



# Racial and Ethnic Disparities in COVID-19: Evidence from Six Large Cities

Joseph Benitez<sup>1</sup> · Charles Courtemanche<sup>2</sup> · Aaron Yelowitz<sup>2</sup>

Received: 22 July 2020 / Revised: 7 October 2020 / Accepted: 14 October 2020 / Published online: 24 October 2020  
© Springer Nature Switzerland AG 2020

## Abstract

As of June 2020, the coronavirus pandemic has led to more than 2.3 million confirmed infections and 121 thousand fatalities in the USA, with starkly different incidence by race and ethnicity. Our study examines racial and ethnic disparities in confirmed COVID-19 cases across six diverse cities—Atlanta, Baltimore, Chicago, New York City, San Diego, and St. Louis—at the ZIP code level (covering 436 “neighborhoods” with a population of 17.7 million). Our analysis links these outcomes to six separate data sources to control for demographics; housing; socioeconomic status; occupation; transportation modes; health care access; long-run opportunity, as measured by income mobility and incarceration rates; human mobility; and underlying population health. We find that the proportions of Black and Hispanic residents in a ZIP code are both positively and statistically significantly associated with COVID-19 cases per capita. The magnitudes are sizeable for both Black and Hispanic, but even larger for Hispanic. Although some of these disparities can be explained by differences in long-run opportunity, human mobility, and demographics, most of the disparities remain unexplained even after including an extensive list of covariates related to possible mechanisms. For two cities—Chicago and New York—we also examine COVID-19 fatalities, finding that differences in confirmed COVID-19 cases explain the majority of the observed disparities in fatalities. In other words, the higher death toll of COVID-19 in predominantly Black and Hispanic communities mostly reflects higher case rates, rather than higher fatality rates for confirmed cases.

**Keywords** COVID-19 · Coronavirus · Racial disparities · Ethnic disparities · Health disparities

**JEL Classification** I14

## Introduction

By June 2020, the COVID-19 pandemic in the USA led to over 2.3 million confirmed infections, over 121,000 fatalities, and almost 31,000 hospitalizations (CDC, U. S. Centers for

Disease Control and Prevention 2020, 2020c). Like many other issues pertaining to health and economic disparities, the burden of the COVID-19 pandemic falls disproportionately on Black and Hispanic communities. Through June 13, 2020, the rate of hospitalization for Blacks and Hispanics was more than four times as high as for Whites (CDC, U. S. 2020b). The pandemic is taking a substantial toll on physical, mental, and economic health across the USA, but disparities in whom is impacted by the virus are an additional cause for alarm.

Part of the barrier to carefully studying disparities from the pandemic—both their magnitudes and potential explanations—is that current published data on COVID-19 outcomes are coarse (Killeen et al. 2020). Most studies focus on COVID-19 outcomes at highly aggregated levels such as the state (Friedson et al. 2020) or county (Courtemanche et al. 2020). Within a state or county—especially in areas with larger populations—residential segregation by race and ethnicity

---

✉ Aaron Yelowitz  
aaron@uky.edu

Joseph Benitez  
joseph.benitez@uky.edu

Charles Courtemanche  
courtemanche@uky.edu

<sup>1</sup> Department of Health Management and Policy, College of Public Health, University of Kentucky, Lexington, KY 40536, USA

<sup>2</sup> Department of Economics, Gatton Business & Economics Building, University of Kentucky, Lexington, KY 40506-0034, USA

can be quite stark, meaning that analyzing disparities at such levels misses an important part of the variation. Partway through the pandemic, a number of local and state governments began producing COVID-19 statistics in a more disaggregated fashion—most commonly reporting confirmed cases by ZIP code. This paper provides one of the first attempts to systematically investigate racial/ethnic disparities in COVID-19 using these newly available ZIP code level data. Specifically, we utilize data on confirmed COVID-19 cases from six cities—New York, Chicago, Atlanta, Baltimore, San Diego, and St. Louis—as well as data on fatalities from New York and Chicago.

Our analysis links these COVID-19 outcomes to six separate data sources to control for ZIP code level demographics, housing, socioeconomic status, occupational choices, transportation modes, health care access, long-run opportunity (income mobility and incarceration rates), human mobility, and population health disparities. This rich set of covariates allows us to investigate the extent to which mechanisms that have received popular attention—such as income; education; living in densely populated communities; reliance on public transportation; representation in forward-facing, essential jobs; mobility during lockdowns; pre-pandemic health; and access to health care—contribute to racial and ethnic disparities (Harrison 2020; Hubler et al. 2020; Oppel et al. 2020; CDC, U. S. 2020b). We find statistically significant and economically meaningful disparities for both Blacks and Hispanics at the ZIP code level in confirmed cases, and most of the disparity remains unexplained even after including extensive controls. Without additional covariates, a 10 percentage point increase in a ZIP code's share of Black residents is associated with 9.2 additional confirmed COVID-19 cases per 10,000 residents, while a similar change in the Hispanic share is associated with 20.6 additional cases. Both are sizable changes relative to the average confirmed case rate of 153 per 10,000 population. Using decompositions that are insensitive to the ordering of the covariates (Gelbach 2016), we find that at least part of these disparities can be explained by differences in long-run opportunity (income mobility and incarceration rates), human mobility as measured by cell phone activity, and demographics. However, even with an extensive set of controls, more than half of the disparity in COVID-19 cases remains unexplained. For the two cities where we are also able to examine COVID-19 fatalities, we find that differences in confirmed COVID-19 cases strongly predict the observed disparities, in fact entirely eliminating the entire association for proportion Hispanic and the majority of the association for proportion Black.

The remainder of the paper is arranged as follows. We first examine the rapidly evolving COVID-19 literature with respect to disparities due to race and ethnicity. Next, we discuss our data collection effort. We then provide an empirical model and findings, and then conclude.

## Literature Review

Despite growing recognition about racial and ethnic disparities in COVID-19, our study is one the first to systematically investigate their size and possible explanations at a geographic level narrower than the county.<sup>1</sup> Below, we summarize the literature on racial disparities pertaining to COVID-19 infections, testing, and deaths, with the caveat that the literature is rapidly evolving. We then discuss, based on previous work, limitations on county-level analyses and underlying mechanisms through which disparities could occur.

### Nationwide, County-Level Analyses

Several recent studies utilize data on COVID-19 outcomes from US counties and analyze racial and ethnic disparities. McLaren (2020) collects county-level data on COVID-19 mortality from the entire USA and links to county characteristics from the American Community Survey (ACS). By May 19, 2020, the unadjusted disparity shows a 10 percentage point increase in the Black share corresponding to an increase of 37.6 additional fatalities per million, with a 10 percentage point increase in the Hispanic share associated with 9.6 additional fatalities. With additional controls—especially for public transit—the disparities decrease and in some cases become insignificant (suggesting potential mechanisms for the observed disparity). The study concludes that the disparity for Blacks is very robust to the inclusion of additional covariates, although the one for Hispanics is more fragile.

Knittel and Ozaltun (2020) also examine COVID-19 death rates at the county-level and find mixed evidence of disparities. As of May 27, 2020, in a model with detailed controls but excluding state fixed effects, a 10 percentage point increase in the proportion Black is associated with a large and statistically significant increase of 126.2 deaths per million residents. With state fixed effects included, the estimated disparity shrinks to a still sizeable 46.8 deaths per million but becomes statistically insignificant. The disparity for proportion Hispanic is not statistically significant in either model and the magnitudes are much smaller: 18.8 and 9.6 deaths per million in regressions without and with state fixed effects, respectively.

Desmet and Wacziarg (2020) examine both confirmed COVID-19 cases and fatalities. Outcomes are measured as logarithm of cases or fatalities (plus one) for May 26, 2020, where logarithm of population is included as a control,

<sup>1</sup> In work concurrent to ours, Figueroa et al. (2020) examine disparities in Massachusetts towns/cities through early May 2020 and found that a 10 percentage point increase in the Black population is associated with a 31.2 increase in COVID-19 cases per 10,000, while a 10 percentage point increase in the Hispanic population is associated with an increase of 25.8 cases per 10,000. Additional controls attenuate the association between the COVID-19 rate and Hispanic population—but not Black population.

implying the other estimates can be interpreted as the determinants of cases and deaths in per capita terms. They find highly significant and virtually identical disparities in cases and fatalities for proportion Black and somewhat smaller standardized coefficients for proportion Hispanic.

### Racial Segregation Within US Counties

There is somewhat limited racial and ethnic variation across counties in the USA, although the share of minority-majority counties has been increasing since 2000 (Krogstad 2019). Fewer than 5% of all US counties (151 out of 3143) have either Black, Hispanic, or indigenous people as the majority. As of 2018, there were 72 majority-Black counties, primarily located in the southeastern U.S.; there were 69 counties where Hispanics are the majority, predominantly in the southwestern U.S. (Schaeffer 2019). Especially in large US counties, there can be significant residential segregation, which may hide racial disparities in COVID-19 that can be more precisely measured at a more localized level. This limitation is well-understood; for example, McLaren (2020) notes “much of the relevant variation exists at the zip code level.” On a national level, there are 658 counties that contain 10 or more residential ZIP codes entirely within the county.<sup>2</sup> More than one-third of these counties have large dispersion—25 percentage points or more—in the share Black or Hispanic between the ZIP codes with the highest and lowest shares.

Neither the southeast nor southwest regions were the initial hotspots for COVID-19 spread, raising the question of whether observed correlations between county racial composition and COVID-19 death rates are confounded by the staggered timing in which the disease reached different parts of the country. The sensitivity of the results of Knittel and Ozaltun (2020) to the inclusion of state fixed effects is suggestive of such confounding. Yet including state fixed effects in county-level disparities regressions could itself be problematic, as it controls away much of the identifying variation, reducing the precision of the estimates. Accordingly, their coefficient for proportion Black becomes statistically insignificant after including state fixed effects despite remaining sizeable at 46.8 deaths per million. Estimating racial disparities with both credibility and precision therefore appears to require data with finer-grained geographic detail.

In our analysis of six cities, nearly 31% of ZIP codes we analyze were minority-majority “neighborhoods” (to borrow the terminology of Almagro and Orane-Hutchinson (2020)), who refer to ZIP codes within New York City as

neighborhoods). While analyses at even a more finely grained level than ZIP code (e.g., Census Tract) would be desirable, ZIP codes offer much more heterogeneity with respect to race and ethnicity than counties, while also having satisfactory indicators related to long-run economic opportunity, human mobility, and other key demographic and health-related information.

### Potential Explanations and Related Evidence

Several factors may explain racial and ethnic disparities in the spread of COVID-19 and subsequent COVID-19 outcomes. We next explain these theories and associated empirical evidence. Many of the relevant studies examine New York City in isolation, since it was initially the hardest hit in the USA, and because of its relatively early posting of ZIP code level data. Of course, New York City differs from the rest of the USA in many respects, even relative to other large cities, so the extent to which these findings are generalizable is unclear.

One possibility may be the nature of jobs. Many features about occupations, commuting, and the workplace could contribute to the spread of COVID-19, and may disproportionately affect people of color. Almagro and Orane-Hutchinson (2020) argue that occupation is a key explanatory variable for understanding the early transmission of COVID-19 in New York City, and since minority workers are more likely than others to be front-line employees or unable to work from home, they are more likely to be working outside the home during those hours. Accordingly, Coven and Gupta (2020) find that Black and Hispanic neighborhoods (measured at the ZIP code level) in New York City exhibited more daytime work activity than other neighborhoods during the pandemic using mobile location data sourced from VenPath. Selden and Berdahl (2020) examine the Medical Expenditure Panel Survey and find that Black adults in every age group were more likely than White adults to have health risks associated with severe COVID-19 illness, and that Blacks at high risk for severe illness were 1.6 times as likely as Whites to live in households containing health-sector workers.

Another phenomenon observed in New York City during the early stages of the pandemic was temporary relocation outside the city. Ability to relocate depends on whether one’s job can be done remotely and also whether one has the financial resources to do so. Perhaps for these reasons, Coven and Gupta (2020) find that the propensity to leave the city was strongly negatively associated with the proportion Black in the Census Tract.

Health care access may also be an important factor, and the barriers to receiving COVID-19 testing at the onset of

<sup>2</sup> Calculations conducted with a crosswalk between ZIP codes and counties from the Missouri Census Data Center, along with information on racial and ethnic composition from Social Explorer.

the pandemic are well known. However, Schmitt-Grohé et al. (2020) found that testing services were evenly shared across the income distribution in New York City's 177 ZIP codes. Borjas (2020) finds access to testing was roughly uniform across the share of the ZIP code that was minority, but the conditional probability of a positive test result was far greater in neighborhoods with larger Black populations.

It is also possible that residential segregation as measured at the ZIP code level, and the consequences from it, led to long-lasting effects on health that made minority communities particularly vulnerable to COVID-19. Link and Phelan (1995) argue that social factors are a fundamental cause of disease that, because they embody access to important resources, affect multiple disease outcomes through multiple mechanisms, and consequently maintain an association with disease even when intervening mechanisms change. Phelan and Link (2015) argue that racial inequalities in health endure primarily because racism is a fundamental cause of racial differences in socioeconomic status, and in turn socioeconomic status is a fundamental cause of health inequalities. Chetty et al. (2019) find Black Americans have much lower rates of upward mobility and higher rates of downward mobility than Whites, leading to persistent income disparities across generations. The Black-White gap persists even among boys who grow up in the same neighborhood. Logan and Parman (2018) demonstrate that premature mortality among Blacks is rooted in historical segregation. Using person-level data, the authors apply a comprehensive measure of segregation extending the analysis of structural factors in racial health disparities. Wiemers et al. (2020) highlight disparities in potential COVID-19 complications by constructing a vulnerability index from the Panel Study of Income Dynamics, finding that Blacks are drastically more vulnerable than other groups among people aged 45 and older. National estimates find that COVID-19 related hospitalizations among Blacks and Hispanics are more than four times that of Whites (CDC, U. S. 2020b). Price-Haywood et al. (2020) found that more than 70% of patients who were hospitalized or died of COVID-19 were Black, compared to an overall population representation of 31%.

## Data

To analyze racial and ethnic disparities, we combine ZIP code level data on COVID-19 outcomes from state and local government websites with data from (1) the 2018 ACS 5-year sample, (2) the 2010 Census, (3) the Opportunity Atlas, (4) SafeGraph mobility data, (5) health professional shortage areas published by the Health Resources & Services Administration, and (6) conditional life expectancy published by the Centers for Disease Control and Prevention.

## COVID-19 Data

From state and local websites, we gathered COVID-19 data at the ZIP code level for six metropolitan areas: New York City, Chicago, Atlanta, San Diego, St. Louis, and Baltimore.<sup>3</sup> The ZIP codes include both the city proper, and in some cases the surrounding county. Cross-sectional data was gathered from June 6, 2020, to June 9, 2020, for these localities.<sup>4</sup> Once merged with all other sources, the full analysis uses 436 ZIP codes, with 177 in New York City, 58 in Chicago, 49 in Atlanta, 95 in San Diego, 21 in St. Louis, and 36 in Baltimore. Overall, there are approximately 17.7 million people living in these ZIP codes, with nearly half residing in New York City.

Our primary outcome variable is confirmed COVID-19 cases per 10,000 population. Although serological surveys provide strong evidence that confirmed cases are an undercount of total infections, confirmed case numbers still have clear clinical and economic significance. Nationally, the fatality and hospitalization rates for confirmed cases were roughly 5% and 10%, respectively, by June 2020.<sup>5</sup> Even after discharge from a hospital, persistent symptoms may remain (Carfi et al. 2020). In addition, confirmed infections (which tend to be more severe than those that remain undetected) undoubtedly lead to lost earnings, family strain, psychological distress, and potentially harmful long-term consequences (Eisenberg et al. Forthcoming). Kniesner and Sullivan (Forthcoming) estimate economic losses from COVID-19 at \$46,000 per non-fatal case, by applying value per statistical

<sup>3</sup> The ZIP code data encompass different geographies, based on what data were available at the time. For New York City, the data covered all of the city's 177 ZIP codes across all five boroughs. For Chicago, the data included all 60 ZIP codes in the city proper, but not other parts of Cook County or DuPage County. For Atlanta, there were 59 ZIP codes within Fulton County; the data did not include the DeKalb county portion of the city or the portions of Fulton county outside the city limits. For San Diego, there were 100 ZIP codes across all of San Diego County. For St. Louis, the data included all 22 ZIP codes in St. Louis City, but not the surrounding St. Louis County. For Baltimore, there were initially 40 ZIP codes covering parts of Baltimore City, Baltimore County, and Anne Arundel County.

<sup>4</sup> Chicago, Baltimore, and Atlanta archive their COVID-19 reports, while the other cities (New York, San Diego, and St. Louis) only retain current data. Publicly available ZIP-code-level COVID-19 data can be found from the following URLs (accessed September 27, 2020):

Atlanta: <https://www.fultoncountyga.gov/covid-19/epidemiology-reports>

Baltimore: <https://coronavirus.maryland.gov/datasets/mdcovid19-master-zip-code-cases/data>

Chicago: <https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v/data>

New York City: <https://github.com/nychealth/coronavirus-data>

San Diego: <https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/Epidemiology/COVID-19%20Summary%20of%20Cases%20by%20Zip%20Code.pdf>

St. Louis: <https://www.stlouis-mo.gov/government/departments/health/communicable-disease/covid-19/data/zip.cfm>

<sup>5</sup> University of Minnesota. COVID-19 Hospitalization Tracking Project. <https://carlsonschool.umn.edu/mili-misrc-covid19-tracking-project>. Accessed June 16, 2020

life and relative severity/injury estimates from the Department of Transportation.

All six localities provided counts of COVID-19 cases, which are scaled into counts per 10,000 population. When weighted by population, Table 1 shows the median ZIP code had 143 cumulative cases per 10,000 population by early June, translating into a cumulative measured infection rate of about 1.4%. Measured infection rates varied substantially, with 12 cases per 10,000 (0.1%) in the lowest decile and 315 cases per 10,000 (3.2%) in the highest decile. In the aggregate, New York City had the highest rate of confirmed cases (2.3%), followed by Chicago (1.7%) and Baltimore (0.9%). The other three localities had confirmed case rates varying from 0.26 to 0.48%. In the aggregate, there were more than 271,000 confirmed COVID-19 cases in these 6 cities, approximately 14% of all cases nationally by that point.<sup>6</sup>

We also conduct auxiliary analyses using a subsample of only Chicago and New York City, cities that provide additional data that allow us to investigate two important issues. First, do observed disparities in confirmed COVID-19 cases accurately reflect disparities in illnesses, or are they confounded by geographic variation in availability of tests and criteria for obtaining them? Note that we will include city fixed effects in all our models, which alleviates this concern to some extent. Nonetheless, since Chicago and New York City report tests run by ZIP code, they enable us to control for testing more directly. Second, are racial and ethnic disparities in COVID-19 fatalities the result of a higher likelihood of catching the virus, a greater risk of dying conditional on catching it, or some combination of both? Answering this question requires data on COVID-19 fatalities—not just cases—and Chicago and New York City are the only cities in our sample that report deaths by ZIP code. The bottom panel of Table 1 shows data for those two cities (and 235 ZIP codes). In these cities, the cumulative fatality rate from COVID-19 was 0.17% by early June.

### Census Bureau Data

We merged this information to the 2018 ACS 5-year sample, as well as to the 2010 Decennial Census using Social Explorer (which provides summary statistics at the ZIP code level).<sup>7</sup> The ACS contains a rich set of variables on demographics, economic outcomes, and housing characteristics. Demographic variables include percent male, percent foreign born, and percent aged 18–44, 45–64, 65–74, and 75+ (children under 18 are omitted). Housing variables include density, percent renters, percent vacant units, percent overcrowded

**Table 1** Summary statistics

	(1) COVID-19 cases per 10k pop	(2) COVID-19 fatalities per 1 m pop	(3) COVID-19 tests per 10k pop	(4) Percent Black	(5) Percent Hispanic	(6) Percent other race
Mean	153	–	–	23.6	25.9	13.8
SD	113	–	–	27.7	21.4	12.9
10th percentile	12	–	–	1.6	4.3	2.6
50th percentile	143	–	–	10.1	17.9	9.4
90th percentile	315	–	–	74.0	63.0	34.3
N	436	436	436	436	436	436
Mean	219	1727	898	23.9	29.1	14.7
SD	87	964	210	27.1	21.7	14.3
10th percentile	103	620	637	1.6	6.9	2.4
50th percentile	217	1653	880	12.6	21.3	9.4
90th percentile	336	2898	1164	69.1	66.6	38.0
N	235	235	235	235	235	235

Top panel includes all 6 cities; bottom panel includes only Chicago and New York City. COVID-19 statistics measured between June 6, 2020, and June 9, 2020. Race/ethnicity variables obtained from 2018 ACS 5-year sample

(1.5 or more persons per bedroom), and percent of units with 0 or 1 bedrooms. The 2010 Census—although dated—provides information about group quarters, specifically percent of population in nursing homes, correctional facilities, college dormitories, and military barracks. Returning to the ACS 5-year sample, our socioeconomic variables include percent in education bins (dropout, high school, some college, bachelor’s degree, the group beyond a bachelor’s degree is omitted as a reference category), income inequality as measured by the Gini coefficient, and percent in poverty bins (0–49% FPL, 50–74%, 75–99%, 100–149%, 150–199%, 200%+ is omitted). Occupation variables include percent of workers in service occupations, sales, farming, construction, production, or transport (managerial occupations omitted). Transportation variables include percent of workers of workers who use a car, percent who use public transportation, and percent with long commuting times (60+ min). Finally, one of our measures of health access—percent without health insurance—comes from the ACS.

<sup>6</sup> <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>

<sup>7</sup> <https://www.socialexplorer.com/>

## Opportunity Atlas Data

The Opportunity Atlas is a collaboration of the Census Bureau, Harvard University, and Brown University that uses anonymous data following 20 million Americans from childhood to their mid-30s, with many outcomes measured at the Census Tract level (which we aggregate up to ZIP code).<sup>8</sup> As noted by Chetty et al. (2018), neighborhoods matter at a very granular level, where neighborhoods even one mile away have very little predictive power for child outcomes. We focus on two key variables, which represent long-run opportunity. The first is average annual household income ranking in 2014–2015 for children (in their mid-30s) who grew up in the area, based on having a low-income parent (25th percentile). The second is fraction of male children who grew up in the area who were in prison or jail on April 1, 2010. We aggregate Census Tracts to the ZIP code level using a crosswalk provided by the Missouri Census Data Center.<sup>9</sup> We follow the spirit of Courtemanche et al. (2017) by assigning Census Tract to the ZIP code where the plurality of residents live. In practice, approximately 53% of Tracts nationally map into one ZIP code only, and roughly 75% of Tracts have at least 80% of their population in one ZIP code.

## SafeGraph Data

Many recent COVID-19 studies examine mobility using data from SafeGraph, which provides access to their data through free, non-commercial agreements.<sup>10</sup> Following Gupta et al. (2020), we compute the fraction of cell phone devices that were detected to be entirely at home during the day, aggregating from the Census Block Group level to the ZIP code level. We aggregate Census Block Groups to ZIP codes using a crosswalk provided by the Missouri Census Data Center. We computed daily averages for each ZIP code, and then averaged across all days for the months of March 2020, April 2020, and May 2020. Again, following Courtemanche et al. (2017), we assign Census Block Groups to the ZIP code where the plurality of residents live. In practice, approximately 73% of Census Block Groups nationally map into one ZIP code only, and roughly 85% of Tracts have at least 80% of their population in one ZIP code.

## Health Professional Shortage Area Data

We incorporate information on each ZIP code's status as being designated as a health professional shortage area (HPSA) for federal fiscal year 2020.<sup>11</sup> HPSAs are designated by the

Health Resources and Services Administration (HRSA) to signify areas as medically underserved. The Centers for Medicare and Medicaid Services (CMS) provide HPSA designation status at the ZIP code level to signal to eligible health care professionals (e.g., physicians) if the location where they practice medical care is eligible for enhanced Medicare reimbursements per the 2005 Medicare Modernization Act (CMS, U.S. Centers for Medicare, and Medicaid Services 2020). This feature of the program creates a financial incentive for delivering care in medically underserved settings with historically higher uninsured rates and limited access to care. We use this as a proxy measure to capture differences in access to primary care and mental health services. HPSAs can be entire counties, but are most commonly smaller portions of a county—this is particularly the case in larger cities.

## Centers for Disease Control and Prevention Data

Our population health variables are conditional life expectancies obtained from the U.S. Small-area Life Expectancy Estimates Project (USALEEP).<sup>12</sup> The files contain conditional life expectancies for different age bins at the Census Tract level; in our model, we include conditional life expectancies for ages 65–74, 75–84, and 85 plus, and aggregate from the Tract level to ZIP code. Many commentators attribute disparities in COVID-19 cases and deaths to underlying health conditions such as elevated rates of chronic illnesses among Blacks and Latinos (Artiga et al. 2020). We use variation in life expectancy—and focus on the elderly who are most vulnerable to COVID-19—to control for variation in underlying health status as well as the risk factors leading to differences in preventable mortality.

Table 2 shows, along some margins, large differences in neighborhood characteristics depending on racial and ethnic composition. Out of the 436 ZIP codes, 188 are majority White, 84 are majority Black, 49 are majority Hispanic, and 115 are none of these. With respect to demographics, Hispanic neighborhoods have much higher representation of foreign-born individuals. With respect to housing, there are more renters in majority-Black and majority-Hispanic neighborhoods. Lower educational attainment and higher poverty levels are also common attributes of these neighborhoods, as are lower levels of income mobility—an indicator of long-run opportunity. At least some types of workers whose jobs do not easily transfer online—those in service occupations—are more prevalent in predominantly Black and Hispanic neighborhoods. Also common in predominantly Black and Hispanic neighborhoods are larger dependence on public transit as a key mode of transportation (McLaren 2020) and longer commuting times. Cell phone mobility measures are relatively similar, on average, across neighborhoods. Health care access

<sup>8</sup> [https://opportunityinsights.org/data/?geographic\\_level=99&topic=0&paper\\_id=1652#resource-listing](https://opportunityinsights.org/data/?geographic_level=99&topic=0&paper_id=1652#resource-listing)

<sup>9</sup> <http://mcdc.missouri.edu/applications/geocorr2018.html>

<sup>10</sup> <https://www.safegraph.com/covid-19-data-consortium>

<sup>11</sup> <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HPSAPSAPhysicianBonuses>

<sup>12</sup> <https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html>

**Table 2** Summary statistics by neighborhood type

	Majority White	Majority Black	Majority Hispanic	All other ZIPs
Confirmed COVID-19 case per 10k pop	99.9 (6.6)	171.7 (11.1)	237.4 (16.7)	161.9 (10.3)
% White	64.9 (0.7)	10.3 (1.1)	12.7 (1.6)	31.4 (1.4)
% Black	7.0 (0.5)	74.6 (1.6)	15.2 (1.6)	15.9 (1.4)
% Hispanic	14.3 (0.6)	10.5 (1.2)	64.7 (1.3)	28.4 (1.3)
% other	13.8 (0.5)	4.6 (0.4)	7.4 (0.8)	24.3 (1.7)
Demographic controls				
% age 18–44	41.5 (0.8)	38.6 (0.6)	41.4 (0.5)	41.2 (0.6)
% age 45–64	24.7 (0.3)	25.0 (0.3)	22.7 (0.3)	24.6 (0.3)
% age 65–74	8.3 (0.2)	7.7 (0.2)	6.3 (0.2)	7.3 (0.2)
% age 75+	6.4 (0.2)	5.8 (0.2)	4.5 (0.2)	5.6 (0.2)
% foreign-born	21.8 (0.8)	18.7 (1.8)	37.3 (1.3)	34.9 (1.4)
% male	48.8 (0.2)	45.7 (0.3)	49.2 (0.3)	49.0 (0.2)
Housing controls				
Density	25,835 (2448)	22,888 (2685)	40,125 (4724)	25,712.6 (2103.1)
% college group quarters	1.0 (0.2)	0.6 (0.3)	0.1 (0.1)	0.7 (0.3)
% correctional group quarters	0.2 (0.1)	0.5 (0.3)	0.6 (0.3)	0.3 (0.2)
% military group quarters	0.3 (0.3)	0.0 (0.0)	0.4 (0.3)	0.1 (0.1)
% nursing home group quarters	0.5 (0.0)	0.5 (0.1)	0.4 (0.1)	0.5 (0.1)
% unit is 0 or 1 bedroom	28.2 (1.4)	22.0 (1.3)	28.6 (2.1)	25.2 (1.4)
% 1.5+ occupants per bedroom	1.8 (0.1)	1.5 (0.1)	4.7 (0.4)	2.8 (0.2)
% renter	49.6 (1.4)	59.5 (1.9)	71.4 (2.9)	56.9 (1.7)
% vacant	9.3 (0.4)	14.0 (0.9)	7.2 (0.5)	8.1 (0.3)
Socioeconomic				
Gini coefficient	47.5 (0.4)	48.2 (0.6)	46.2 (0.6)	46.2 (0.5)
% HS dropout	8.6 (0.5)	15.8 (0.5)	30.0 (0.9)	17.5 (0.7)
% HS graduate/GED	17.0 (0.7)	30.2 (0.6)	27.5 (0.8)	23.9 (0.7)
% Some college	20.9 (0.6)	28.8 (0.5)	23.9 (0.7)	23.5 (0.6)
% Bachelor's degree	30.0 (0.7)	15.3 (0.6)	12.6 (0.7)	21.8 (0.6)
% 0–49 FPL	5.5	11.1	10.1	7.5

**Table 2** (continued)

	Majority White	Majority Black	Majority Hispanic	All other ZIPs
	(0.2)	(0.6)	(0.7)	(0.4)
% 50–74 FPL	2.7 (0.2)	5.8 (0.3)	7.0 (0.4)	4.0 (0.2)
% 75–99 FPL	3.5 (0.2)	6.1 (0.2)	7.8 (0.3)	5.1 (0.2)
% 100–149 FPL	6.3 (0.2)	11.0 (0.3)	14.0 (0.4)	10.0 (0.3)
% 150–199 FPL	6.1 (0.2)	10.1 (0.2)	12.1 (0.3)	9.1 (0.2)
Opportunity Atlas controls				
Income mobility	44.7 (0.5)	33.7 (0.5)	39.6 (0.5)	44.0 (0.7)
Male incarceration	2.9 (0.2)	8.3 (0.4)	3.5 (0.3)	3.0 (0.2)
Occupational controls				
% sales occupation	20.9 (0.2)	22.4 (0.3)	21.0 (0.4)	21.6 (0.2)
% service	14.3 (0.4)	26.4 (0.6)	30.9 (1.0)	22.1 (0.5)
% construction	4.8 (0.3)	5.3 (0.2)	9.0 (0.5)	6.8 (0.3)
% transport	4.4 (0.2)	10.3 (0.3)	10.1 (0.4)	7.2 (0.3)
% production	2.4 (0.1)	3.2 (0.2)	5.7 (0.5)	3.6 (0.2)
% farming	0.1 (0.0)	0.1 (0.0)	0.3 (0.1)	0.2 (0.0)
Transportation controls				
% travel by car	56.9 (2.3)	53.2 (2.5)	46.2 (4.4)	51.7 (2.6)
% travel by public transit	26.1 (1.8)	37.5 (2.4)	42.5 (4)	36.4 (2.4)
% 60+ min commute	12.6 (0.8)	23.1 (1.2)	22.4 (1.5)	20.2 (1.1)
SafeGraph mobility controls				
% home all day, 3/2020	33.8 (0.3)	35.1 (0.4)	37.3 (0.5)	37.2 (0.4)
% home all day, 4/2020	47.4 (0.5)	46.9 (0.7)	51.0 (0.9)	53.2 (0.7)
% home all day, 5/2020	43.8 (0.5)	43.3 (0.6)	45.9 (0.8)	48.6 (0.7)
Health access controls				
% uninsured	6.2 (0.2)	9.8 (0.4)	13.4 (0.6)	10.0 (0.4)
Mental health HPSA	0.9 (0.7)	13.8 (3.8)	12.4 (4.8)	1.0 (0.9)
Primary care HPSA	0.0 (0.1)	4.7 (2.3)	0.0 (0.3)	1.0 (0.9)
Population health controls				
Life expectancy, age 65–74	20.3 (0.2)	18.9 (0.2)	19.9 (0.2)	20.1 (0.3)
Life expectancy, age 75–84	13.0 (0.2)	12.5 (0.1)	13.0 (0.2)	13.0 (0.2)
Life expectancy, age 85+	7.3	7.4	7.5	7.4



**Table 2** (continued)

	Majority White	Majority Black	Majority Hispanic	All other ZIPs
	(0.1)	(0.1)	(0.2)	(0.2)
Atlanta zip code	11.4 (2.3)	15.1 (3.9)	0.0 (0.0)	7.1 (2.4)
Baltimore zip code	8.9 (2.1)	14.9 (3.9)	0.0 (0.0)	2.4 (1.4)
Chicago zip code	12.2 (2.4)	22.9 (4.6)	21.8 (6.0)	11.6 (3.0)
New York City zip code	39.5 (3.6)	41.6 (5.4)	55 (7.2)	58.4 (4.6)
San Diego zip code	25.3 (3.2)	0.0 (0.0)	23.2 (6.1)	19.4 (3.7)
St. Louis zip code	2.8 (1.2)	5.5 (2.5)	0.0 (0.0)	1.1 (1.0)
Number of ZIP Codes	188	84	49	115

Standard errors in parentheses. There are 436 ZIP codes from 6 cities displayed. All summary statistics are weighted by population

is worse for Black and Hispanic neighborhoods according to both percent uninsured and mental health HPSA, while population health—proxied by conditional life expectancy—is fairly similar across neighborhood types, especially from age 75 onward. Finally, racial composition and segregation varies by city. None of the ZIP codes in San Diego are majority Black, while Atlanta, Baltimore, and St. Louis have no ZIP codes that are majority Hispanic.

## Empirical Model and Findings

### Model

We estimate linear models of the following form:

$$\begin{aligned} covid_{z,j} = & \beta_0 + \beta_1 PCT\_BLACK_{z,j} \\ & + \beta_2 PCT\_HISPANIC_{z,j} + \beta_3 OTHER_{z,j} \\ & + \beta_4 X_{z,j} + \delta_j + \varepsilon_{z,j} \end{aligned} \tag{1}$$

where  $covid_{z,j}$  represents either confirmed COVID-19 cases per 10,000 population or COVID-19 fatalities per 1,000,000 population in ZIP code  $z$  in city  $j$ . The key explanatory variables include the percentage Black and percentage Hispanic in each ZIP code (we also include percentage other race, with percentage White omitted).<sup>13</sup> We successively include additional neighborhood characteristics in  $X_{z,j}$ . City fixed effects

<sup>13</sup> In the subsequent tables, we focus attention on percent black and percent Hispanic. The other race category includes several groups, some of which are not traditionally disadvantaged. Appendix Table 8 shows all the coefficients from our base model and full model.

are given by  $\delta_j$  and  $\varepsilon_{z,j}$  is the error term. All observations are weighted by population in the ZIP code. Standard errors are heteroscedasticity-robust.<sup>14</sup>

Any observational analysis like this identifies correlations, not causation (Knittel and Ozaltun 2020). With many observable neighborhood differences—including some characteristics that intuitively might represent greater likelihood of COVID-19 transmission such as density, occupations, modes of transport, or health care access—it may be tempting to attach causal stories. However, these variables could influence each other or be influenced by unobservable characteristics of the ZIP code, preventing causal inference. Our principal goal, therefore, is simply to assess the extent to which such measured characteristics can collectively explain the racial and ethnic disparities in COVID-19 burden, and the extent to which these disparities remain unexplained by conventional measures.

Our model’s city fixed effects account for some potential confounders, such the arrival of the virus, weather patterns, and lockdown policies, but not unobserved heterogeneity at the sub-city level. A number of recent studies control for unobserved heterogeneity at granular levels (Schuetz et al. 2008; Price 2013), recognizing that micro-geographies are an important determinant of individual health outcomes (Arcaya et al. 2016). It is not possible to include ZIP code fixed effects (or

<sup>14</sup> We also ran regressions clustering the standard errors at the city level, and this made the standard errors implausibly small in some cases. Since the number of clusters (6) is too small for cluster-robust inference to be reliable, we err on the side of caution and report the (larger) standard errors obtained without clustering. Also note that, since we do not include any city-level variables, there is no particular reason to think that standard errors obtained without clustering are systematically flawed (Cameron and Miller 2015).

fixed effects at a narrower level such as Census Tract) in our models because they would be perfectly collinear with the ZIP code-level covariates such as race and ethnicity.<sup>15</sup>

## Size of Disparities

Our primary results are presented in Table 3, which examines confirmed COVID-19 infections per 10,000 population. In the top panel, we examine 436 ZIP codes across the 6 cities, and successively include covariates for demographics, housing, socioeconomic status, opportunity, occupation, transportation, human mobility, health access, and population health. The base model in column (1)—which only includes the race/ethnicity variables, city fixed effects, and a constant term—shows large, highly significant health disparities for both Blacks and Hispanics. A 10 percentage point increase in a ZIP code's Black share is associated with 9.2 additional confirmed COVID-19 cases per 10,000 population, while a similar change in the Hispanic share is associated with 20.6 additional COVID-19 cases. Both are very sizable changes relative to the average confirmed cases rate of 153 per 10,000 population.

As the analysis moves from one column to the next, sets of covariates are successively included. For example, including “demographics” in column (2) adds six additional variables for percent male, percent foreign, and percent in each of four age bins to the specification. Across the remaining columns, the measured disparity remains sizable and significant, regardless of the set of controls that are included. The full model in column (10) shows disparities that are roughly 60% as large as in column (1). Thus, a key insight is that even with an extensive set of controls for factors that should plausibly affect the transmission of the virus, more than half of the overall disparity in confirmed COVID-19 cases remains unexplained.

In the bottom panel, we focus on the 235 ZIP codes in Chicago and New York, since those localities provide additional data on COVID-19. The initial overall disparities in column (1) are larger, and a 10 percentage point increase in a ZIP code's Black share is associated with 12.4 additional confirmed COVID-19 cases per 10,000 population, while a similar change in the Hispanic share is associated with 24.8

additional COVID-19 cases. Again, both are sizable changes relative to the average confirmed cases rate of 219 per 10,000 population for these two cities. As neighborhood characteristics are added in the remaining columns, in some instances, the disparity falls and in others it rises. With the full set of controls in column (10), there remain sizable disparities for both Blacks and Hispanics, and again more than half of the overall disparity remains unexplained. It should be noted that the “health access” variables in column (9) now include COVID-19 tests per capita at the ZIP code level.

Next, we turn to Table 4, where we examine COVID-19 fatalities at the ZIP code level for Chicago and New York, an outcome that was a key focus in McLaren (2020) and Knittel and Ozaltun (2020). To maintain comparability, we scale COVID-19 fatalities per million population. In column (1), we estimate models with race/ethnicity controls, city fixed effects, and a constant term, finding large and statistically significant disparities. A 10 percentage point increase in the Black (Hispanic) share is associated with 143 (149) additional fatalities per million, from a baseline of 1727 per million in these two cities. In column (2), we include as an additional control COVID-19 cases (per 10,000) in each ZIP code. Thus, we ask the extent to which variation in fatalities is simply explained by greater numbers of confirmed COVID-19 cases, versus the extent to which factors beyond additional cases matter. The results show that the coefficients fall by 50–100%, suggesting that racial and ethnic disparities in the spread of infection are an extremely important determinant for resultant fatalities. For proportion Hispanic, the coefficient estimate is very close to zero, while for proportion Black, the coefficient estimate is about half as large. In columns (3) and onward, additional covariates similar to the ones in the bottom panel of Table 3 are added. With sufficient controls, the original racial and ethnic disparities become insignificant.

## Mechanisms

In a highly cited paper, Gelbach (2016) shows that the sequence in which covariates are entered into the model can lead to very different conclusions about their relative importance. In our context, the extent to which different sets of neighborhood characteristics “explain” the observed racial and ethnic disparities may be sensitive to the sequence in which they are entered into the model. For example, the changes in racial and ethnic disparities from adding demographic covariates may have been different than those shown in Tables 3 and 4 if we had added housing variables first, since demographic and housing variables are correlated. In order to create a path-independent explanation of the influence of each set of neighborhood characteristics, Gelbach prescribes omitted variable bias equations. Essentially, we estimate omitted variable bias on the coefficient of interest from the exclusion of each sets of neighborhood controls one at a time from the full model (in

<sup>15</sup> We considered controlling for geography with Census Public Use Microdata Area (“PUMA”) fixed effects instead of city fixed effects. PUMAs are geographically contiguous areas of at least 100,000 people; our 436 ZIP codes across the 6 cities map into 130 PUMAs. PUMA fixed effects are intuitively appealing since they avoid perfect collinearity while capturing unobserved heterogeneity at a much more detailed geographic level than the city. Unfortunately, in practice, PUMA lines often coincide with racial and ethnic divides of a city, and there is insufficient within-PUMA variation to credibly estimate disparities. Specifically, PUMA fixed effects and the ZIP code-level controls explain over 97% of the variation in a ZIP code's percent Black and Hispanic, which is far beyond typical rules of thumb for severe multicollinearity.

**Table 3** COVID-19 cases per 10k population

6 cities (Atlanta, Baltimore, Chicago, New York City, San Diego, St. Louis), <i>N</i> = 436 ZIP codes										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
% Black	0.92*** (0.13)	0.94*** (0.13)	0.97*** (0.16)	0.86*** (0.19)	0.85*** (0.21)	0.54** (0.23)	0.65*** (0.21)	0.76*** (0.21)	0.72*** (0.22)	0.59*** (0.23)
% Hispanic	2.06*** (0.21)	1.29*** (0.23)	1.67*** (0.27)	1.51*** (0.38)	1.53*** (0.38)	1.26*** (0.38)	1.42*** (0.35)	1.32*** (0.35)	1.29*** (0.34)	1.22*** (0.33)
Adj <i>R</i> <sup>2</sup>	0.76	0.81	0.85	0.85	0.85	0.86	0.87	0.88	0.88	0.88
2 cities (Chicago, New York City), <i>N</i> = 235 ZIP codes										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
% Black	1.24*** (0.19)	1.10*** (0.21)	1.45*** (0.22)	1.18*** (0.30)	0.91*** (0.34)	0.89** (0.36)	1.16*** (0.35)	1.32*** (0.34)	1.26*** (0.28)	1.26*** (0.32)
% Hispanic	2.48*** (0.26)	1.47*** (0.28)	2.31*** (0.32)	1.81*** (0.53)	1.71*** (0.54)	1.65*** (0.54)	1.84*** (0.54)	1.76*** (0.53)	1.31*** (0.35)	1.29*** (0.38)
Adj <i>R</i> <sup>2</sup>	0.44	0.63	0.72	0.73	0.73	0.73	0.75	0.75	0.91	0.91
Demographics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opportunity Atlas	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Occupational	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Transportation	No	No	No	No	No	No	Yes	Yes	Yes	Yes
SafeGraph	No	No	No	No	No	No	No	Yes	Yes	Yes
Health access	No	No	No	No	No	No	No	No	Yes	Yes
Population health	No	No	No	No	No	No	No	No	No	Yes

All specifications include city fixed effects, “% Other non-white”, and a constant term. Unit of observation is ZIP code. *Demographic variables* include percentage in ZIP code who are male, foreign born, or in age bins (18–44, 45–64, 65–74, 75+). *Housing variables* include density, percentage who are renters, percent of units vacant, percentage who are in nursing homes, correctional facilities, college dorms, or military barracks (2010 Census), percent overcrowded (1.5+ per room), and percent with 0 or 1 bedroom sizes. *Socioeconomic variables* include percent in education bins (dropout, high school, some college, bachelor’s degree), Gini coefficient, percent in poverty bins (0–49% FPL, 50–74%, 75–99%, 100–149%, 150–199%). *Opportunity Atlas variables* include income mobility and male incarceration (Opportunity Atlas). *Occupation variables* include percent of workers in service occupations, sales, farming, construction, production, or transport. *Transportation variables* include percent of workers of workers who use a car, percent who use public transportation, and percent with long commuting times (60+ minutes). *Safegraph variables* include percent who on average remained at home all day in each month from March to May 2020 (Safegraph). *Health access variables* include health professional shortage areas (HRSA; mental health, primary care), percent without health insurance, and COVID-19 tests per capita (Chicago and New York City only). *Population health variables* include conditional life expectancy (CDC, ages 65–74, 75–84, 85+). All control variables obtained from 2018 ACS 5-year sample unless otherwise indicated. All regressions weighted by ZIP code population from 2018 ACS 5-year sample. Heteroscedasticity-robust standard errors in parentheses. \*\*\* *p* < 0.01, \*\* *p* < 0.05, \* *p* < 0.10

Table 3, column 10). The influence of each set of neighborhood characteristics therefore becomes a function of the correlation between the covariate and race/ethnicity in addition to the covariate’s coefficient in the full specification.

Table 5 shows the results of the Gelbach decomposition for our full set of 6 cities on COVID-19 cases (Table 3, top panel), as well for the 2 cities on cases and fatalities (Table 3, bottom panel; Table 4).<sup>16</sup> The baseline coefficient comes from the model that only includes racial/ethnic composition, city fixed effects, and a constant term. The “explained difference” is the reduction that occurs from moving from the first column to the last column (e.g., from 0.92 to 0.59 for “% Black” in the top panel of Table 3), and the remaining rows show the contribution of each set of neighborhood characteristics to the explained reduction (as

well as their statistical significance). In the model of COVID-19 cases with all 6 cities in column (1), much of the explained difference in the disparity for Blacks can be attributed to the long run opportunity variables (income mobility and male incarceration rates for those born between 1978 and 1983). Although other sets of neighborhood characteristics occasionally have large magnitudes (e.g., the occupation controls), they are not statistically significant. For the Hispanic disparity, the opportunity variables again contribute to part of the explained difference; however, demographics (which includes the fraction foreign born) and human mobility (from SafeGraph) are larger factors.

The findings, when restricted to Chicago and New York, are somewhat different in for cases and fatalities (columns 2 and 3 of Table 5, respectively). For confirmed COVID-19 cases, the combined addition of all the covariates does little to change the estimate for proportion Black; however, several

<sup>16</sup> This was implemented with the “b1x2” command in Stata.

**Table 4** COVID-19 fatalities per 1m population

	2 cities (Chicago, New York City), $N=235$ ZIP codes										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
% Black	14.3*** (2.0)	6.7*** (2.0)	5.8*** (2.0)	3.9** (2.2)	7.3** (3.1)	2.7 (4.0)	4.6 (4.2)	2.0 (4.2)	3.6 (4.2)	1.6 (4.5)	-0.1 (4.5)
% Hispanic	14.9*** (2.6)	-0.4 (2.9)	-0.4 (2.7)	-4.4 (2.7)	3.8 (3.8)	1.4 (4.0)	2.6 (4.5)	0.9 (4.2)	1.2 (4.3)	-0.1 (4.3)	-0.5 (4.3)
Adj $R^2$	0.43	0.61	0.69	0.79	0.81	0.81	0.83	0.84	0.84	0.84	0.84
Cases per 100k Pop	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opportunity Atlas	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Occupational	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Transportation	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
SafeGraph	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Health access	No	No	No	No	No	No	No	No	No	Yes	Yes
Population health	No	No	No	No	No	No	No	No	No	No	Yes

All specifications include city fixed effects, “% Other non-white”, and a constant term. Unit of observation is ZIP code. *Demographic variables* include percentage in ZIP code who are male, foreign born, or in age bins (18–44, 45–64, 65–74, 75+). *Housing variables* include density, percentage who are renters, percent of units vacant, percentage who are in nursing homes, correctional facilities, college dorms, or military barracks (2010 Census), percent overcrowded (1.5+ per room), and percent with 0 or 1 bedroom sizes. *Socioeconomic variables* include percent in education bins (dropout, high school, some college, bachelor’s degree), Gini coefficient, and percent in poverty bins (0–49% FPL, 50–74%, 75–99%, 100–149%, 150–199%). *Opportunity Atlas variables* include income mobility and male incarceration (Opportunity Atlas). *Occupation variables* include percent of workers in service occupations, sales, farming, construction, production, or transport. *Transportation variables* include percent of workers of workers who use a car, percent who use public transportation, and percent with long commuting times (60+ minutes). *Safegraph variables* include percent who on average remained at home all day in each month from March to May 2020 (Safegraph). *Health access variables* include health professional shortage areas (HRSA; mental health, primary care), percent without health insurance, and COVID-19 tests per capita (Chicago and New York City only). *Population health variables* include conditional life expectancy (CDC, ages 65–74, 75–84, 85+). All control variables obtained from 2018 ACS 5-year sample unless otherwise indicated. All regressions weighted by ZIP code population from 2018 ACS 5-year sample. Heteroscedasticity-robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

factors offset each other to result in essentially the zero change. Controlling for health care access and human mobility is associated with statistically significant reductions in the disparity, but this is offset by statistically significant increases in the disparity from housing and socioeconomic characteristics and insignificant but similarly sized increases from occupation and transportation. For proportion Hispanic, the key driver explaining much of the reduction in confirmed cases appears to be health care access, which in the context of the two cities includes COVID-19 tests per capita.<sup>17</sup>

Finally, for COVID-19 fatalities (column 3 of Table 5), adding covariates from the base model to the full model entirely explains the disparities for proportion Black and Hispanic. The most important factor in explaining COVID-19 fatalities is simply the rate of infection. Differences in confirmed cases explain approximately two-thirds of the disparity in fatalities for Blacks, and the entire disparity for Hispanics. Other factors appear to have large and somewhat

offsetting effects for both groups (e.g., socioeconomic controls and occupational controls), but a key implication is that understanding the core causes of COVID-19 cases can potentially explain the alarming subsequent differences in mortality.

## Robustness Checks

We explore several robustness checks in Tables 6 and 7. As illustrated in Table 2, one key advantage of analyzing ZIP codes rather than counties is that residential segregation is much more stark. Overall, 327 of the 436 ZIP codes are highly segregated, and we re-estimate the models of COVID-19 cases restricted to neighborhoods that are the majority of one race or ethnicity in the top panel of Table 6. Overall, the findings—both from the base model and full model—are very similar to Table 3. Next, we test the sensitivity of the results to excluding one city at a time in Table 7, presenting findings from the base specification and full specification. In all cases, there are sizable racial and ethnic disparities in the base specification. The findings on Hispanic disparities are robust across all specifications; there are large disparities that are partially explained by neighborhood characteristics.

<sup>17</sup> Appendix Table 8 shows that among health access variables for New York and Chicago, tests per capita is highly significant for confirmed cases.

**Table 5** Accounting for change in coefficients

Coefficient		(1)	(2)	(3)
		COVID-19 cases per 10k population 6 cities	COVID-19 cases per 10k population 2 cities	COVID-19 fatalities per 1m population 2 cities
% Black	Baseline Coefficient	0.92*** (0.13)	1.24*** (0.19)	14.3*** (2.0)
	Explained difference	0.33 (0.24)	-0.03 (0.36)	14.4*** (4.9)
	Cases per 100k Pop	N/A	N/A	9.3*** (1.9)
	Demographics	-0.06 (0.09)	0.09 (0.13)	1.3 (1.9)
	Housing	-0.14 (0.12)	-0.25** (0.11)	2.6 (1.8)
	Socioeconomic	-0.23 (0.34)	-0.71* (0.40)	-15.0** (6.2)
	Opportunity Atlas	0.39*** (0.13)	0.09 (0.11)	1.1 (1.9)
	Occupational	0.32 (0.32)	0.21 (0.33)	9.4* (5.4)
	Transportation	0.09 (0.13)	0.21 (0.14)	0.5 (2.0)
	SafeGraph	0.07 (0.05)	0.12** (0.05)	0.9 (0.7)
	Health access	0.04 (0.09)	0.34** (0.15)	1.2 (1.3)
	Population health	-0.16 (0.10)	-0.13 (0.12)	3.0* (1.6)
% Hispanic	Baseline coefficient	2.06*** (0.21)	2.48*** (0.26)	14.9*** (2.6)
	Explained difference	0.84** (0.36)	1.19** (0.46)	15.4*** (4.9)
	Cases per 100k Pop	N/A	N/A	18.7*** (3.4)
	Demographics	0.55*** (0.18)	0.38* (0.19)	-2.4 (2.9)
	Housing	-0.32 (0.25)	-0.37* (0.20)	1.7 (3.4)
	Socioeconomic	-0.09 (0.58)	-0.40 (0.59)	-19.1** (9.2)
	Opportunity Atlas	0.17*** (0.06)	0.05 (0.05)	0.6 (0.7)
	Occupational	0.25 (0.54)	0.29 (0.54)	9.5 (9.1)
	Transportation	0.10 (0.22)	0.09 (0.14)	0.7 (1.5)
	SafeGraph	0.25*** (0.10)	0.25*** (0.09)	1.5 (1.3)
	Health access	0.10 (0.22)	1.00*** (0.26)	2.2 (3.3)
	Population health	-0.10 (0.07)	-0.09 (0.0)	2.0* (1.2)

This table follows the corrective procedure of Gelbach (2016) for decomposing the change in coefficients from Tables 3 and 4. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

For proportion Black, the baseline disparity was smaller than that for Hispanics, and the explained part was modest as well in Table 3.

### Conclusion

In this study, we use ZIP code level data to understand the factors contributing to racial and ethnic disparities in COVID-19 burden. We find strong evidence that predominantly Black and Hispanic neighborhoods were disproportionately at risk to COVID-19 infections and mortality—although the disparities were larger among Hispanic neighborhoods. Even though our study was limited to six cities, these cities include the first (New York) and third (Chicago) largest in the USA, and case counts from these cities accounted for 14% of all confirmed US cases through June 9, 2020. The ZIP code level data allow us to examine a much wider range of variation in racial and ethnic composition than other studies using county-level data, and we also contribute

to the literature by exploring numerous possible explanations for the disparities using decomposition methods. Differences in social mobility, demographics, and long-run opportunity arose as important contributors to COVID-related disparities. However, a significant share of the disparity in cases for Blacks and Hispanics remains unexplained despite the inclusion of an exhaustive list of covariates. For fatalities, the majority of the disparities appear to be driven by the differences in cases, as opposed to differential case fatality rates. This is an important result, as it implies that interventions targeting reducing the spread of the virus in minority communities might be a more effective use of scarce resources than those targeting health care utilization once infected.

Our inability to explain most of the disparities in COVID-19 spread is perhaps surprising since we control for the risk factors largely associated with social deprivation. Recent studies have used area deprivation indexes (ADI) to explain disparities in avoidable hospitalizations and readmissions (Kind and Buckingham 2018; Jencks et al. 2019; Hu et al. 2018) and

**Table 6** COVID-19 cases per 10k population robustness checks

	6 cities (Atlanta, Baltimore, Chicago, New York City, San Diego, St. Louis), $N = 327$ Highly Segregated ZIP codes									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
% Black	0.92*** (0.14)	0.85*** (0.14)	0.88*** (0.19)	0.87*** (0.21)	0.91*** (0.23)	0.65*** (0.24)	0.72*** (0.23)	0.81*** (0.22)	0.78*** (0.23)	0.68*** (0.24)
% Hispanic	2.18*** (0.23)	1.17*** (0.24)	1.44*** (0.29)	1.07** (0.45)	1.08** (0.45)	0.88** (0.40)	1.06*** (0.38)	0.97*** (0.37)	0.95** (0.38)	0.92** (0.37)
Adj $R^2$	0.76	0.83	0.86	0.87	0.87	0.88	0.89	0.89	0.89	0.89
Demographics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opportunity Atlas	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Occupational	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Transportation	No	No	No	No	No	No	Yes	Yes	Yes	Yes
SafeGraph	No	No	No	No	No	No	No	Yes	Yes	Yes
Health access	No	No	No	No	No	No	No	No	Yes	Yes
Population health	No	No	No	No	No	No	No	No	No	Yes

Specifications include city fixed effects, “% Other non-white”, and a constant term. Unit of observation is ZIP code. *Demographic variables* include percentage in ZIP code who are male, foreign born, or in age bins (18–44, 45–64, 65–74, 75+). *Housing variables* include density, percentage who are renters, percent of units vacant, percentage who are in nursing homes, correctional facilities, college dorms, or military barracks (2010 Census), percent overcrowded (1.5+ per room), and percent with 0 or 1 bedroom sizes. *Socioeconomic variables* include percent in education bins (dropout, high school, some college, bachelor’s degree), Gini coefficient, and percent in poverty bins (0–49% FPL, 50–74%, 75–99%, 100–149%, 150–199%). *Opportunity Atlas variables* include income mobility and male incarceration (Opportunity Atlas). *Occupation variables* include percent of workers in service occupations, sales, farming, construction, production, or transport. *Transportation variables* include percent of workers of workers who use a car, percent who use public transportation, and percent with long commuting times (60+ minutes). *Safegraph variables* include percent who on average remained at home all day in each month from March to May 2020 (Safegraph). *Health access variables* include health professional shortage areas (HRSA; mental health, primary care), percent without health insurance, and COVID-19 tests per capita (Chicago and New York City only). *Population health variables* include conditional life expectancy (CDC, ages 65–74, 75–84, 85+). All control variables obtained from 2018 ACS 5-year sample unless otherwise indicated. All regressions weighted by ZIP code population from 2018 ACS 5-year sample. Heteroscedasticity-robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

difficulty managing chronic illnesses (Kurani et al. 2020; Zhang et al. 2020; Camacho et al. 2017; Durfey et al. 2019). ADIs are composite measures generally made up of weighted combinations of education, labor market composition, income, income inequality, and housing market information such as home ownership (Singh 2003). Rather than use an ADI, we allow for each of our key covariates to enter into the regression separately, which is theoretically more flexible and should allow for greater explanatory power.

With that said, if our lengthy list of explanatory variables does not explain the disparities in COVID-19 cases, then what does? Are we not more successful in explaining the disparities because we are considering the wrong theories, or because available data are inadequate to fully test the existing theories? Structural racism may influence multiple mechanisms that are difficult to quantify (Poteat et al. 2020; Braveman and Gottlieb 2014). For example, racial biases could influence clinical decision making, which would not be captured by crude access measures such as uninsured rates (Jones 2001). Obermeyer et al. (2019) found that several clinical decision-support algorithms were less likely to refer Black patients for advanced care and screening than White patients presenting

with identical symptoms. With respect to the COVID pandemic, there are reports of Blacks being denied screenings after seeking care for COVID-19 symptoms at the early stages of this pandemic (Gathright 2020; Samuels 2020; Shamus 2020; Patton 2020; Mitropoulos and Moseley 2020). Additionally, perhaps conditional life expectancies are inadequate measures of underlying health risks. The premature deaths linked to COVID-19 complications are associated with poorer underlying health status and whether the person had a pre-existing condition such as diabetes, obesity, or hypertension, and we are unable to measure those directly.<sup>18</sup> As another example, available socioeconomic information may not truly capture the source of disadvantage; percent with a high school or college degree says little about the quality of the schools attended, for instance.

<sup>18</sup> The Centers for Disease Control and Prevention has several state and national level tools to monitor chronic diseases such as the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS), but the only publicly available resource at the county-level we are aware of is the U.S. Diabetes Surveillance System, and we are not aware of any at a narrower level than county.

**Table 7** COVID-19 cases per 10k population (leave one city out)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
% Black	0.98*** (0.15)	0.64*** (0.25)	0.98*** (0.14)	0.68*** (0.26)	0.90*** (0.16)	0.46** (0.23)	0.64*** (0.12)	0.45 (0.28)	1.02*** (0.15)	0.84*** (0.25)	0.95*** (0.14)	0.65*** (0.23)
% Hispanic	2.10*** (0.22)	1.24*** (0.34)	2.06*** (0.22)	1.31*** (0.37)	1.65*** (0.23)	1.17*** (0.36)	2.05*** (0.27)	1.09** (0.46)	2.39*** (0.26)	1.51*** (0.38)	2.06*** (0.21)	1.22*** (0.33)
Adj R <sup>2</sup>	0.73	0.87	0.76	0.88	0.78	0.90	0.78	0.85	0.67	0.85	0.75	0.88
Full controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Leave out	Atlanta		Baltimore		Chicago		New York City		San Diego		St. Louis	

All specifications include city fixed effects, “% Other non-white”, and a constant term. Unit of observation is ZIP code. *Demographic variables* include percentage in ZIP code who are male, foreign born, or in age bins (18–44, 45–64, 65–74, 75+). *Housing variables* include density, percentage who are renters, percent of units vacant, percentage who are in nursing homes, correctional facilities, college dorms, or military barracks (2010 Census), percent overcrowded (1.5+ per room), and percent with 0 or 1 bedroom sizes. *Socioeconomic variables* include percent in education bins (dropout, high school, some college, bachelor’s degree), Gini coefficient, and percent in poverty bins (0–49% FPL, 50–74%, 75–99%, 100–149%, 150–199%). *Opportunity Atlas variables* include income mobility and male incarceration (Opportunity Atlas). *Occupation variables* include percent of workers in service occupations, sales, farming, construction, production, or transport. *Transportation variables* include percent of workers who use a car, percent who use public transportation, and percent with long commuting times (60+ minutes). *Safegraph variables* include percent who on average remained at home all day in each month from March to May 2020 (Safegraph). *Health access variables* include health professional shortage areas (HRSA; mental health, primary care), percent without health insurance, and COVID-19 tests per capita (Chicago and New York City only). *Population health variables* include conditional life expectancy (CDC, ages 65–74, 75–84, 85+). All control variables obtained from 2018 ACS 5-year sample unless otherwise indicated. All regressions weighted by ZIP code population from 2018 ACS 5-year sample. Heteroscedasticity-robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$

To call the COVID-19 pandemic “the Great Equalizer” is a misnomer (Mein 2020; Kim et al. 2020), and the pandemic’s key risk factors are unevenly distributed across communities. Even though we are able to explain some of the racial and ethnic disparities as attributable to different concentrations of socioeconomic risk factors, the fact that most of the case disparities remains unexplained demonstrates the difficulty of addressing deeply embedded racial and ethnic inequalities in health outcomes. Although more study is needed, our results suggest that policies enacted to curb COVID-19’s spread should consider how they would overcome the structural barriers to improvement across different groups. More superficial interventions such as economic stimulus or expanding health insurance coverage are unlikely to be fully adequate.

**Acknowledgments** Benitez acknowledges support from the Robert Wood Johnson Foundation Policies for Action Program. Yelowitz acknowledges support from University of Kentucky’s Institute for the Study of Free Enterprise. We thank Chris Bollinger, Bill Hoyt, Teresa Waters, James Ziliak, and seminar participants at University of Kentucky

for helpful comments. We thank SafeGraph for providing access to the mobility data. Data, programs, and log files are available at: <http://yelowitz.com/replication/jerp2020covid.zip>.

**Code Availability** All data, Stata programs, log files, and source documents will be stored at <http://yelowitz.com/replication/jerp2020covid.zip>.

**Authors’ Contributions** All authors contributed to the data collection, analysis, and write-up.

**Funding** Yelowitz received financial support from University of Kentucky’s Institute for the Study of Free Enterprise. Neither Benitez nor Courtemanche received financial support.

**Data Availability** All data, Stata programs, log files, and source documents will be stored at <http://yelowitz.com/replication/jerp2020covid.zip>.

**Compliance with Ethical Standards**

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Appendix**

**Table 8** Base and full models for main specifications

	Confirmed COVID-19 case per 10k pop 6 cities		Confirmed COVID-19 case per 10k pop 2 cities		COVID-19 fatalities per 1m pop 2 cities	
	Base	Full	Base	Full	Base	Full
% Black	0.923*** (0.134)	0.593*** (0.226)	1.236*** (0.191)	1.262*** (0.318)	14.3*** (2)	– 0.1 (4.5)

**Table 8** (continued)

	Confirmed COVID-19 case per 10k pop 6 cities		Confirmed COVID-19 case per 10k pop 2 cities		COVID-19 fatalities per 1m pop 2 cities	
	Base	Full	Base	Full	Base	Full
% Hispanic	2.057*** (0.213)	1.218*** (0.332)	2.484*** (0.263)	1.29*** (0.375)	14.9*** (2.6)	− 0.4 (4.3)
% other	0.511* (0.305)	− 0.987** (0.415)	0.886** (0.409)	− 0.403 (0.391)	13.5** (5.6)	2.7 (5)
Atlanta zip code	− 175.251*** (8.597)	− 264.171*** (26.535)				
Baltimore zip code	− 108.232*** (10.873)	− 180.516*** (19.534)				
Chicago zip code	− 63.606*** (10.514)	− 122.915*** (15.539)	− 62.821*** (10.139)	− 52.887*** (19.687)	− 1191.6*** (98.2)	− 127.2 (238.7)
San Diego zip code	− 200.835*** (7.221)	− 238.392*** (22.822)				
St. Louis zip code	− 150.886*** (13.499)	− 231.067*** (28.413)				
Confirmed COVID-19 case per 10k pop						7.5*** (1)
% Male		4.006*** (1.338)		− 0.165 (1.167)		− 18 (24.5)
% foreign-born		1.164** (0.452)		0.757* (0.402)		4 (5.6)
% age 18–44		− 1.464* (0.756)		− 0.313 (0.915)		− 17.2 (13.2)
% age 45–64		− 2.45** (1.097)		− 1.181 (1.204)		− 19 (21.1)
% age 65–74		0.812 (2.074)		0.363 (2.553)		32.9 (31.9)
% age 75+		− 0.953 (2.077)		− 2.912 (2.229)		60.7** (29.1)
Density		0 (0)		0 (0)		0 (0)
% renter		− 0.39 (0.423)		− 0.559 (0.448)		4.3 (7.1)
% vacant		1.052* (0.628)		− 1.214 (0.809)		7.1 (8.9)
% correctional group quarters		0.705 (1.377)		2.396*** (0.743)		− 27.7** (13.3)
% nursing home group quarters		16.452*** (4.956)		6.054** (2.827)		278.9*** (100.2)
% college group quarters		2.686*** (0.853)		1.53 (0.962)		23.4 (16.7)
% military group quarters		− 0.577 (0.856)		274.535** (121.418)		158.8 (2996.4)
% 1.5+ occupants per bedroom		1.54 (2.865)		− 1.542 (1.66)		17.4 (23.4)
% unit is 0 or 1 bedroom		1.193*** (0.415)		0.865** (0.34)		− 7.2 (5.1)
% HS dropout		− 0.061 (1.384)		0.091 (1.284)		− 3.4 (16.7)
% HS graduate/GED		1.997** (0.997)		0.71 (0.86)		9.5 (13.3)
% some college		− 2.392** (1.004)		− 3.827*** (1.101)		− 7.5 (16.6)
% bachelor's degree		1.148		1.03		49.2***



**Table 8** (continued)

	Confirmed COVID-19 case per 10k pop 6 cities		Confirmed COVID-19 case per 10k pop 2 cities		COVID-19 fatalities per 1m pop 2 cities	
	Base	Full	Base	Full	Base	Full
Gini coefficient		(1.024)		(1.208)		(17.5)
		– 135.128 (90.609)		27.099 (85.101)		2261.5 (1523.7)
% 0–49 FPL		1.003 (1.205)		1.075 (1.217)		– 13.9 (18.1)
% 50–74 FPL		1.588 (2.075)		– 0.391 (2.012)		– 62.3** (29.9)
% 75–99 FPL		– 0.952 (1.776)		0.127 (1.641)		38.7 (26.8)
% 100–149 FPL		– 0.176 (1.569)		– 0.46 (1.456)		10.5 (27.8)
% 150–199 FPL		– 1.166 (1.552)		0.25 (1.521)		3.9 (24.6)
Income mobility		– 199.703*** (54.695)		– 69.522 (74.58)		– 915.1 (1114.3)
Male incarceration		116.042 (121.102)		– 18.72 (149.98)		– 340.2 (1867.8)
% service		0.588 (1.071)		– 0.655 (0.983)		37.3** (16.7)
% sales occupation		1.038 (1.061)		2.468** (1.146)		7.2 (17)
% farming		– 18.273*** (6.386)		– 19.859 (21.675)		228.5 (470.4)
% construction		– 2.91* (1.517)		0.973 (1.594)		7.9 (27.1)
% production		2.855 (3.301)		1.551 (2.702)		– 64.2* (34.3)
% transport		1.645 (1.662)		2.592** (1.173)		4.2 (18.9)
% travel by car		0.988** (0.49)		0.983 (0.601)		– 10.6 (8.9)
% travel by public transit		– 1.304*** (0.484)		– 0.39 (0.388)		– 0.5 (6.3)
% 60+ min commute		1.262** (0.503)		0.735 (0.483)		7.5 (7.1)
% home all day, 3/2020		151.357 (125.94)		91.833 (114.989)		331.7 (1513.7)
% home all day, 4/2020		300.218*** (95.738)		191.483** (92.184)		2052.3 (1289.9)
% home all day, 5/2020		– 325.512** (146.591)		– 110.497 (117.307)		– 2551* (1536)
Mental health HPSA		27.891 (18.825)		3.732 (14.132)		– 86.5 (158)
Primary care HPSA		– 34.379** (14.534)		– 14.927 (20.405)		704.6*** (175.1)
% uninsured		0.262 (1.306)		3.397*** (1.17)		21.9 (19.1)
COVID-19 tests per 10k pop				0.223*** (0.023)		– 0.6** (0.3)
Life expectancy, age 65–74		7.129 (6.275)		8.421 (8.349)		– 180.6* (100.6)
Life expectancy, age 75–84		0.575		– 6.116		121.7

**Table 8** (continued)

	Confirmed COVID-19 case per 10k pop 6 cities		Confirmed COVID-19 case per 10k pop 2 cities		COVID-19 fatalities per 1m pop 2 cities	
	Base	Full	Base	Full	Base	Full
Life expectancy, age 85+		(8.211) – 3.781 (3.658)		(11.206) 0.23 (4.2)		(119.3) 45.5 (57.8)
Life expectancy missing		58.425 (65.566)		62.816 (72.579)		– 1693.1 (1209.2)
Constant term	145.318*** (12.63)	– 27.46 (129.995)	119.721*** (16.532)	– 217.922 (141.339)	1047.4*** (180.7)	798.8 (2703.6)

## References

- Almagro M, Orane-Hutchinson A. The determinants of the differential exposure to COVID-19 in New York City and their evolution over time. *Covid Economics*. 2020;13 <https://cepr.org/sites/default/files/news/CovidEconomics13.pdf>.
- Arcaya MC, Tucker-Seeley RD, Kim R, Schnake-Mahl A, So M, Subramanian SV. Research on neighborhood effects on health in the United States: a systematic review of study characteristics. *Soc Sci Med*. 2016;168:16–29. <https://doi.org/10.1016/j.socscimed.2016.08.047> <http://www.sciencedirect.com/science/article/pii/S027795361630483X>.
- Artiga S, Garfield R, Orgera K. Communities of color at higher risk for health and economic challenges due to COVID-19. Kaiser Family Foundation; 2020. <https://www.kff.org/coronavirus-covid-19/issue-brief/communities-of-color-at-higher-risk-for-health-and-economic-challenges-due-to-covid-19/>
- Borjas, George J. 2020. "Demographic determinants of testing incidence and COVID-19 infections in New York City neighborhoods." *National Bureau of Economic Research Working Paper Series No. 26952*. <https://doi.org/10.3386/w26952>. <http://www.nber.org/papers/w26952>.
- Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Reports*. 2014;129(1suppl2):19–31. <https://doi.org/10.1177/00333549141291S206>.
- Camacho FT, Tan X, Alcalá HE, Shah S, Anderson RT, Balkrishnan R. Impact of patient race and geographical factors on initiation and adherence to adjuvant endocrine therapy in medicare breast cancer survivors. *Medicine*. 2017;96(24):e7147–7.
- Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. *J Hum Resour*. 2015;50(2):317–72. <https://doi.org/10.3368/jhr.50.2.317>.
- Carfi A, Bernabei R, Landi F, for the Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent symptoms in patients after acute COVID-19. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.12603>.
- CDC, U. S. Centers for Disease Control and Prevention. "COVID-19 in racial and ethnic minority groups." GA: Atlanta; 2020b.
- CDC, U. S. Centers for Disease Control and Prevention. "COVID-19 laboratory-confirmed hospitalizations." GA: Atlanta; 2020c.
- CDC, U. S. Centers for Disease Control and Prevention. "Coronavirus Disease 2019 (COVID-19): Cases & Deaths in the U.S.". GA: Atlanta; 2020a.
- Chetty R, Friedman JN, Hendren N, Jones MR, Porter SR. The opportunity atlas: mapping the childhood roots of social mobility. National Bureau of Economic Research Working Paper Series. 2018;25147. <https://doi.org/10.3386/w25147> <http://www.nber.org/papers/w25147>.
- Chetty R, Hendren N, Jones MR, Porter SR. Race and economic opportunity in the United States: an intergenerational perspective. *The Quarterly Journal of Economics*. 2019;135(2):711–83 <https://doi.org/10.1093/qje/qjz042>.
- CMS, U.S. Centers for Medicare & Medicaid Services. Physician health professional shortage area bonuses. Accessed July. 2020;8:2020 <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HPSAPSAPhysicianBonuses>.
- Courtemanche C, Marton J, Ukert B, Yelowitz A, Zapata D. Early impacts of the affordable care act on health insurance coverage in Medicaid expansion and non-expansion states. *Journal of Policy Analysis and Management*. 2017;36(1):178–210. <https://doi.org/10.1002/pam.21961> <https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.21961>.
- Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Aff*. 2020;39(7):1237–46. <https://doi.org/10.1377/hlthaff.2020.00608>.
- Coven J, Gupta A. Disparities in mobility responses to COVID-19. NYU Working Paper; 2020. <https://static1.squarespace.com/static/56086d00e4b0fb7874bc2d42/t/5ebf201183c6f016ca3abd91/1589583893816/DemographicCovid.pdf>
- Desmet K, Wacziarg R. Understanding spatial variation in COVID-19 across the United States. National Bureau of Economic Research Working Paper Series. 2020;27329. <https://doi.org/10.3386/w27329> <http://www.nber.org/papers/w27329>.
- Durfey SNM, Kind AJH, Buckingham WR, DuGoff EH, Trivedi AN. Neighborhood disadvantage and chronic disease management. *Health services research* 54 Suppl. 2019;1(Suppl 1):206–16.
- Eisenberg MD, Barry CL, Schilling C, Kennedy-Hendricks A. Financial risk for COVID-19-like respiratory hospitalizations in consumer-directed health plans. *American Journal of Preventive Medicine*. Forthcoming. <https://doi.org/10.1016/j.amepre.2020.05.008>.
- Figueroa JF, Wadhwa RK, Lee D, Yeh RW, Sommers BD. Community-level factors associated with racial and ethnic disparities in COVID-19 rates in Massachusetts. *Health Affairs*. 2020;0(0). <https://doi.org/10.1377/hlthaff.2020.01040> <https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.01040>.
- Friedson AI, McNichols D, Sabia JJ, Dave D. Did California's shelter in place order work? Early evidence on coronavirus-related health benefits. *National Bureau of Economic Research Working Paper Series*. 2020;26992. <https://doi.org/10.3386/w26992> <http://www.nber.org/papers/w26992>.
- Gathright, Jenny. 2020. D.C.'s Black residents make up less than half the population, 80% Of COVID-19 deaths : NPR.

- Gelbach JB. When do covariates matter? And which ones, and how much? *J Labor Econ*. 2016;34(2):509–43. <https://doi.org/10.1086/683668> <https://www.journals.uchicago.edu/doi/abs/10.1086/683668>.
- Gupta S, Nguyen TD, Rojas FL, Raman S, Lee B, Bento A, et al. Tracking public and private responses to the COVID-19 epidemic: evidence from state and local government actions. *National Bureau of Economic Research Working Paper Series*. 2020;27027. <https://doi.org/10.3386/w27027> <http://www.nber.org/papers/w27027>.
- Harrison, David. 2020. "Public transit use is associated with higher coronavirus death rates, researchers find." *The Wall Street Journal*, June 28, 2020, 2020. <https://www.wsj.com/articles/public-transit-use-is-associated-with-higher-coronavirus-death-rates-researchers-find-11593342001>.
- Hu J, Kind AJH, Nerenz D. Area deprivation index predicts readmission risk at an urban teaching hospital. *American journal of medical quality : the official journal of the American College of Medical Quality*. 2018;33(5):493–501.
- Hubler, Shawn, Thomas Fuller, Anjali Singhvi, and Juliette Love. 2020. "Many Latinos couldn't stay home. Now virus cases are soaring in their communities." *The New York Times*, June 26, 2020, 2020. <https://www.nytimes.com/2020/06/26/us/corona-virus-latinos.html>.
- Jencks SF, Schuster A, Dougherty GB, Gerovich S, Brock JE, Kind AJH. Safety-net hospitals, neighborhood disadvantage, and readmissions under Maryland's all-payer program. *Ann Intern Med*. 2019;171(2):91–8 <https://www.acpjournals.org/doi/abs/10.7326/M16-2671>.
- Jones CP. Invited commentary: "race," racism, and the practice of epidemiology. *American Journal of Epidemiology*. 2001;154(4):299–304. <https://doi.org/10.1093/aje/154.4.299>.
- Killeen, Benjamin D., Jie Ying Wu, Kinjal Shah, Anna Zapaishchikova, Philipp Nikutta, Aniruddha Tamhane, Shreya Chakraborty, Jinchai Wei, Tiger Gao, Mareike Thies, and Mathias Unberath. 2020. "A county-level dataset for informing the United States' response to COVID-19" [https://github.com/JieYingWu/COVID-19\\_US\\_County-level\\_Summaries/blob/master/data/interventions.csv](https://github.com/JieYingWu/COVID-19_US_County-level_Summaries/blob/master/data/interventions.csv).
- Kim EJ, Marrast L, Conigliaro J. COVID-19: magnifying the effect of health disparities. *J Gen Intern Med*. 2020:1–1.
- Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible—the neighborhood atlas. *N Engl J Med*. 2018;378(26):2456–8.
- Kniesner TJ, Sullivan RS. The forgotten numbers: a closer look at Covid-19 non-fatal valuations. *Journal of Risk and Uncertainty*. Forthcoming.
- Knittel, Christopher R, and Bora Ozaltun. 2020. "What does and does not correlate with COVID-19 death rates." *medRxiv*: 2020.06.09.20126805. <https://doi.org/10.1101/2020.06.09.20126805>. <https://www.medrxiv.org/content/medrxiv/early/2020/06/11/2020.06.09.20126805.full.pdf>.
- Krogstad, Jens Manuel. 2019. "Reflecting a demographic shift, 109 U.S. counties have become majority nonwhite since 2000." Accessed July 3, 2020. <https://www.pewresearch.org/fact-tank/2019/08/21/u-s-counties-majority-nonwhite/>.
- Kurani SS, McCoy RG, Lampman MA, Doubeni CA, Finney Rutten LJ, Inselman JW, et al. Association of neighborhood measures of social determinants of health with breast, cervical, and colorectal cancer screening rates in the US Midwest. *JAMA Netw Open*. 2020;3(3):e200618–8.
- Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995:80–94.
- Logan TD, Parman JM. Segregation and mortality over time and space. *Soc Sci Med*. 2018;199:77–86. <https://doi.org/10.1016/j.socscimed.2017.07.006> <http://www.sciencedirect.com/science/article/pii/S0277953617304288>.
- McLaren, John. 2020. Racial disparity in COVID-19 deaths: seeking economic roots with Census data. <http://www.nber.org/papers/w27407>.
- Mein SA. COVID-19 and health disparities: the reality of "the great equalizer". *J Gen Intern Med*. 2020:1–1.
- Mitropoulos, Arielle, and Mariya Moseley. 2020. Beloved Brooklyn teacher, 30, dies of coronavirus after she was twice denied a COVID-19 test—ABC News.
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–53. <https://doi.org/10.1126/science.aax2342> <https://science.sciencemag.org/content/sci/366/6464/447.full.pdf>.
- Oppel Jr., Richard A., Robert Gebeloff, K.K. Rebecca Lai, Will Wright, and Mitch Smith. 2020. "The fullest look yet at the racial inequity of Coronavirus." *The New York Times*, July 5, 2020, 2020. <https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html>.
- Patton, Stacey. 2020. Black America is suffering from the coronavirus—the Washington post.
- Phelan JC, Link BG. Is racism a fundamental cause of inequalities in health? *Annu Rev Sociol*. 2015;41:311–30.
- Poteat T, Millett GA, Nelson LRE, Beyrer C. Understanding COVID-19 risks and vulnerabilities among black communities in America: the lethal force of syndemics. *Ann Epidemiol*. 2020;47:1–3 <http://www.sciencedirect.com/science/article/pii/S1047279720301770>.
- Price GN. Hurricane Katrina as an experiment in housing mobility and neighborhood effects: were the relocated poor Black evacuees better-off? *The Review of Black Political Economy*. 2013;40(2):121–43. <https://doi.org/10.1007/s12114-012-9151-5>.
- Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among Black patients and White patients with Covid-19. *N Engl J Med*. 2020. <https://doi.org/10.1056/nejmsa2011686>.
- Samuels, Alex. 2020. Black Texans already face health care disparities: the coronavirus is making it worse.
- Schaeffer K. Rising number of U.S. counties are majority Hispanic or Black. Accessed July. 2019;3:2020 <https://www.pewresearch.org/fact-tank/2019/11/20/in-a-rising-number-of-u-s-counties-hispanic-and-black-americans-are-the-majority/>.
- Schmitt-Grohé S, Teoh K, Uribe M. Covid-19: testing inequality in New York City. *National Bureau of Economic Research Working Paper Series*. 2020;27019. <https://doi.org/10.3386/w27019> <http://www.nber.org/papers/w27019>.
- Schuetz J, Been V, Ellen IG. Neighborhood effects of concentrated mortgage foreclosures. *J Hous Econ*. 2008;17(4):306–19. <https://doi.org/10.1016/j.jhe.2008.09.004> <http://www.sciencedirect.com/science/article/pii/S1051137708000338>.
- Selden TM, Berdahl TA. COVID-19 and racial/ethnic disparities in health risk, employment, and household composition. *Health Aff*. 2020;39(9):1624–32. <https://doi.org/10.1377/hlthaff.2020.00897> <https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.00897>.
- Shamus, Kristen Jordan. 2020. Family ravaged by coronavirus begged for tests, hospital care but was repeatedly denied.
- Singh GK. Area deprivation and widening inequalities in US mortality, 1969–1998. *Am J Public Health*. 2003;93(7):1137–43 <https://pubmed.ncbi.nlm.nih.gov/12835199> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447923/>.
- Wiemers EE, Abrahams S, Marwa AF, Joseph Hotz V, Schoeni RF, Seltzer JA. *Disparities in vulnerability to severe complications from COVID-19 in the United States*. 2020; <http://www.nber.org/papers/w27294>.
- Zhang Y, Ancker JS, Hall J, Khullar D, Wu Y, Kaushal R. Association between residential neighborhood social conditions and health care utilization and costs. *Med Care*. 2020;58(7):586–93.