* (1999).

50. Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 — Studies Needed. *N Engl J Med* **382**, 1194–1196 (2020).

51. Chin, T. *et al*. U.S. county-level characteristics to inform equitable COVID-19 response. *medRxiv* (2020) doi:10.1101/2020.04.08.20058248.

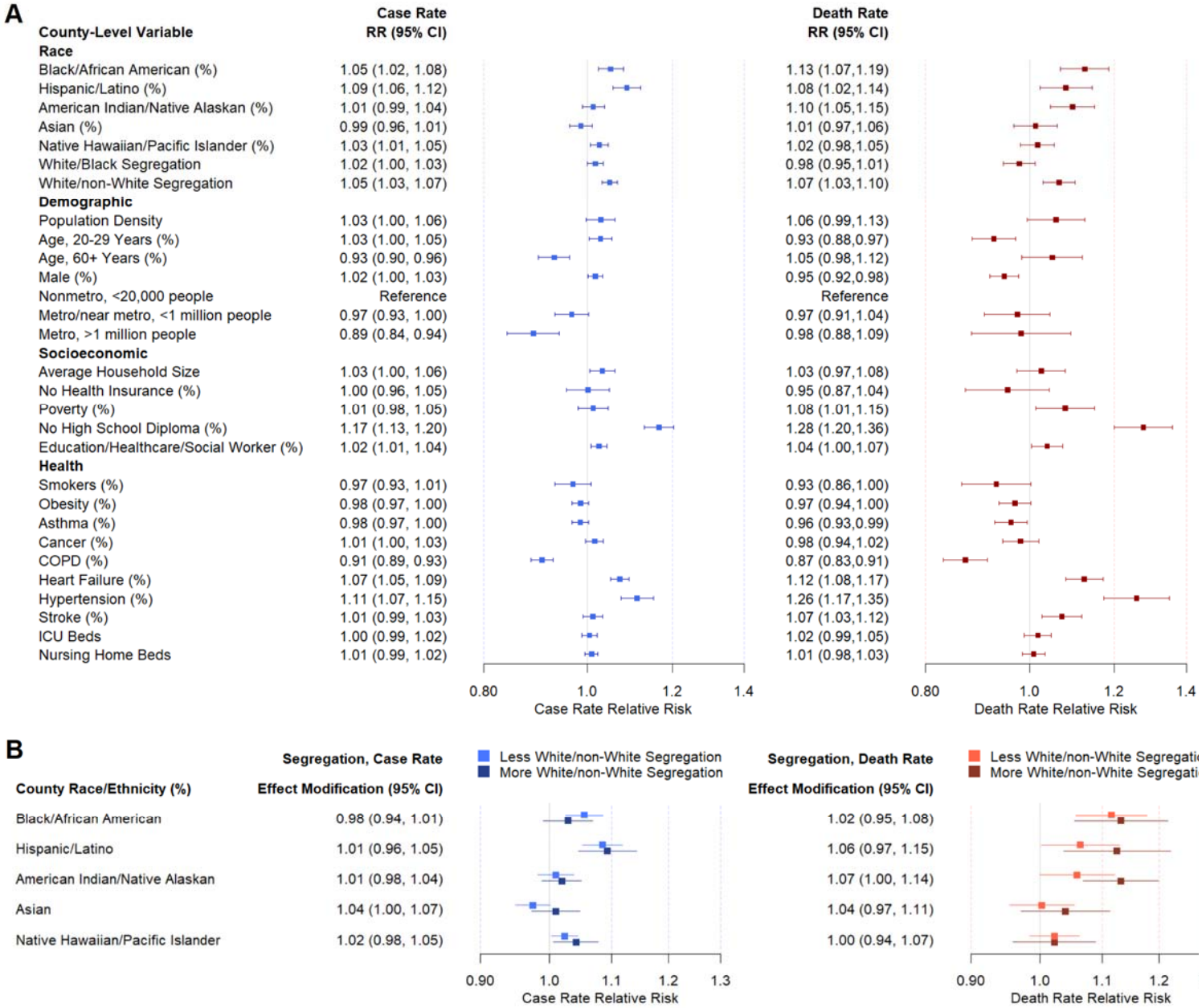
medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 1. Case rate, death rate, racial variable heatmaps. (A) Observed case rates and (B) observed death rates from 1/22/20-11/8/20, and (C) race and ethnicity percentage and residential racial segregation index of dissimilarity (from 0 complete integration to 100 complete segregation) heatmaps.



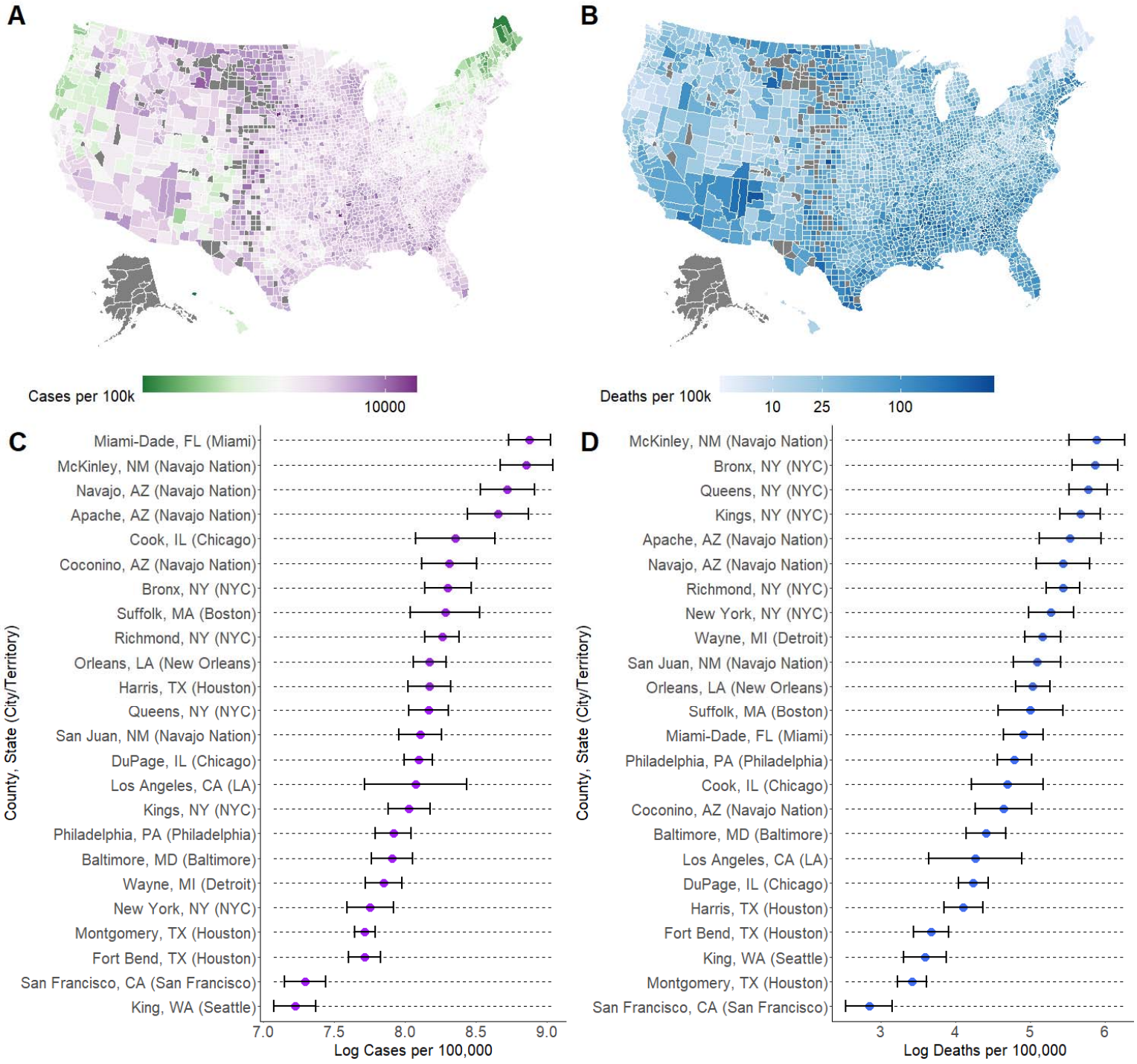
medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 2. Adjusted cumulative case and death rate relative risks. (A) Adjusted relative risks of demographic, socioeconomic, and health comorbidity factors on cumulative COVID-19 case and death rates through 11/8/20. (B) Adjusted race/ethnicity relative risks stratified by counties with White/non-White segregation index of dissimilarity below and above the 75th percentile for case and death rates. Boxes are point estimates and error bars mark 95% confidence intervals. Relative risks are for a one standard deviation increase in a variable, except for the metro/nonmetro categorical variable.



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 3. County estimated cumulative rates. Estimated cumulative case (A) and death (B) rate heat maps from 1/22/20-11/8/20. Estimated rates are similar to observed rates. Estimated case (C) and death (D) rates for selected counties. The estimated rates were calculated using fitted Poisson mixed models. Circles are point estimates; error bars are 95% CIs. Gray counties had missing covariates.



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Comparison of univariate regression (unadjusted) and multiple regression (adjusted) race/ethnicity analyses. Univariate/unadjusted includes log population offset, state fixed effects, and county random effects. Fully adjusted also includes all other variables in the primary models. The coefficients in the fully adjusted models are the same as those in Figure 2A. Bold indicates confidence interval does not include 1.

	Cumulative Case Rate		**Cumulative Death Rate**	
Variable	**Unadjusted**	**Fully Adjusted**	**Unadjusted**	**Fully Adjusted**
Black (%)	**1.21 (1.18, 1.24)**	**1.05 (1.02, 1.08)**	**1.24 (1.19, 1.30)**	**1.13 (1.07, 1.19)**
Hispanic (%)	**1.30 (1.27, 1.33)**	**1.09 (1.06, 1.12)**	**1.23 (1.18, 1.28)**	**1.08 (1.02, 1.14)**
American Indian (%)	**1.05 (1.03, 1.07)**	1.01 (0.99, 1.04)	**1.12 (1.07, 1.17)**	**1.10 (1.05, 1.15)**
Asian (%)	**1.09 (1.07, 1.11)**	0.99 (0.96, 1.01)	1.01 (0.98, 1.05)	1.01 (0.97, 1.06)
Native Hawaiian (%)	**1.11 (1.09, 1.13)**	**1.03 (1.01, 1.05)**	**1.04 (1.00, 1.09)**	1.02 (0.98, 1.05)

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methods

Data Sources

USA Facts is a non-profit organization providing data about government tax revenues, expenditures, and outcomes[14]. Area Health Resources Files is a part of the federal government's Health Resources & Services administrations that includes data on population characteristics, economics, hospital utilization, and more[44]. County Health Rankings & Roadmaps is a collaboration between the Robert Wood Johnson Foundation and University of Wisconsin that provides local community health data[45]. SafeGraph is a company that provides point of information, location, and geographic data[17]. The COVID Tracking Project is a collaborative volunteer-run effort to track the ongoing COVID-19 outbreak in the United States[16].

Outcome Variables

County-level cumulative COVID-19 cases and deaths as of 11/8/20 were directly obtained from USA Facts[46]. USA Facts aggregates data from the Centers of Disease Control and state and local public health agencies. County-level data were confirmed by referencing state and local agencies. Daily percentage time home mobility metrics for each US county from 3/8/20-9/30/20 were obtained from SafeGraph[17].

Covariates

Demographic variables were obtained from Area Health Resource Files, and included population density, median age, and proportion of males in a given county. County-level population distribution by race/ethnicity, including Black/African American,

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Asian, and Hispanic/Latino proportions, were directly obtained from the 2018 US Census Bureau estimates [15]. County residential racial segregation indices of dissimilarity were obtained from County Health Rankings & Roadmaps[47]. These indices were originally calculated from data from US Census tracts from the American Community Survey 2014-2018. Counties with less than 100 Black/non-White residents had the index of dissimilarity set to be equal to 1. The US Department of Agriculture provided rural-urban continuum codes, and these codes were grouped into three categories: metro, population ≥ 1 million (code 1); metro or near metro, population 20,000 to 1 million (codes 2-4); nonmetro, population < 20,000 (codes 5-9).[48]

Socioeconomic variables were obtained from Area Health Resource Files and County Health Rankings & Roadmaps, and included average household size, percentage of individuals between 18-64 years old without health insurance, percentage in poverty, percentage of people aged >25 years without a high school diploma, and percentage of people working in education/health care/social assistance.

Prevalence rates for several comorbidities were obtained from County Health Rankings & Roadmaps, and Centers for Disease Control and Prevention. Comorbidities included county-level percentages for: smoking, asthma, cancer, chronic obstructive pulmonary disease, heart failure, hypertension, and stroke. Kaiser News provided total nursing home beds. Additional details are in **Table S1**.

Statistical Analyses

Cumulative COVID-19 county case and county death counts were each modeled using a Poisson mixed effects model using log(county population size) as an offset.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Models included fixed effects for covariates and indicator variables for each state to account for differences in state geography, testing rates, and other sources of variability. County specific random effects were included to account for overdispersion. Log transformations were applied to racial percentage and population density to reduce the influence of outlier values (**Table S1**). All continuous variables were scaled to have mean 0 and standard deviation 1.

We explored if race relative risks varied by residential segregation by including race percentage and White/non-White segregation interaction terms. Case and death rates were fit with one model each, and results controlled for all covariates in the primary models. We were interested in these interactions because we hypothesized certain counties may have a lower percentage of certain racial/ethnic populations but a large amount of racial segregation that would predispose these racial/ethnic communities to worse COVID-19 outcomes.

Predicted rates were calculated from the cumulative case and death rate models used. Predictions included estimated fixed and random effects (Best Linear Unbiased Prediction) and confidence intervals incorporated estimated standard errors for both fixed and random effects.

Various exploratory and sensitivity analyses were run. We compared multiple regression results with univariate/minimally adjusted regression results. We obtained univariate/minimally adjusted relative rates adjusting for a log population size offset, state fixed effects, and county random effects. Multiple regression/adjusted relative risks additionally controlled for all other covariates in the primary model (**Extended Table**

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

S2). We examined Pearson's correlation between county race percentage and socioeconomic and disease rate variables.

We investigated potential non-linear associations between county race/ethnicity percentages and cases and deaths by categorizing county race/ethnicity percentages into quartiles. We partly performed this analysis to explore the absence of an adjusted association between county Hispanic/Latino percentage and deaths in the primary analyses. We obtained minimally adjusted relative rates adjusting for a log population size offset, state fixed effects, and county random effects. Adjusted relative risks additionally controlled for all other covariates in the primary model (**Extended Table S2**), though the county race/ethnicity variables were categorized into quartiles instead of continuous. We also assessed adding county-level diabetes and kidney disease percentage to improve modeling. Minimally adjusted and adjusted relative risks obtained from the diabetes and kidney disease analyses were similar to the above analyses. Bayesian information criterion (BIC) were compared to evaluate model fit.

County-level SafeGraph daily percentage time at home was modeled in a longitudinal analysis using a generalized estimating equations (GEE) model with an identity link. Percentage time at home was clustered by county with repeated measures for each day from 3/8/20-9/30/20. The same covariates as the primary analyses were used (**Table S2**) without the county-specific random effects or log population size offset; fixed effects for weekday were also added. A cubic spline basis for time with knots every 14 days and an auto-regressive-1 working correlation structure were used to account for serial correlation among repeated measurements over time in each county. Robust sandwich standard errors were calculated. We also included average

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

percentage time home from 3/8/20-9/30/20 as a covariate in modeling case and death rates.

We also performed exploratory Case Fatality Rate (CFR) and Infection Fatality Rate (IFR) analyses to characterize associations of death rates with covariates among infected cases (elaborated on below).

We also explored including state testing rates as a covariate for modeling. However, we found this did not change any estimated effects adjusted for demographic, socioeconomic, and comorbidity variables because we already controlled for fixed state effects using state dummy variables. Adding state testing rates only changed the estimated state fixed effects through re-parametrization.

All analyses were conducted in R. The following packages were used in formatting data: data.table, dplyr. The following packages were used in formatting results and creating plots: ggplot2, usmap, gridExtra, tidyverse, plyr. The following packages were used in modeling: glmnet, geepack, geeM, lme4, splines. Code for these analyses is available as described in the code availability section.

<u>Exploratory Case Fatality Rate and Infection Fatality Rate Regression</u>

We conducted exploratory Case Fatality Rate (CFR) and Infection Fatality Rate (IFR) regression analysis to investigate COVID-19 health disparities on the death rate among infected subjects. CFR were calculated by dividing the number of deaths by the number of reported cases in each county. CFR regression was performed in a similar way to the cumulative death rate regression by fitting Poisson mixed models except for using an offset for log(total reported cases) instead of log(population size). IFR were

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

calculated by dividing the number of deaths by the number of infected cases in each county.

Since county-specific number of total infected cases were not observed and would likely be underestimated by using the numbers of reported cases, we estimated the county-specific total number of infected cases by dividing the number of total reported cases by a constant ascertainment rate of cases. Since the county-specific testing data and ascertainment rates were not available, we modeled them using state-specific fixed effects and county-specific random effects and allowed the ascertainment rates to vary between states and counties.

To define the CFR model, assume R_{ij} is the number of reported cases and μ_{ij} is the mean number of cumulative deaths in county *j* of state *i*. Then the CFR Poisson mixed model can be written as

$$\ln(\mu_{ij}) = \ln(R_{ij}) + \theta_i + X'_{ij}\delta + e_{ij},$$

where θ_i is the state effect, X_{ij} is a vector of covariates, and the e_{ij} are county-specific random effects to account for overdispersion.

For the IFR model, let I_{ij} be the unobserved total number of infected cases in county j of state i. Suppose a_{ij} is the unobserved ascertainment rate for state *i* and county *j*. The estimated ascertainment rate has been estimated to be between 4% to 20%[24]. We have $I_{ij} = \frac{R_{ij}}{a_{ij}}$. Let $a_{ij} = c_i d_{ij}$, where c_i is the overall ascertainment rate for state i and d_{ij} is the multiplicative departure of the ascertainment rate of county j from the state level acertainment rate c_i. Then the IFR Poisson mixed model can be written as

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$\ln(\mu_{ij}) = \ln(I_{ij}) + X'_{ij}\alpha + \beta_i + b_{ij} = \ln(R_{ij}) - \ln(c_i) + \beta_i + X'_{ij}\alpha - \ln(d_{ij}) + b_{ij},$$

where β_i is the state effect. Write $\gamma_i = -\ln(c_i) + \beta_i$ and $u_{ij} = -\ln(d_{ij}) + b_{ij}$. Then we have

$$\ln(\mu_{ij}) = \ln(R_{ij}) + \gamma_i + X'_{ij}\alpha + u_{ij}.$$

Assuming u_{ij} to be county-specific random effects following a normal distribution $N(0, \tau)$, it follows that the IFR regression model is identical to the CFR regression model except that the estimated state effects and the county-specific random effects can also be interpreted as capturing the state and county-level ascertainment rates. Therefore, the identical results for CFR and IFR are presented in **Extended Data Fig. S7**. Future work can investigate allowing the county-specific ascertainment rate to be better estimated when county-specific testing data are available.

The CFR/IFR results had similar directions as the primary death rate results but there were few significant variables, possibly because of differential underestimation of asymptomatic and mildly asymptomatic cases by race/ethnicity[23], selection bias associated with both the subjects who were tested, e.g., symptomatic subjects were more likely to be tested and to test positive, large fluctuations in the numbers of tests from county to county, and insufficient testing capacity.

Additional data collection, such as county-level test data and race/ethnicity specific case and death counts, is needed to better estimate the number of infected cases by estimating county-specific ascertainment rates more accurately, in order to make IFR analysis results more reliable.

Limitations and Strengths

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

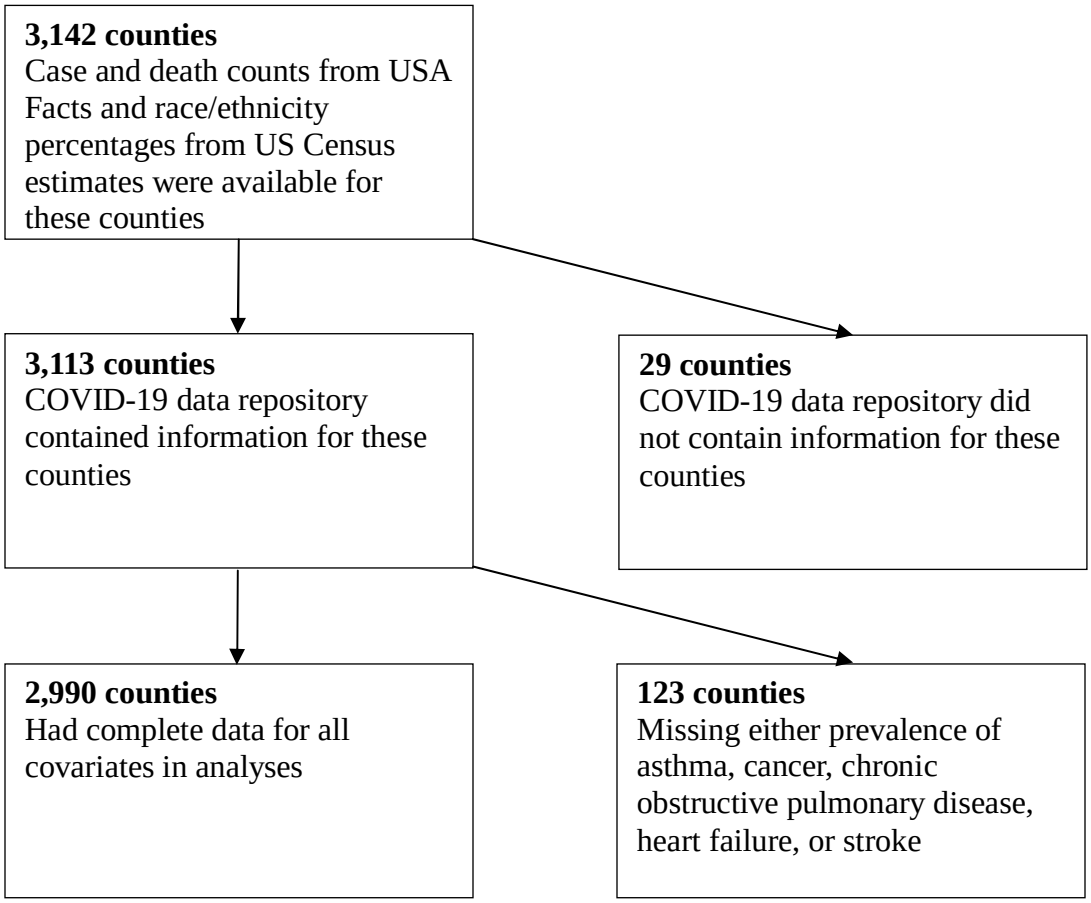
As an ecological study, there are some limitations to our study. We analyzed aggregated county-level data which were subject to confounding bias and bias due to data aggregation [49]. Individual level data from electronic health records or case control or community studies would provide valuable information[31,50]. Additional variables such as household and community exposures to COVID-19 cases, occupation (e.g., health care workers or essential workers), public transportation usage, COVID-19 symptoms, living resources, and medications could improve modeling for studying disparities and provide more control of confounders[51].

However, there are also many strengths to our study. We used nationally compiled county data from various reputable sources, and we showed that many findings from previous studies and reports on the city and state level generalize to the national level. Our analyses also adjust for regional and state-to-state variation, and while many studies adjust for demographic and either socioeconomic or health variables, we present results jointly adjusting for all three domains, as well as county resident mobility.

Many other studies have focused solely on Black/African American or a more general non-White ethnicity, but we also incorporate and study Hispanic/Latinos, Asians, and American Indians/Native Alaskan effects separately. We also explore the effects of residential racial segregation indices of dissimilarity calculated from US Census tract level data. We also studied the effects of race/ethnicity and other demographic variables, socioeconomic and comorbidity factors on social distancing using mobility-based time at home measures.

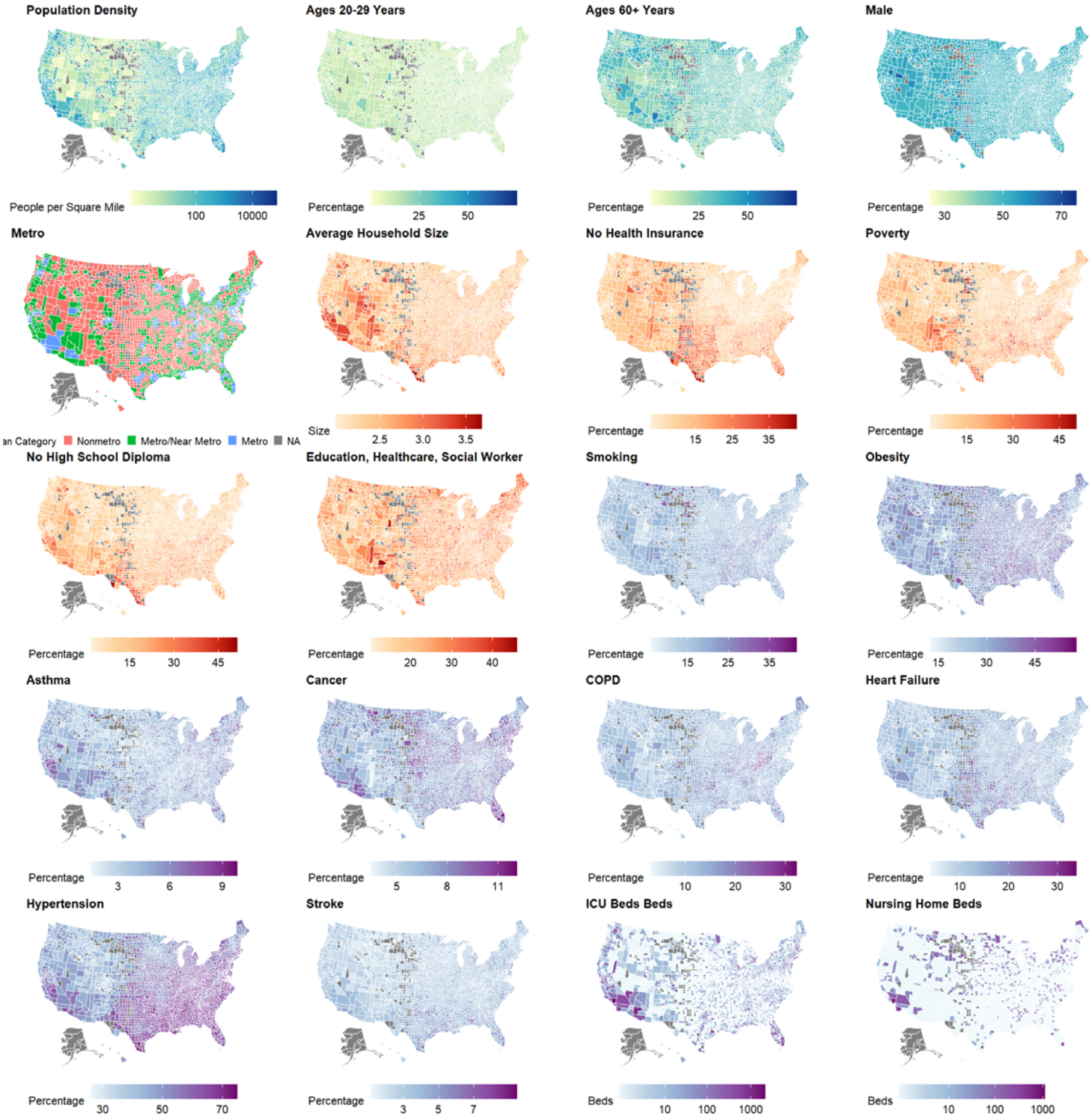
medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. S1. Study sample size flow chart.



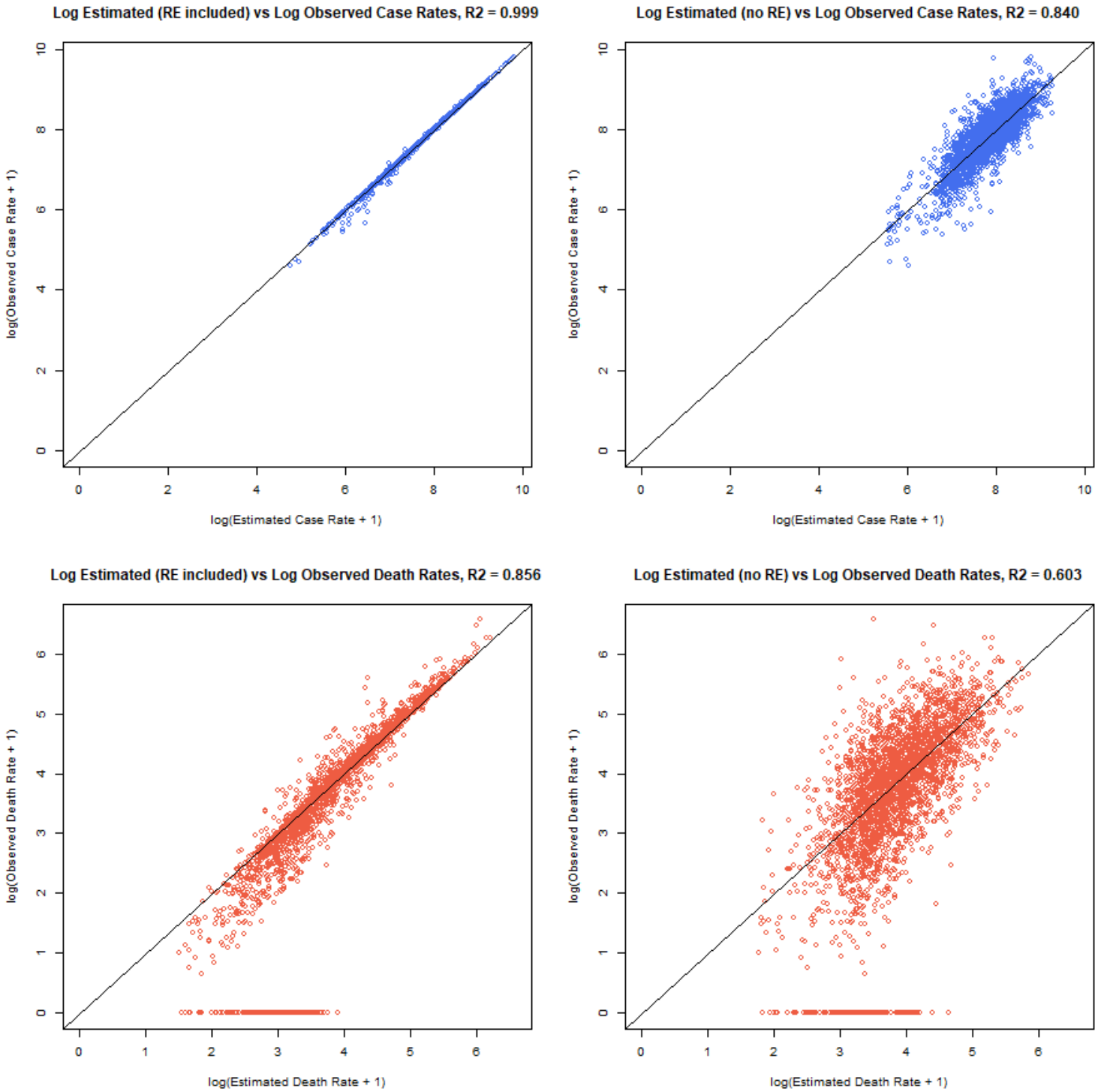
medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. S2. Covariate heatmaps. Demographic (yellow, aqua, blue), socioeconomic (orange, red), and health (blue, purple) heatmaps. Gray counties have missing covariates.



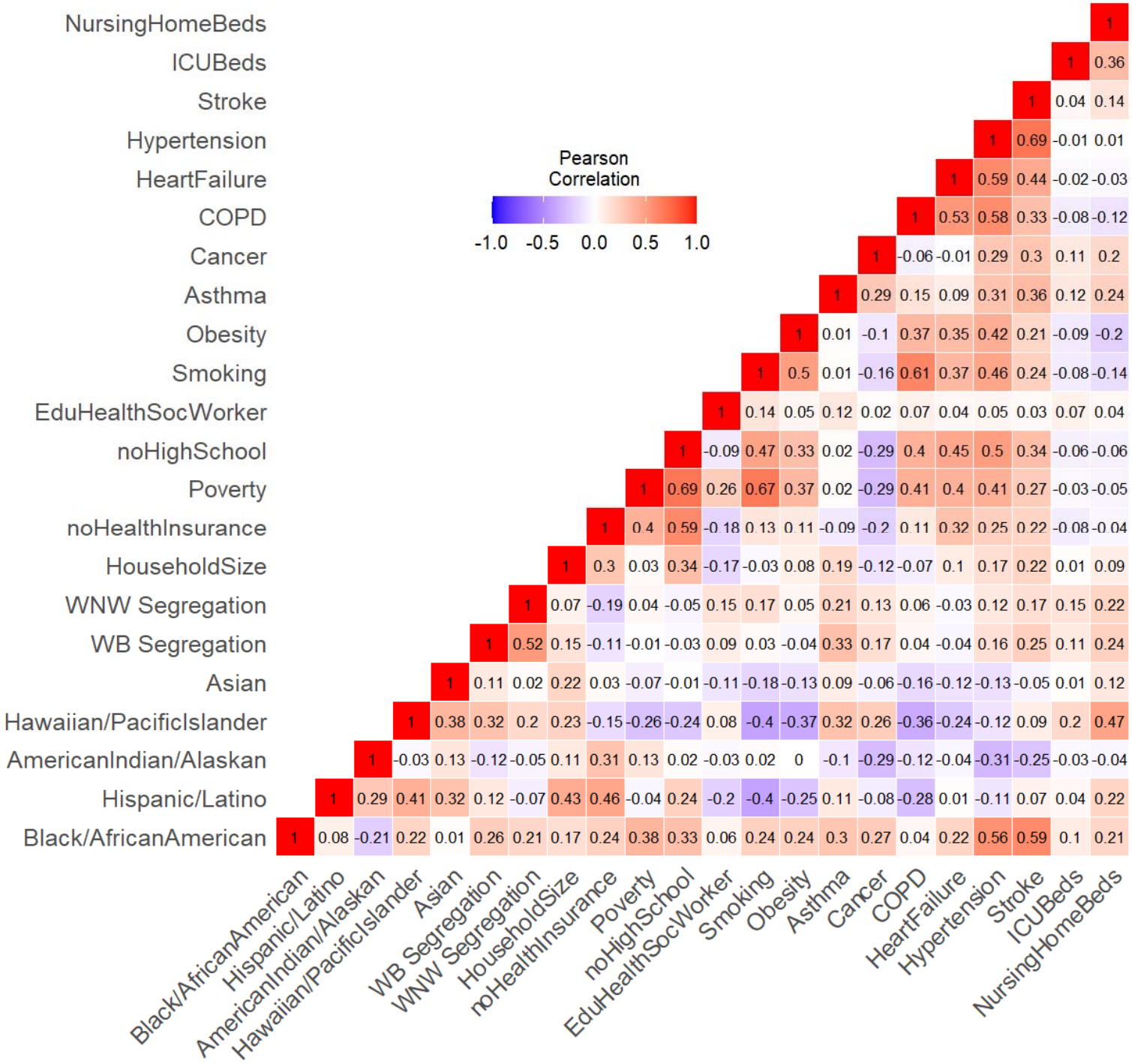
medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. S3. Estimated vs observed cumulative case and death rates. Observed and estimated (with and without random effects) log case rates (blue) and log death rates (red). Dark line has slope 1 intercept 0. RE = random effects; R2 = Pearson's correlation coefficient.



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

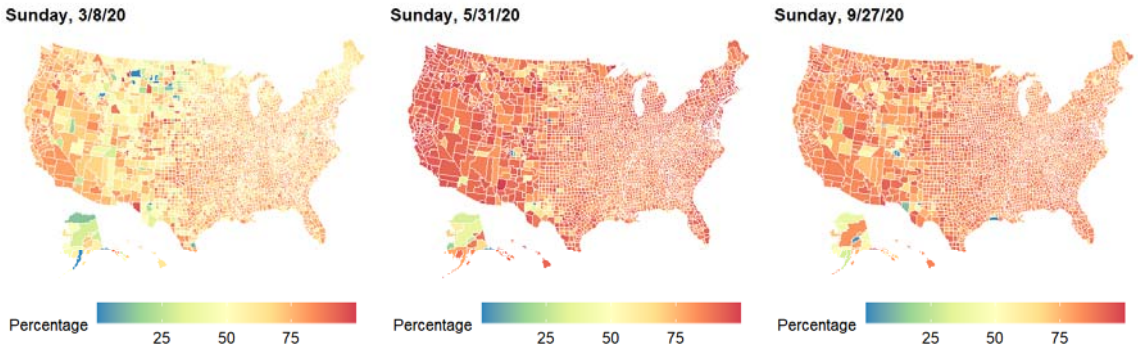
Fig. S4. Covariate correlations. Heatmap of correlations between racial, socioeconomic, and health covariates.



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

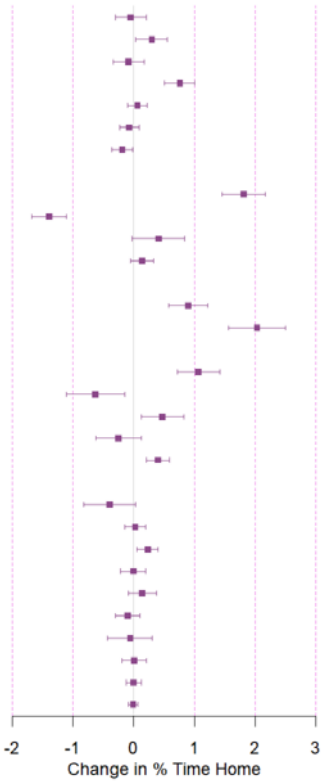
Fig. S5 Adjusted time at home effects. (A) Heatmaps of county residents' percentage time at home at selected dates. Modeling accounts for all dates. (B) Adjusted effects of county-level covariates on the change in percentage time at home for US counties from 3/8/20-9/30/20 using longitudinal analysis. Positive values are associated with more time at home, and negative values with less time at home. Boxes are point estimates and error bars mark 95% CIs. Effects are for a one standard deviation increase in a variable, except for the metro/nonmetro categorical variable.

A



B

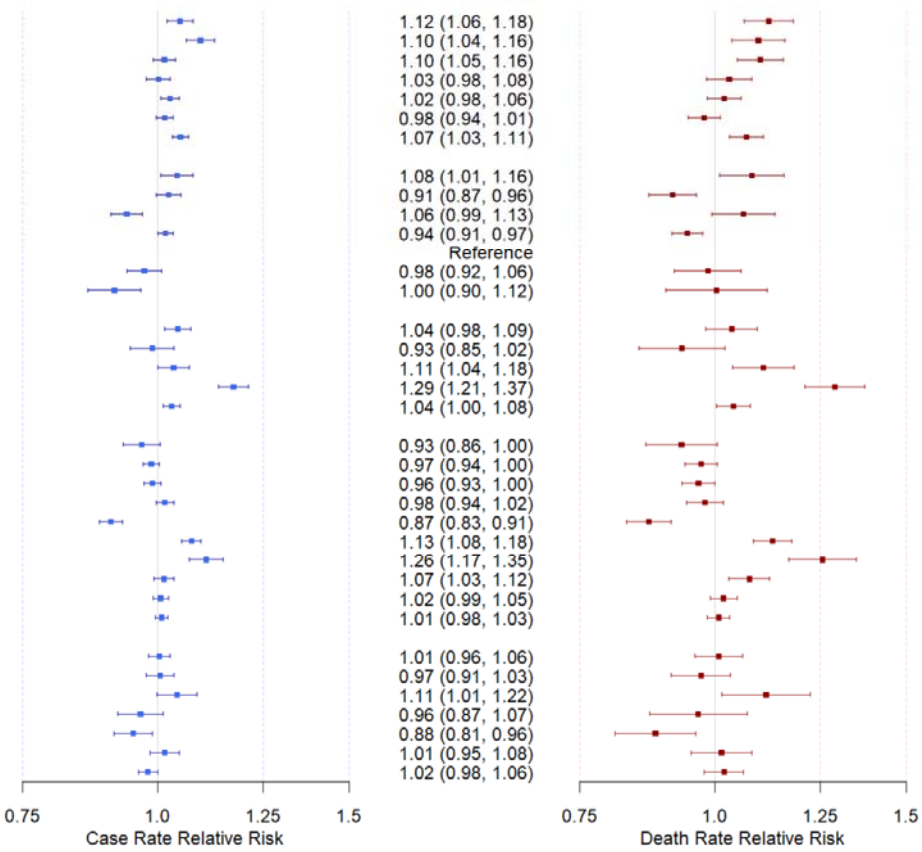
County-Level Variable	% Time Home Change (95% CI)
Race	
Black/African American (%)	-0.05 (-0.30, 0.20)
Hispanic/Latino (%)	0.30 (0.04, 0.56)
American Indian/Native Alaskan (%)	-0.08 (-0.34, 0.17)
Asian (%)	0.76 (0.51, 1.01)
Native Hawaiian/Pacific Islander (%)	0.06 (-0.10, 0.22)
White/Black Segregation	-0.07 (-0.23, 0.09)
White/non-White Segregation	-0.19 (-0.36, -0.02)
Demographic	
Population Density	1.81 (1.46, 2.16)
Age, 20-29 Years (%)	-1.39 (-1.68, -1.10)
Age, 60+ Years (%)	0.41 (-0.02, 0.84)
Male (%)	0.14 (-0.05, 0.33)
Nonmetro, <20,000 people	Reference
Metro/near metro, <1 million people	0.90 (0.58, 1.22)
Metro, >1 million people	2.03 (1.56, 2.50)
Socioeconomic	
Average Household Size	1.07 (0.72, 1.42)
No Health Insurance (%)	-0.63 (-1.11, -0.15)
Poverty (%)	0.47 (0.12, 0.82)
No High School Diploma (%)	-0.25 (-0.62, 0.12)
Education/Healthcare/Social Worker (%)	0.40 (0.21, 0.59)
Health	
Smokers (%)	-0.39 (-0.82, 0.04)
Obesity (%)	0.02 (-0.15, 0.19)
Asthma (%)	0.23 (0.06, 0.40)
Cancer (%)	-0.01 (-0.21, 0.20)
COPD (%)	0.14 (-0.09, 0.37)
Heart Failure (%)	-0.10 (-0.30, 0.10)
Hypertension (%)	-0.06 (-0.43, 0.31)
Stroke (%)	0.01 (-0.20, 0.21)
ICU Beds	0.00 (-0.12, 0.12)
Nursing Home Beds	-0.01 (-0.09, 0.07)



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. S6. Time at home covariate for case and death rates. Adjusted relative risks of covariates on cumulative COVID-19 case and death rates through 11/8/20 additionally controlling for monthly average percentage time home from 3/8/20-9/30/20. Boxes are point estimates and error bars mark 95% CIs. Relative risks are for a one standard deviation increase in a variable, except for the metro/nonmetro categorical variable.

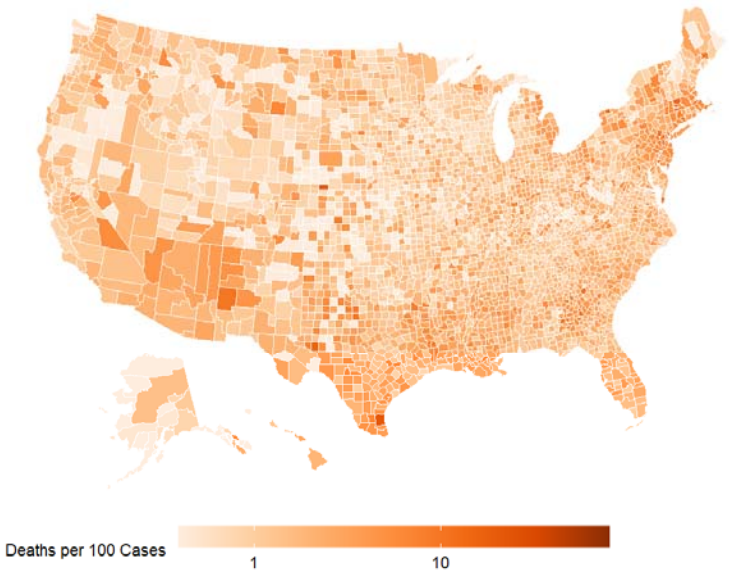
County-Level Variable	Case Rate RR (95% CI)	Death Rate RR (95% CI)
Race		
Black/African American (%)	1.05 (1.02, 1.08)	1.12 (1.06, 1.18)
Hispanic/Latino (%)	1.09 (1.06, 1.13)	1.10 (1.04, 1.16)
American Indian/Native Alaskan (%)	1.01 (0.99, 1.04)	1.10 (1.05, 1.16)
Asian (%)	1.00 (0.98, 1.03)	1.03 (0.98, 1.08)
Native Hawaiian/Pacific Islander (%)	1.03 (1.01, 1.05)	1.02 (0.98, 1.06)
White/Black Segregation	1.01 (1.00, 1.03)	0.98 (0.94, 1.01)
White/non-White Segregation	1.05 (1.03, 1.07)	1.07 (1.03, 1.11)
Demographic		
Population Density	1.04 (1.01, 1.08)	1.08 (1.01, 1.16)
Age, 20-29 Years (%)	1.02 (1.00, 1.05)	0.91 (0.87, 0.96)
Age, 60+ Years (%)	0.94 (0.90, 0.97)	1.06 (0.99, 1.13)
Male (%)	1.02 (1.00, 1.03)	0.94 (0.91, 0.97)
Nonmetro, <20,000 people	Reference	Reference
Metro/near metro, <1 million people	0.97 (0.94, 1.01)	0.98 (0.92, 1.06)
Metro, >1 million people	0.91 (0.86, 0.96)	1.00 (0.90, 1.12)
Socioeconomic		
Average Household Size	1.04 (1.01, 1.07)	1.04 (0.98, 1.09)
No Health Insurance (%)	0.99 (0.94, 1.03)	0.93 (0.85, 1.02)
Poverty (%)	1.03 (1.00, 1.07)	1.11 (1.04, 1.18)
No High School Diploma (%)	1.17 (1.14, 1.21)	1.29 (1.21, 1.37)
Education/Healthcare/Social Worker (%)	1.03 (1.01, 1.05)	1.04 (1.00, 1.08)
Health		
Smokers (%)	0.97 (0.93, 1.00)	0.93 (0.86, 1.00)
Obesity (%)	0.99 (0.97, 1.00)	0.97 (0.94, 1.00)
Asthma (%)	0.99 (0.97, 1.01)	0.96 (0.93, 1.00)
Cancer (%)	1.01 (1.00, 1.03)	0.98 (0.94, 1.02)
COPD (%)	0.90 (0.88, 0.93)	0.87 (0.83, 0.91)
Heart Failure (%)	1.07 (1.05, 1.10)	1.13 (1.08, 1.18)
Hypertension (%)	1.11 (1.07, 1.15)	1.26 (1.17, 1.35)
Stroke (%)	1.01 (0.99, 1.03)	1.07 (1.03, 1.12)
ICU Beds	1.01 (0.99, 1.02)	1.02 (0.99, 1.05)
Nursing Home Beds	1.01 (0.99, 1.02)	1.01 (0.98, 1.03)
Mobility		
Average Time Home in March (%)	1.00 (0.98, 1.03)	1.01 (0.96, 1.06)
Average Time Home in April (%)	1.00 (0.98, 1.03)	0.97 (0.91, 1.03)
Average Time Home in May (%)	1.04 (1.00, 1.09)	1.11 (1.01, 1.22)
Average Time Home in June (%)	0.96 (0.92, 1.01)	0.96 (0.87, 1.07)
Average Time Home in July (%)	0.95 (0.91, 0.99)	0.88 (0.81, 0.96)
Average Time Home in August (%)	1.01 (0.98, 1.05)	1.01 (0.95, 1.08)
Average Time Home in September (%)	0.98 (0.96, 1.00)	1.02 (0.98, 1.06)



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

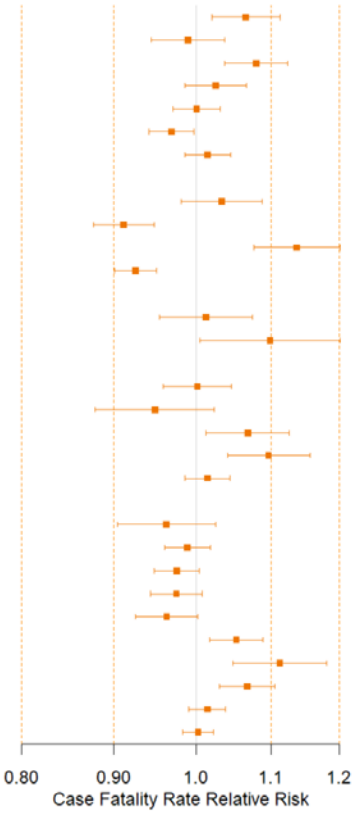
Fig. S7. Case/infection fatality rate analysis. (A) Heatmap of cumulative case fatality rates by county. (B) Case fatality relative risks. Dots are point estimates and error bars mark 95% CIs. Regression coefficient estimates in infection fatality rate models are identical (**Online Methods**)

A



B

County-Level Variable	Case Fatality Rate RR (95% CI)
Race	
Black/African American (%)	1.06 (1.02,1.11)
Hispanic/Latino (%)	0.99 (0.94,1.04)
American Indian/Native Alaskan (%)	1.08 (1.04,1.12)
Asian (%)	1.02 (0.99,1.07)
Native Hawaiian/Pacific Islander (%)	1.00 (0.97,1.03)
White/Black Segregation	0.97 (0.94,1.00)
White/non-White Segregation	1.01 (0.98,1.04)
Demographic	
Population Density	1.03 (0.98,1.09)
Age, 20-29 Years (%)	0.91 (0.88,0.95)
Age, 60+ Years (%)	1.14 (1.08,1.20)
Male (%)	0.93 (0.90,0.95)
Nonmetro, <20,000 people	Reference
Metro/near metro, <1 million people	1.01 (0.95,1.07)
Metro, >1 million people	1.10 (1.00,1.20)
Socioeconomic	
Average Household Size	1.00 (0.96,1.05)
No Health Insurance (%)	0.95 (0.88,1.02)
Poverty (%)	1.07 (1.01,1.13)
No High School Diploma (%)	1.10 (1.04,1.16)
Education/Healthcare/Social Worker (%)	1.01 (0.99,1.04)
Health	
Smokers (%)	0.96 (0.90,1.02)
Obesity (%)	0.99 (0.96,1.02)
Asthma (%)	0.98 (0.95,1.00)
Cancer (%)	0.97 (0.94,1.01)
COPD (%)	0.96 (0.93,1.00)
Heart Failure (%)	1.05 (1.02,1.09)
Hypertension (%)	1.11 (1.05,1.18)
Stroke (%)	1.07 (1.03,1.11)
ICU Beds	1.01 (0.99,1.04)
Nursing Home Beds	1.00 (0.98,1.02)



medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S1. List of county-level variables, transformations, and sources.

Variable Name	Date	Transformation	Source and Description
Outcomes			
Cumulative COVID-19 Cases	11/8/20		USA Facts
Cumulative COVID-19 Deaths	11/8/20		USA Facts
Daily Time at Home (%)	3/8/20-9/30/20		SafeGraph
Race/ethnicity			
Black/African American (%)	2018	log(x+1)	US Census Bureau
Asian (%)	2018	log(x+1)	US Census Bureau
Hispanic/Latino (%)	2018	log(x+1)	US Census Bureau
American Indian/Native Alaskan (%)	2018	log(x+1)	US Census Bureau
White Black Segregation Index	2014-2018		County Health Rankings & Roadmaps
White non-White Segregation Index	2014-2018		County Health Rankings & Roadmaps
Demographics			
Population Size	2019		US Census Bureau
Population Density	2019	log(x)	US Census Bureau
Age, 20-29 years (%)	2019		US Census Bureau
Age, 60+ years (%)	2018		US Census Bureau
Male (%)	2018		US Census Bureau
Rural Urban Continuum Code	2013		US Department of Agriculture
Socioeconomic			
Average Household Size	2010		Area Health Resources Files
No Health Insurance, 18-64 years (%)	2017		Area Health Resources Files
Poverty (%)	2017		Area Health Resources Files
No High School Diploma, 25+ years (%)	2013-17		Area Health Resources Files
Education, Health Care, Social Assistance Workers (%)	2013-17		Area Health Resources Files
Health			
Smokers (%)	2017		County Health Rankings & Roadmaps
Obesity (%)	2017		County Health Rankings & Roadmaps
Asthma (%)	2017		County Health Rankings & Roadmaps
Cancer (%)	2017		County Health Rankings & Roadmaps
COPD (%)	2017		County Health Rankings &

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

			Roadmaps
Heart Failure (%)	2017		County Health Rankings & Roadmaps
Hypertension (%)	2017		County Health Rankings & Roadmaps
Stroke (%)	2017		County Health Rankings & Roadmaps
Nursing Home Beds	2017		Kaiser Health News

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S2. Primary model county-covariates.

Type	Category	Variables
Outcome	Count	Total cases, total deaths
Offset	Rate Denominator	log(Population Size)
Fixed Effects	State Differences	State
	Racial/ethnic	Black (%) Asian (%) Hispanic (%) American Indian (%) White Black Segregation Index White non-White Segregation Index
	Demographic	Population Density Age 20-29 years Age 65+ years (%) Male (%) Rural-Urban (categorical)
	Socioeconomic	Average household size No health insurance (%) Poverty (%) No high school diploma (%) Education, health care, social assistance worker (%)
	Health	Smokers (%) Obesity (%) Asthma (%) Cancer (%) Chronic obstructive pulmonary disease (%) Heart Failure (%) Hypertension (%) Stroke (%) Nursing home beds
Random Effects	Overdispersion	County

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S3. American Indian Case Rate Analysis. Results below additionally includes log population offset, state fixed effects, and county random effects**.** Bold indicates confidence interval does not include 1.

Variables	**American Indian Case Rate Relative Risk**
American Indian	**1.05 (1.03, 1.07)**
American Indian + Black + Hispanic + Asian + Native Hawaiian	**1.03 (1.01, 1.05)**
American Indian + No Health Insurance	0.99 (0.96, 1.01)

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S4. Asian Case and Death Rate Analysis. Results below additionally includes log population offset, state fixed effects, and county random effects**.** Bold indicates confidence interval does not include 1.

Variables	**Asian Case Rate Relative Risk**
Asian	**1.09 (1.07, 1.11)**
Asian + Black + Hispanic + American Indian + Native Hawaiian	**0.98 (0.96, 0.99)**
Asian + No Health Insurance	**1.12 (1.10, 1.14)**

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S5. Native Hawaiian Death Rate Analysis. Results below additionally includes log population offset, state fixed effects, and county random effects**.** Bold indicates confidence interval does not include 1.

Variables	**Native Hawaiian Death Rate Relative Risk**
Native Hawaiian	**1.04 (1.00, 1.09)**
Native Hawaiian + Black + Hispanic + American Indian + Asian	0.99 (0.95, 1.03)
Native Hawaiian + No Health Insurance	1.03 (0.99, 1.07)

medRxiv preprint doi: https://doi.org/10.1101/2020.12.02.20234989; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table S6. Categorical race/ethnicity analysis. Minimally adjusted includes log population offset, state fixed effects, and county random effects. Fully adjusted also includes all other variables in the primary models. Bold indicates confidence interval does not include 1.

	Total Case Rate		**Total Death Rate**	
Variable	**Minimally adjusted**	**Fully Adjusted**	**Minimally adjusted**	**Fully Adjusted**
Black (%)				
1st quartile	(reference)	(reference)	(reference)	(reference)
2nd quartile	**1.14 (1.09, 1.19)**	1.03 (0.98, 1.07)	0.96 (0.88, 1.05)	1.01 (0.92, 1.10)
3rd quartile	**1.30 (1.23, 1.36)**	**1.06 (1.00, 1.11)**	1.04 (0.95, 1.15)	1.05 (0.95, 1.16)
4th quartile	**1.57 (1.48, 1.67)**	**1.13 (1.06, 1.21)**	**1.47 (1.31, 1.64)**	**1.21 (1.07, 1.38)**
Hispanic (%)				
1st quartile	(reference)	(reference)	(reference)	(reference)
2nd quartile	**1.07 (1.02, 1.12)**	1.02 (0.98, 1.06)	1.04 (0.96, 1.13)	1.08 (0.99, 1.17)
3rd quartile	**1.16 (1.10, 1.21)**	1.03 (0.98, 1.08)	1.07 (0.98, 1.17)	**1.12 (1.02, 1.24)**
4th quartile	**1.58 (1.49, 1.68)**	**1.15 (1.07, 1.23)**	**1.38 (1.24, 1.54)**	**1.21 (1.07, 1.38)**
American Indian (%)				
1st quartile	(reference)	(reference)	(reference)	(reference)
2nd quartile	1.04 (0.99, 1.09)	0.97 (0.93, 1.01)	1.02 (0.94, 1.12)	0.97 (0.90, 1.05)
3rd quartile	1.05 (1.00, 1.12)	0.96 (0.92, 1.01)	1.01 (0.92, 1.12)	0.95 (0.86, 1.04)
4th quartile	**1.13 (1.06, 1.21)**	0.96 (0.90, 1.02)	**1.15 (1.02, 1.29)**	0.98 (0.87, 1.11)
Asian (%)				
1st quartile	(reference)	(reference)	(reference)	(reference)
2nd quartile	1.05 (1.00, 1.09)	1.01 (0.97, 1.05)	0.96 (0.88, 1.04)	1.01 (0.93, 1.09)
3rd quartile	**1.10 (1.05, 1.16)**	1.02 (0.97, 1.06)	0.96 (0.88, 1.04)	1.01 (0.93, 1.11)
4th quartile	**1.22 (1.16, 1.28)**	1.00 (0.95, 1.06)	0.95 (0.87, 1.03)	1.03 (0.92, 1.15)
Native Hawaiian (%)				
1st quartile	(reference)	(reference)	(reference)	(reference)
2nd quartile	**1.06 (1.01, 1.11)**	1.02 (0.98, 1.05)	0.98 (0.91, 1.07)	1.00 (0.93, 1.08)
3rd quartile	**1.06 (1.01, 1.11)**	1.00 (0.97, 1.05)	0.94 (0.87, 1.03)	0.99 (0.91, 1.07)
4th quartile	**1.25 (1.19, 1.31)**	1.04 (1.00, 1.10)	1.05 (0.95, 1.15)	1.00 (0.91, 1.10)