

NBER WORKING PAPER SERIES

RACIAL DISPARITY IN COVID-19 DEATHS:
SEEKING ECONOMIC ROOTS WITH CENSUS DATA.

John McLaren

Working Paper 27407
<http://www.nber.org/papers/w27407>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by John McLaren. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data.

John McLaren

NBER Working Paper No. 27407

June 2020

JEL No. I14,J15

ABSTRACT

This note seeks the socioeconomic roots of racial disparities in COVID-19 mortality, using county-level mortality, economic, and demographic data from 3,140 counties. For all minorities, the minority's population share is strongly correlated with total COVID-19 deaths. For Hispanic/Latino and Asian minorities those correlations are fragile, and largely disappear when we control for education, occupation, and commuting patterns. For African Americans and First Nations populations, the correlations are very robust. Surprisingly, for these two groups the racial disparity does not seem to be due to differences in income, poverty rates, education, occupational mix, or even access to healthcare insurance. A significant portion of the disparity can, however, be sourced to the use of public transit.

John McLaren

Department of Economics

University of Virginia

P.O. Box 400182

Charlottesville, VA 22904-4182

and NBER

jmclaren@virginia.edu

1 Introduction

The higher mortality rates for African Americans and other minority groups in the COVID-19 pandemic has been the subject of much public concern. Centers for Disease Control (2020) surveys much of what is known about the phenomenon, and Wood (2020) provides a journalistic survey. Gross et al. (2020) find that the mortality rates for African Americans are more than triple the rates for whites after correcting for age, and the rates for Hispanic/Latinos are almost double the rates for non-Hispanic whites. The Navajo Nation has one of the highest mortality rates in the nation.¹ The sources of these disparities, however, are more difficult to analyze than establishing the existence of such disparities. If the disparities can be explained in terms of socioeconomic factors such as income, access to insurance, and occupation, that knowledge can be useful in designing policies to reduce the disparities. A number of authors have called for this analysis of sources as an urgent priority for research. Chowkwanyun and Reed (2020) argue that without an analysis of the sources, the political discourse can gravitate toward ‘biologic explanations’ or explanations based on racial stereotypes which are harmful in themselves and get in the way of policy solutions.² Wood (2020) summarizes several commentators on the topic, arguing that on COVID-19 racial disparity, “Outrage is warranted. But outrage unaccompanied by analysis is a danger in itself.” Timothy Freeman, pastor at Trinity African Methodist Episcopal Zion Church in Washington, DC, suggested that ‘COVID-19 is affecting black and brown people in disproportionate numbers, “and not just because we’re black and brown, but because of the social and economic conditions people are forced to live in...”’ Johnson (2020).

In this important task of finding the sources of the disparity, a major difficulty is data limitations. As Gross et al. (2020) report, only 28 state health systems break down their

¹See, for example, Simon Romero and Jack Healy, “Tribal Nations Face Most Severe Crisis in Decades as the Coronavirus Closes Casinos,” *New York Times*, May 11, 2020.

²A recent example of this is the remarks made by Republican Ohio State Senator Steve Huffman, who ignited controversy when he asked, in discussing the racial disparity in mortality rates, “Could it just be that African Americans or the colored population do not wash their hands as well as other groups or wear a mask or do not socially distance themselves?” (Mogan Gstalter, “Ohio lawmaker asks if ‘colored population’ is hit hard by coronavirus because they ‘don’t wash their hands as well,’” *The Hill*, June 11, 2020.)

COVID-19 data reports by race. Even when the case information is broken down by race, confidentiality rules prevent information about the patient’s economic variables from being matched up with the clinical information.³ Ideally, one would have a large sample of COVID-19 patients and non-COVID-19 patients with full information about employment, education, occupation, income, and so on for each individual but that is not available.⁴

This note provides an approach to get around these problems partially, using geographic variation in aggregate variables by county.⁵ The data used are county aggregates of COVID-19 mortality over time, matched with socioeconomic and demographic data by country. The mortality data are not broken down by race, but in principle one can still test for and measure racial disparity by looking for correlation between the size of a county’s minority population share and the county’s mortality rate. If there were no difference in mortality between a minority group and the rest of the population, the size of the minority share would have no effect on the county’s mortality rate, so this correlation provides indirect evidence of a disparity in mortality rates. The exercise follows two steps: First, confirm the existence of a racial disparity by using a regression to show that minority population shares are correlated with mortality rates. Second, control for a range of county-level socio-economic factors to see if the racial disparity weakens or disappears. If it does, then that stands as evidence that the socioeconomic factor is part of the reason for the racial mortality discrepancy.

We apply this simultaneously to four minorities, as defined by the Census: ‘Black or African American,’ which will be abbreviated here as ‘African American;’ ‘Hispanic/Latino;’ ‘Asian;’ and ‘American Indian and Alaska Native,’ which will be abbreviated here as ‘First

³For example, Price-Haywood et al. (2020) obtained detailed medical data on thousands of COVID-19 patients at a hospital in Louisiana for a study on racial disparity, but the only information they have on the socioeconomic status of each patient is the insurance type and the patient’s zip code. On the other hand, Wiemers et al. (2020) are able to link pre-existing medical risk factors for COVID-19 with socioeconomic factors using the Panel Study of Income Dynamics.

⁴Wood (2020) quotes Merlin Chowkwanyun on the difficulty of identifying the sources of the disparity: ‘That means collecting data that currently do not exist. “When a patient comes to a clinic, they can self-report their race, but how do they self-report socioeconomic status? How do you actually gather that information?”’

⁵In general terms, this is a direction suggested by Chowkwanyun and Reed (2020), and is similar in spirit to the work of the Harvard Public Health Disparities Geocoding Project.

Nations.’ To anticipate results, the main findings are:

(i) For all four minorities, after controlling for state-level effects, there is a strong positive correlation across counties between the minority’s population share and COVID-19 deaths. (ii) For Hispanic/Latino and Asian minorities those correlations are fragile, and largely disappear when we correct for education, occupation, and commuting patterns. (iii) By contrast, for African Americans and First Nations populations, the correlations are very robust. Regardless of what other factors are controlled for, a racial disparity in mortality rates persists. (iv) Surprisingly, for these two groups the racial disparity does *not* seem to be due to differences in income, poverty rates, education, occupational mix, or even access to healthcare insurance, which has been hypothesized by many observers to be a key source of the disparity. (v) A significant portion of the disparity can, however, be sourced to the use of public transit.

As a consequence, for African American and First Nations minorities, a large portion of the disparities is not explained by the socioeconomic factors that we are able to measure. Other candidates, which are beyond the scope of this note, are listed in Centers for Disease Control (2020), and include:

(i) Concentration of workers in essential services. This study uses fairly broad occupational categories due to data limitations, but identifying the number of workers in essential services in each location at each time requires finer occupational data. Of course, ‘essential service’ is defined by state governments and the criteria can vary over time.

(ii) Differential incidence of pre-existing conditions that can make a COVID-19 infection more dangerous. This is a major factor identified in case-study data by Price-Haywood et al. (2020), for example. Wiemers et al. (2020) extensively study the socioeconomic determinants of pre-existing risk factors for COVID-19.

(iii) Differential availability of paid sick leave. In a workplace where workers have no paid sick leave, workers with symptoms are likely to stay on the job and infect others. This could be an extremely important channel for the disparity.

(iv) Residential segregation. There is an earlier literature on the effects of residential segregation on healthcare outcomes over and above a household’s own income, education and other socioeconomic indicators. For example, Hearst et al. (2008) look for an effect of segregation on infant mortality and find no effect after other factors have been controlled for.

(v) Discrimination in healthcare services. Hall et al. (2015) survey numerous studies that demonstrate evidence of implicit racial or ethnic bias in healthcare providers and find some effect on healthcare outcomes. Hoffman et al. (2016) report evidence that a substantial fraction of health professionals undervalue the pain experienced by African-American patients due to racial attitudes, and there are claims that African Americans who have had adverse experiences with healthcare professionals have trouble trusting them and so are slower to seek treatment (Johnson (2020)).

(vi) ‘Weathering,’ which is defined as “advanced aging caused by bodily wear and tear from fight-or-flight responses to external stressors, especially racial discrimination.” (Chowkwanyun and Reed (2020)).

Related work.

The method in this note is close to Desmet and Wacziarg (2020), who also use county-level data to relate COVID-19 outcome data to socioeconomic county characteristics. The main difference is that they do not focus on the racial disparity. Gross et al. obtain county-level data from the 28 states whose governments release COVID-19 mortality data broken down by race in order to measure the racial disparity. Unlike the present note, their focus is not to try to identify socioeconomic roots of the disparity, but rather to provide a more accurate measure of the disparity, and especially after correcting for age. Price-Haywood et al. (2020) obtained confidential data on 3,481 COVID-19 patients at a hospital in Louisiana. For each patient, the data include medical history and the treatment and outcome of the COVID-19 infection over time, as well as the type of insurance and the zip code of residence. That last piece of information is used as a noisy measure of socioeconomic status, as the

zip codes served by the hospital are very varied in income and other indicators. They find an overwhelming racial disparity in COVID-19 outcomes; the vast majority of COVID-19 patients and deaths were African American, although two-thirds of the local population were white. However, once a patient is admitted, controlling for patient specific information (including pre-existing conditions), they found that race had no effect on the conditional survival probability. Wiemers et al. (2020) identify the key pre-existing medical risk factors for COVID-19 and use data from the 2017 wave of the Panel Study of Income Dynamics to study the socioeconomic factors that predict these risk factors. Race, age, income, and education all emerge as important predictors of these risk factors.

2 Data.

All data in this note are at the county level, from 50 states plus the District of Columbia and Puerto Rico. The ideal would be individual data but that is not publicly available, so the present goal is to work with the finest possible geographic units. Mortality data are available at the level of zip codes, which are typically smaller than counties, but the economic and demographic variables are typically not available at that level.⁶ The data include 3,140 counties after counties with missing data have been dropped, covering a total population of 322,179,225. Counties are extremely heterogenous in size, with a highly skewed distribution. The smallest is Kalawao County, Hawaii, with 75 people, and the largest is Los Angeles County, with more than 10 million. The mean size is 102,605 and the median is 25,732. For this reason, all descriptive statistics reported in Table 1 and all regressions are weighted by population.

⁶This is unfortunate, because it is well known that much of the relevant variation exists at the zip code level. See the use of zip codes locally in the service area of one hospital in Louisiana by Price-Haywood et al. (2020), and the anecdotal discussion of zip codes in Rubin et al. (2020).

2.1 Mortality data.

Problems in measuring the number of COVID-19 cases are well-known, with large disparities in the availability of testing and criteria for testing. For that reason, this note focusses on mortality data, which is less likely to be distorted by such problems. Of course, even cause of death is subject to judgement calls, as for example when a patient with a severe COVID-19 infection dies of a heart attack that may have been triggered by the infection. The disease-specific deaths rates are therefore themselves imperfect measures of the prevalence of the virus but are likely the best available measure.

Data on COVID-19 deaths by county are from usafacts.org, a private non-profit group that compiles them from local sources. The data are cumulative deaths per day in each county; here, we use cumulative deaths on 1/22 (the first day in the data), 2/22, 3/22, 4/22, and 5/19 (the last day in the data at the time of the download). To find deaths during each month, we take first differences. For two counties, this results in a negative death recorded for May. Lauderdale County, AL, population 92,585, lists 3 cumulative deaths as of April but only 2 as of May. Craighead County, Arkansas records 1 death as of April but none as of May. Both of these anomalies are likely cases of re-classification of a death either by cause of death or jurisdiction in which it is filed. Some small number of re-classifications likely occur throughout the data over time, in both directions, but only negative reclassifications are visible to researchers, so eliminating these cases from the data would introduce a bias.

All mortality figures are scaled as deaths per million. Table 1 shows that the population-weighted average mortality was 135.72 per million in April and 277.67 in May. These figures are sharply skewed to the right, with three counties of New York City showing above 800 per million while also several less-publicized small counties such as Early County, GA and Morgan County, CO had deaths above 1,000 per million. At the same time, the median county had no deaths, while the population-weighted median was 36 in April and 85 in May.

2.2 Demographic and economic data.

The demographic and economic variables come from the American Community Survey (ACS) of the US Census Bureau. The most recent round of the ACS with complete publicly-available information required for this exercise is the 2018 survey; the five-year data are used here, meaning that all data are an average from the five years of the ACS ending in 2018. The reason is that the 2018 figures alone have the vast majority of counties suppressed due to Census confidentiality rules.

(i) Racial and ethnic variables. This note examines four minority groups as defined by the Census: ‘Black or African American,’ which will be abbreviated as ‘African American;’ ‘Hispanic Or Latino,’ (which is listed under the Ethnicity categories rather than race, unlike the others listed here); ‘Asian;’ and ‘American Indian and Alaska Native,’ which will be abbreviated as ‘First Nations.’ Table 1 shows that the African-American and Hispanic/Latino populations shares are much larger than the other two, and there is substantial variation in all four across counties.

(ii) Income, Education, and Health Insurance. In order to check whether or not incidence of the disease is driven by narrow economic factors, the following are extracted from the ACS, summarized in rows 8-14 of Table 1: median household income in each county; the fraction of adults 25 years of age or older without high-school diploma; with high-school diploma but no further education; with some college (including Associate’s degree from community college) but no four-year degree; and with a four-year college degree; and the poverty rate in each county. The ACS also asks respondents if they have any healthcare insurance, and the fraction who respond in the negative for each county is included in the data here; the average nationwide stands at 9.37%.

(iii) Commuting. The ACS reports detailed information on how workers get to work, from which we extract two variables that seem potentially relevant to transmission of the virus: The fraction of workers who share a vehicle to get to work, and the fraction who use public transit (not including taxis) to do so. Both means of transport involve being

inside a vehicle with other people for the length of the ride and could therefore potentially help spread the virus. The public-transit fraction is dramatically skewed. Even the 75th percentile county (weighted, as always, by population) has only a 3.89% share who use public transit, significantly below the mean of 4.80%. The highest share is held by Kings County, NY, at 61.36%.

The ability to work from home. Since physical distancing is a key strategy for preventing spread of the virus, workers who are able to work from home may well have an advantage in staying healthy. Dingel and Neiman (2020) have attempted to quantify the likelihood that a given US worker is able to work from home, using the O*NET surveys conducted by the US Department of Labor, which codify differences across US occupations. For example, one survey question asks workers if they work outside every day; another question asks if the worker uses “vehicles, mechanized devices, or equipment” on the job. Occupations for which a majority of respondents answer affirmatively to either of these questions are coded by Dingel and Neiman as occupations that cannot be performed at home. Using these and numerous other criteria, Dingel and Neiman go through the O*NET surveys, coding each job as work-at-home compatible or not, and then use Bureau of Labor Statistics (BLS) data on numbers of workers in each job to aggregate up to broader categories. The 22 two-digit categories are used here because they are what are released by the Census Bureau at the county level. This note uses the weighted average of the Dingel-Neiman index across those categories for each county, where the weight for each occupation category is that category’s share of employment in that county. The last row of Table 1 shows that by this measure about 39% of workers can work from home, a figure that varies from county to county with the occupational make-up of employment. We also use the 22 broad BLS occupational categories separately.

3 Empirical approach and basic results

The basic estimation is based on the equation:

$$mortality_t^c = \beta_0 + \sum_m \beta_m D(m, t) + \beta^R ROS_t^c + \sum_k \beta_0^k share_c^k + \sum_k \sum_m \beta_m^k share_c^k D(m, t) + \epsilon_{c,t}, \quad (1)$$

where $mortality_t^c$ is COVID-19 deaths per million in county c during month t ; $D(m, t)$ is a dummy variable that takes a value of 1 if $m = t$ and 0 otherwise, where $m, t \in \{March, April, May\}$; ROS_t^c is mortality per million during month t in the rest of the state to which county c belongs; $share_c^k$ is the share of minority group k in the population of county c in the 2018 ACS figures; and $\epsilon_{c,t}$ is an error term. This regression establishes the basic correlations between COVID-19 deaths and minority populations.

In this equation, any positive value for β_m^k implies a higher-than-average mortality rate for minority k , since a mere increase in the population share of a given minority would not change the overall mortality rate if all racial and ethnic groups had the same mortality rate. The month-dummy terms $\beta_m D(m, t)$ can capture any national trend, while the rest-of-state term $\beta^R ROS_t^c$ will capture any state-level shock. States vary in innumerable ways that are not in the data and that may be relevant to the pandemic; for example, coastal states are likely to receive more foreign visitors from pandemic-afflicted countries in Europe and east Asia, and, as Desmet and Wazciarg (2020) show, locations close to an airport that carries a large number of passengers from the most afflicted countries tend to have more cases. State-level policy to deal with the pandemic also varies in numerous ways. With ROS_t^c included in the regression, the identifying variation will be *cross-county* variation over time, holding state-level factors constant. Note: Because the District of Columbia has only one county, it has no equivalent to this term and so it is omitted from the regressions.

Subsequent regressions add additional terms of the form:

$$\sum_m \beta_m^X X_c D(m, t),$$

where X_c is a characteristic of county c , to see if the minority effects β_m^k estimated in the first regression are the result of correlations with other measurable county characteristics. If part of the reason for a non-zero coefficient β_t^k in regression (1) is that variable X_c drives differences in COVID-19 deaths and it is correlated with race or ethnicity, then adding that term to the regression should reduce the estimated magnitude of β_t^k .

4 Results.

4.1 Basic regression.

Column (1) of Table 2 shows the simplest form of regression (1), with only a constant plus dummies for the months March, April and May. The time trend is obvious from the sharply increasing, and highly significant, coefficient over time.⁷ Column (2) adds the control for state-level shocks, in the form of rest-of-state mortality (ROS_t^c in (1)). This variable enters with a coefficient of 0.92, strongly significant, and eliminates the effect of the time dummies while quintupling the R^2 . The implication is that the time trends in the data are state-level, not national.

Column (3) presents the results with the minority population shares added, the main variable of interest. Because there are hardly any deaths in February and March, the first two coefficients for each of the four categories are small in magnitude and mostly insignificant, but for each of them the April and May coefficients are positive, large, and highly significant. This implies that across the board, controlling for state effects, counties with large minority populations were hit harder by the pandemic. Column (4) controls for the size of the county ('population'), to allow for the possibility that the patterns may be driven by differences between large urban counties and small rural ones. The population terms have infinitesimal estimated coefficients, which are small even when multiplied by the standard deviation of population and are mostly insignificant. In addition, they do not cause any meaningful

⁷As a reminder, this and all regressions are weighted by county population.

change in the minority-share coefficients.

Henceforth, in all regressions, for completeness and consistency all variables in column (4) will be included, but the un-interacted and March-interacted minority population shares, the un-interacted month dummies, and the population terms will all be omitted from tables for brevity.

To get an idea of the magnitudes, we can think about these estimates in two ways. First, we can ask how much of an increase in death rates is predicted by a one-standard-deviation increase in each minority share, and how this compares with the standard deviation in death rates. For the case of African Americans, a one-standard-deviation increase in the population share corresponds in April to increased deaths amounting to $(4.894)(12.67) = 62$ per million, which is 21% of the standard deviation of deaths in April of 300.57 (taking the standard deviation of the population share and mortality from rows 3 and 1 of Table 1 respectively). Thus, in this narrow sense, about a fifth of the April variation can be accounted for by this one demographic variable. We can call this the ‘variation ratio,’ and compute it for all four minority groups and both April and May, as reported in the first two columns of Table 3.

From the table it can be seen that the variation ratio is smaller for all groups in May, because there was so much more variance in the deaths in May due to explosion of the pandemic in some states but not others. In both months it is largest by a wide margin for African Americans, because that group has both a large coefficient in Table 2 and a relatively large population variance.

The second way of thinking about the magnitudes is to see what they imply for differential mortality rates. Consider a thought experiment in which there are two demographic groups, numbered 1 and 2, with a mortality rate of m_1 and m_2 respectively, and assume that both of these rates are constants. Group 1 comprises s_c percent of the population in county c , so the average mortality rate is:

$$\bar{m}_c = \frac{s_c m_1 + (100 - s_c) m_2}{100}. \tag{2}$$

If the share s_c varies across counties but the mortality rates m_1 and m_2 do not,⁸ we can estimate $\frac{\partial \bar{m}_c}{\partial s_c}$ by regressing \bar{m}_c on s_c as in (1). If we denote the regression coefficient by β , we can write:

$$\frac{\partial \bar{m}_c}{\partial s_c} = \frac{m_1 - m_2}{100} = \beta, \quad (3)$$

and then we can solve (2) and (3) and compute the ratio $\frac{m_1}{\bar{m}_c}$ for group 1:

$$\frac{m_1}{\bar{m}_c} = 1 + \frac{\beta(100 - s_c)}{\bar{m}_c}. \quad (4)$$

We can call this value the ‘differential mortality ratio’ and compute it for every county. The expression is easy to understand. If $\beta = 0$, so that changes in the minority share have no effect on mortality rates, then $\frac{m_1}{\bar{m}_c} = 1$, and there is no disparity. On the other hand, if s_c is close to 100, so that group 1 is almost the whole population, then again $\frac{m_1}{\bar{m}_c}$ will be close to 1. A large value for the differential mortality ratio will be seen when the group is a relatively small minority (s_c is small) and yet addition of a few new members of that minority has a sizable effect on aggregate mortality (β is large relative to \bar{m}_c).

For both months and all four minorities, the population-weighted median value of this differential mortality ratio is listed in the last two columns of Table 3. From the table we can see that the ratio is well above unity for each case, with the lowest value at 2.1 and the highest at 14.91. These figures are on the high side compared to others in the literature, and as later regressions will show, the coefficients on which they are based should be taken as upper bounds. For example, in figures broken down by race and age group released by the government of New York City,⁹ the ratio of deaths per million for African-American or Hispanic/Latino relative to whites varies from 1.58 to 5.95; in the case of Asians it is close to 1 for adults. Gross et al. report age-corrected relative mortality ratios by state for states

⁸There is very little evidence on whether this assumption holds in practice or not. However, Gross et al. (2020) compute state-level COVID-19 mortality rates for minorities relative to non-Hispanic whites, from those 28 states that report deaths by race. This would correspond to $\frac{m_1}{m_2}$ in this thought experiment. The authors find that the ratio is not correlated with differences in the minority share of population across states.

⁹<https://www1.nyc.gov/site/doh/covid/covid-19-data-deaths.page>

whose data is broken down by race, ranging from a ratio of 18 for the African-American-to-white ratio in Wisconsin down to 0.44 for Pennsylvania. Their aggregate estimate for the US is a ratio of 3.57 for African Americans relative to whites, and 1.88 for Hispanic/Latino relative to white. Price-Haywood et al. (2020) obtained confidential data on 3,481 COVID-19 patients at a hospital in Louisiana and find that 70.6% of those who died were black, while 31% of the population were black, implying a differential mortality ratio $\frac{m_1}{m_c}$ of 2.28.

4.2 Controlling for Income, Poverty, and Education.

We now ask the main question of interest: Are the racial and ethnic disparities established in Tables 2 and 3 caused by underlying economic differences between the different groups? Table 4 adds county median household income to the regression in column (1), the county poverty rate in column (2), and the fraction of the county population with no high-school diploma, with only a high-school diploma, and with some college in column (4). (College graduates are the omitted category.) Surprisingly, neither income nor poverty rates have any visible effect on either the minority variables or in themselves. The coefficients on these variables are small and statistically insignificant. The education variables are similar, but with two differences. First, they reduce the magnitude and eliminate the significance of the Hispanic/Latino and Asian variables. Second, they produce negative, sizable, and large coefficients for the share with Some College, suggesting that an abundance of people in that educational category lowers deaths. But this must be interpreted with care. First, note that by the adding-up constraint on education shares, these coefficients imply that reducing all three shares without a college degree by, say, 5 percentage points, increasing the college-graduate share by 15 percentage points, would *increase* COVID-19 deaths. Possibly the Some College effect has to do with its effect on the prevalence of certain occupations, a point to which we will return soon.

The bottom line is that the African-American and First-Nations effects are robust to these controls, but not the other two categories.

4.3 Insurance and Commuting.

The racial and ethnic disparities in health insurance in the US are well-known. As of 2018, 5.4% of white, non-Hispanic Americans had no healthcare insurance, compared to 9.7% of Black Americans and 17.8% of Hispanics (any race) (Berchick et al. (2019), p.14, using the language for the categories in that report). Since persons with no healthcare insurance are likely to delay treatment or have no access to treatment at all, including for pre-existing conditions that may make them more susceptible to COVID-19 death, a natural question is whether or not the racial disparity in mortality is the result of these racial disparities in insurance coverage, and indeed, the CDC has listed this factor as a prime suspect for the mortality disparity (Centers for Disease Control, 2020). Commenting on the COVID-19 problem, Timothy Freeman, pastor at Trinity African Methodist Episcopal Zion Church in Washington, DC, argues that insurance is a major factor: “I have seen diagnostic tests not performed...and hospitalizations cut extremely short – or not happen at all – because of insurance.” (Johnson, 2020).

Column (1) of Table 5 includes the fraction of county residents who have no healthcare insurance. Surprisingly, the variable has no real effect at all. The coefficients on the uninsured share itself are very small and mostly insignificant, and the coefficients on the minority populations are barely changed from their earlier values, some weakened somewhat and others strengthened. This is despite the fact that there is plenty of variation in the uninsured fraction (as shown in Table 1) and the minority shares in the data are highly correlated with the uninsured fraction.¹⁰ We can conclude that access to healthcare insurance is *not* a driver of the racial mortality disparity.¹¹

Column (2) shows the result of including the fraction of employed workers 16 years of

¹⁰Regressing the uninsured fraction on the four minority shares gives a negative coefficient for the Asian share and a positive coefficient for the other three, all at a 0.1% significance level.

¹¹It could be that this is a result of the nature of this particular disease which has had very limited treatment options. Given that most patients who test positive and are not yet in crisis are mostly advised to rest at home and isolate themselves, this may be an unusual case where health insurance has a limited role.

age and older who travel to work with a carpool, and (3) shows the corresponding result for workers who use public transit. The carpooling variable does not change the racial disparity results much, but the public transit variable has a huge effect. The coefficients are generally cut in half or more; the Asian variables change sign or become insignificant. Further, the coefficients on the public transit variable are large, positive, and statistically significant for April and May. A one-standard deviation increase in the public-transit share in April is associated with an increase of 167 deaths per million, which is more than half the standard deviation of deaths in that month.¹² This is a very large effect. We can conclude that a substantial fraction of the racial disparity in mortality is due to the use of public transit.¹³

It is easy to see with a simple thought experiment why the public transit variable reduces the size of the African-American effect. In a cross-section regression of the public-transit fraction on the four minority population shares (omitted for space), all four minority groups have positive and strongly significant coefficients, although the coefficients are sizable in magnitude only for the African-American share (0.21) and the Asian share (0.63). Apparently all four minority groups use public transit more than the population average. A one-standard-deviation increase in the African-American population share is thus associated with a $12.67 \times 0.21 = 2.66$ percentage-point increase in the fraction who use public transit, which in turn from Table 5 is associated with $2.66 \times 16.74 = 44.54$ additional April deaths per million. This effect through increased use of public transit is more than half the direct effect of the same African-American increase from Table 2, which is $12.67 \times 4.894 = 62$.

Note as well that correcting for public transit, the estimate of the differential mortality ratio from (4) drops to 4.44 in April and 3.23 in May for African Americans and 4.38 in April and 4.63 in May for First Nations people. These figures are close to the direct estimates of

¹²The corresponding negative coefficients for carpooling turn out not to be robust. When the carpooling and public transit variables are added together, the carpooling coefficients become much smaller and insignificant or borderline significant, but the public-transit variables are essentially unchanged.

¹³One might wonder if the public transit variable is simply functioning as a proxy for New York City. Only four counties in the US show a majority of workers getting to work by public transit, and they are all part of New York City. Dropping these from the data barely changes the result. It is still possible that public transit is a proxy for effective local density, a variable emphasized by Desmet and Wacziarg (2020).

Gross et al. (2002) for African Americans for the states that report mortality by race.

A final observation about public transit is that it accounts for much of the difference between mortality in New York City and Los Angeles. Both of these coastal mega-cities have had a large volume of travel with the most COVID-19-afflicted countries, but their experience of the pandemic has been strikingly different. To take Kings County as an example, the county that comprises Brooklyn, COVID-19 deaths per million were 1,628 and 810 in April and May respectively, while for Los Angeles the figures were 72 and 117, an order of magnitude smaller. The fraction who use public transit to get to work was 61% for Brooklyn and 6% for Los Angeles, which by the regression estimates would account for 59% and 78% of the differential mortality between the two cities in the two months respectively.

4.4 Occupational effects.

Table 6 reports the results of controlling for the fraction of workers in the county who can work at home, based on Dingel and Neiman (2020). This variable has barely any effect on the racial-disparity variables, whose coefficients are practically unchanged from column (4) in Table 2. The coefficients on the work-at-home variable itself are small and insignificant. It is likely that this variable will have a much greater effect on the economic impact of the quarantine rather than the incidence of the disease per se, but that question is beyond the scope of this note.

Finally, in Table 7, we add to the basic regression the local employment share of each broad BLS occupational category, interacted with a dummy variable for April and May. Once again, the regression is as in Column (4) of Table 2, but with an additional 42 variables added (22 occupational categories, with Category 11, Managerial Occupations, omitted, making 21 regressors, each interacted with two month dummies). The effects on the minority shares are similar to the last column of Table 4 and are omitted here for space. The effects are strongly significant for African Americans and First Nations shares and mostly insignificant otherwise. The main point of interest in this table is that there are some occupations that

are strongly correlated with local mortality. The most obvious example is Category 31: Healthcare Support Occupations, which includes occupations such as Home Health Aides, Nursing Assistants, and Orderlies. The population-weighted standard deviation of this share is 1.15, so a one-standard-deviation increase in the employment share of that category corresponds to $(1.15)(61.87) = 71.15$ extra deaths per million in April and $(1.15)(34.02) = 39.12$ additional ones in May.¹⁴ It is not surprising that the more healthcare professionals there are in a county the more COVID-19 may spread for a great many reasons, but what may be surprising is that only the support staff have a strong correlation with mortality, and not Category 29, which has the actual doctors and nurses. A similar observation applies to Category 39, Personal Care and Support Occupations, which includes among others barbers, manicurists, and fitness instructors.

We conclude that the occupation make-up does account for the racial and ethnic disparity for Hispanic/Latino and Asian population shares, but not for African-American and First Nations people. Further, it is not the number of people in occupations that allow them to work at home that matters, but rather the presence of certain key occupations.

5 Conclusion.

This note attempts to analyze the socio-economic roots of the widely-noted racial disparity in COVID-19 mortality. The data is county-level COVID-19 mortality data from 3,140 counties together with county economic and demographic data. This has the disadvantage that the county mortality data are not broken down by race, and we cannot tie the socioeconomic status of any one patient to that patient's health outcome. The advantage is that it allows

¹⁴This occupation category can give some insight into why the Hispanic/Latino effect essentially disappears once we control for occupations. In a cross-sectional regression of the Category 31 share on the four minority population shares, the Hispanic/Latino coefficient comes to 0.015. A one-standard-deviation increase in the Hispanic/Latino share is therefore associated with a $17 \times 0.015 = 0.255$ percentage-point increase in the category-31 employment share, which from Table 7 implies an increase of $0.255 \times 61.87 = 15.78$ April deaths per million. This is already much more than enough to account for the April effect of the Hispanic/Latino share in Table 2. The natural interpretation is that Hispanic/Latino Americans are disproportionately represented in healthcare support occupations, and this is a large part of the reason for their correlation with the mortality rates.

for a consistent systematic econometric analysis of the whole country.

The main findings as follows:

(i) For all four minorities, there is a strong positive correlation across counties between the minority's population share and COVID-19 deaths. (ii) For Hispanic/Latino and Asian minorities those correlations are fragile, and largely disappear when we correct for education, occupation, and commuting patterns. (iii) By contrast, for African Americans and First Nations populations, the correlations are very robust. Regardless of what other factors are controlled for, a racial disparity in mortality rates persists. (iv) Surprisingly, for these two groups the racial disparity does *not* seem to be due to differences in income, poverty rates, education, occupational mix, or even access to healthcare insurance, which has been hypothesized by many observers to be a key source of the disparity. (v) A significant portion of the disparity can, however, be sourced to the use of public transit, which also explains a large fraction of the difference in mortality between Los Angeles and New York City.

References

- [1] Berchick, Edward R., Jessica C. Barnett, and Rachel D. Upton (2019). “Health Insurance Coverage in the United States: 2018.” *Current Population Reports* P60-267, US Census Bureau.
- [2] Blinder, Alan. 2009. “How Many US Jobs Might be Offshorable?” *World Economics*, 10:2, pp.41-78.
- [3] Blinder, Alan S., and Alan B. Krueger (2013). “Alternative Measures of Offshorability: A Survey Approach.” *Journal of Labor Economics*, 31(S1), pp.97-128.
- [4] Centers for Disease Control (2020). “COVID-19 in Racial and Ethnic Minority Groups.” <https://www.cdc.gov/coronavirus/2019-ncov/>
- [5] Chowkwanyun, Merlin, and Adolph L. Reed, Jr. (2020). “Racial Health Disparities and Covid-19 – Caution and Context.” *The New England Journal Of Medicine*, May 6, 2020.
- [6] Desmet, Klaus and Romain Wacziarg (2020). “Understanding Spatial Variation in COVID-19 across the United States.” NBER Working Paper No. 27329 (June).
- [7] Dingel, Jonathan I. and Brent Neiman (2020). “How Many Jobs Can Be Done At Home?” NBER Working Paper #26948.
- [8] Fothergill, Alice, Enrique G.M. Maestas, and JoAnne DeRouen (1999). “Race, Ethnicity and Disasters in the United States: A Review of the Literature.” *Disasters* 23:2, pp.156-73.
- [9] Gross, Cary P., Utibe R. Essien, Saamir Pasha, Jacob R Gross, Shi-yi Wang, Marcella Nunez-Smith (2020). “Racial and Ethnic Disparities in Population Level Covid-19 Mortality.” Working Paper, Yale School of Medicine, hosted by medRxiv.

- [10] Hall, William J. , Mimi V. Chapman, Kent M. Lee, Yesenia M. Merino, Tainayah W. Thomas, B. Keith Payne, Eugenia Eng, Steven H. Day, and Tamera Coyne-Beasley (2015). “Implicit Racial/Ethnic Bias Among Health Care Professional and Its Influence on Health Care Outcomes: A Systematic Review.” *American Journal of Public Health* 105:12 (December), pp.e60-e76.
- [11] Hearst, Mary O. J. Michael Oakes, and Pamela Jo Johnson (2008). “The Effect of Racial Residential Segregation on Black Infant Mortality.” *American Journal of Epidemiology* 168:11, pp.1247-54.
- [12] Hoffman, Kelly M., Sophie Trawalter, Jordan R. Axt, and M. Norman Oliver (2016). ‘Racial bias in pain assessment.’ *Proceedings of the National Academy of Sciences*, April.
- [13] Johnson, Akilah (2020). “On the Minds of Black Lives Matter Protesters: A Racist Health System.” ProPublica.org, June 5.
- [14] Price-Haywood, Eboni G., Jeffrey Burton, Daniel Fort, and Leonardo Seoane, M.D. (2020). “Hospitalization and Mortality among Black Patients and White Patients with Covid-19.” *The New England Journal Of Medicine*, May 27, 2020.
- [15] Rubin, Eric J., Lindsey R. Baden, Michele K. Evans, and Stephen Morrissey (2020). “Audio Interview: The Impact of Covid-19 on Minority Communities.” *New England Journal of Medicine*, June 11, 2020.
- [16] Wiemers, Emily E., Scott Abrahams, Marwa AlFakhri, V. Joseph Hotz, Robert F. Schoeni, and Judith A. Seltzer (2020). “Disparities In Vulnerability To Severe Complications From Covid-19 In The United States.” NBER Working Paper 27294 (June).
- [17] Wood, Graeme (2020). “What’s Behind the COVID-19 Racial Disparity?” *The Atlantic*, May 27.

Table 1: Descriptive Statistics.

	Population-weighted mean.	Population-weighted standard deviation.	Population-weighted median.
Population.			
April COVID-19 deaths per million.	135.72	300.57	35.64
May COVID-19 deaths per million.	277.67	482.31	85.43
African-American share.	12.6	12.67	8.16
Hispanic-Latino share.	17.82	17.06	11.2
Asian share.	5.45	6.41	3.39
First-nations share.	0.84	3.27	0.36
Median household income.	62,926.2	17,146.8	60,146
Poverty rate.	14.07	5.14	14.1
Share with less than high school.	12.41	5.41	11.5
High-school graduate share.	27.08	5.41	26.7
Share with some college.	29.08	4.83	29.3
Share with college degree.	31.42	11.04	31.4
No health insurance.	9.37	4.41	8.8
Fraction of workers who use a carpool to get to work.	9.15	2.01	9.05
Fraction who use public transit to get to work.	4.80	10.0	1.60
Fraction of workers who can work from home.	38.70	6.25	38.64

Table 2: Basic regressions.

	(1)	(2)	(3)	(4)
Rest-of-state mortality		0.9165*** (0.1136)	0.9002*** (0.09583)	0.9023*** (0.09551)
African-American share			-0.0001294 (0.0002805)	4.191e-06 (0.0002110)
African-American share*Mar			0.02950* (0.01767)	0.02961 (0.01808)
African-American share*Apr			4.879*** (1.120)	4.894*** (1.132)
African-American share*May			3.874*** (0.6970)	3.760*** (0.6825)
Hispanic-Latino share			-0.0006845* (0.0003943)	-0.0001955 (0.0001418)
Hispanic-Latino share*Mar			-0.005815 (0.01399)	-0.005380 (0.01545)
Hispanic-Latino share*Apr			1.883** (0.7952)	1.952** (0.9277)
Hispanic-Latino share*May			1.374*** (0.4154)	0.9615** (0.4774)
Asian share			0.003918 (0.004383)	0.004811 (0.004797)
Asian share*Mar			0.1119 (0.07004)	0.1126* (0.06797)
Asian share*Apr			6.434** (3.123)	6.546** (3.317)
Asian share*May			4.023** (1.685)	3.259* (1.797)
First Nations share			0.0002912 (0.0002975)	0.0002589 (0.0002752)
First Nations share*Mar			0.009367 (0.007401)	0.009362 (0.007425)
First Nations share*Apr			2.922*** (0.8047)	2.925*** (0.7954)
First Nations share*May			3.977*** (1.399)	4.012*** (1.402)

(... continued)

Table 2: Basic regressions (continued).

	(1)	(2)	(3)	(4)
Population				-9.563e-09* (5.660e-09)
Population*Mar				-8.178e-09 (1.009e-07)
Population*Apr				-1.294e-06 (6.389e-06)
Population*May				8.112e-06 (5.948e-06)
Mar	1.148*** (0.2559)	-0.04557 (0.2765)	-0.9097*** (0.2845)	-0.9157*** (0.2943)
Apr	134.5*** (21.54)	8.128 (6.887)	-122.2*** (30.71)	-123.0*** (31.58)
May	141.9*** (12.43)	20.62 (16.36)	-75.77*** (18.48)	-72.55*** (18.58)
Constant	0.006208 (0.006183)	0.0001142 (0.006641)	-0.007528 (0.01034)	-0.01166 (0.01199)
Observations.	12,560	12,560	12,560	12,560
R^2	0.1256	0.5783	0.6445	0.6455

Table 3: Magnitude of the implied disparity.

	Variation ratio		Differential mortality ratio	
	April	May	April	May
African-American	0.2063	0.0988	10.35	5.57
Hispanic-Latino	0.1157	0.0355	4.41	2.1
Asian	0.1396	0.0433	14.91	5.26
First-nation	0.0318	0.0272	7.43	6.55

Table 4: Income, Poverty, and Education.

	(1)	(2)	(3)
Rest-of-state mortality	0.9061*** (0.1012)	0.9032*** (0.09524)	0.8510*** (0.1023)
African-American share*Apr	4.558*** (0.9503)	4.362*** (0.7910)	4.255*** (0.9723)
African-American share*May	3.862*** (0.6315)	3.706*** (0.5483)	3.162*** (0.5941)
Hispanic-Latino share*Apr	1.809** (0.8815)	1.606** (0.8036)	1.111 (0.8361)
Hispanic-Latino share*May	1.010** (0.4610)	0.9272** (0.4472)	0.6715 (0.5425)
Asian share*Apr	8.468* (5.034)	7.289** (3.522)	6.480 (4.522)
Asian share*May	2.638 (2.606)	3.332* (1.927)	1.681 (2.183)
First Nations share*Apr	2.369*** (0.5493)	1.922** (0.7664)	2.913*** (0.7016)
First Nations share*May	4.205*** (1.359)	3.913*** (1.408)	4.692*** (1.310)
Median household income	7.520e-08 (2.210e-07)		
Median household income*Mar	2.009e-06 (9.727e-06)		
Median household income*Apr	-0.001134 (0.001302)		
Median household income*May	0.0003552 (0.0006946)		
Poverty rate		0.0003386 (0.0007656)	
Poverty rate*Mar		0.006026 (0.02312)	
Poverty rate*Apr		3.123 (2.850)	
Poverty rate*May		0.3172 (1.726)	

(... continued)

Table 4: Income, Poverty, and Education (continued).

	(1)	(2)	(3)
Fraction without highschool diploma			-0.0003876 (0.0007871)
Fraction without highschool diploma*Mar			0.001755 (0.02759)
Fraction without highschool diploma*Apr			3.641 (3.113)
Fraction without highschool diploma*May			0.6144 (2.038)
Fraction with only highschool diploma			0.0009165 (0.0007564)
Fraction with only highschool diploma*Mar			-0.04193 (0.04580)
Fraction with only highschool diploma*Apr			1.038 (2.044)
Fraction with only highschool diploma*May			0.3018 (1.269)
Fraction with some college			-0.0004899 (0.0006472)
Fraction with some college*Mar			-0.02414 (0.02684)
Fraction with some college*Apr			-5.321*** (2.039)
Fraction with some college*May			-7.510*** (2.085)
Constant	-0.01624 (0.01601)	-0.01538 (0.01988)	-0.02253 (0.05317)
Observations.	12,560	12,560	12,560
R^2	0.6471	0.6466	0.6578

Table 5: Health Insurance and Commuting.

	(1)	(2)	(3)
Rest-of-state mortality	0.8922*** (0.1015)	0.8222*** (0.08187)	0.5737*** (0.05203)
African-American share*Apr	4.908*** (1.320)	4.743*** (1.030)	1.800** (0.8139)
African-American share*May	4.196*** (0.8045)	3.623*** (0.5955)	1.877*** (0.4864)
Hispanic-Latino share*Apr	1.947 (1.379)	2.895*** (1.086)	0.8792** (0.3999)
Hispanic-Latino share*May	1.642** (0.6683)	1.569*** (0.5087)	0.1474 (0.2734)
Asian share*Apr	6.603* (3.921)	6.747** (3.184)	-2.334 (1.959)
Asian share*May	2.472 (1.973)	3.344** (1.621)	-1.911* (1.034)
First Nations share*Apr	2.888** (1.203)	4.013*** (1.083)	1.531*** (0.4519)
First Nations share*May	5.018*** (1.431)	4.653*** (1.405)	2.626* (1.442)
Uninsured fraction.	0.001677** (0.0007983)		
Uninsured fraction*Mar	0.06606* (0.03410)		
Uninsured fraction*Apr	0.001387 (4.342)		
Uninsured fraction*May	-4.201 (2.805)		

(... continued)

Table 5: Health Insurance and Commuting (continued).

Fraction who use carpool		0.001269 (0.002381)	
Fraction who use carpool*Mar		0.02230 (0.07090)	
Fraction who use carpool*Apr		-25.77*** (7.481)	
Fraction who use carpool*May		-17.78*** (4.985)	
<hr/>			
Fraction who use public transit			-0.001261 (0.001320)
Fraction who use public transit*Mar			0.04202* (0.02173)
Fraction who use public transit*Apr			16.74*** (2.602)
Fraction who use public transit*May			9.713*** (1.101)
<hr/>			
Constant	-0.02195 (0.01626)	-0.02294 (0.03152)	-0.01503 (0.01423)
<hr/>			
R^2	0.6467	0.6658	0.7777
Observations.	12,560	12,560	12,560
<hr/>			

Table 6: The Ability to Work from Home.

	(1)
Rest-of-state mortality	0.9055*** (0.09836)
African-American share*Apr	4.928*** (1.133)
African-American share*May	3.748*** (0.6909)
Hispanic-Latino share*Apr	1.760** (0.8541)
Hispanic-Latino share*May	1.018** (0.4510)
Asian share*Apr	8.000* (4.550)
Asian share*May	2.845 (2.342)
First Nations share*Apr	2.677*** (0.6521)
First Nations share*May	4.096*** (1.379)
Fraction who can work at home	-0.0001754 (0.0008442)
Work-at-home-able fraction*Mar	0.03786 (0.03829)
Work-at-home-able fraction*Apr	-2.747 (3.005)
Work-at-home-able fraction*May	0.7531 (1.585)
Constant	-0.005200 (0.02213)
Observations	12,560
R^2	0.6470

Table 7: Occupational effects.

	April	May
13: Business and Financial Operations Occupations	10.53 (11.63)	15.83** (7.674)
15: Computer and Mathematical Occupations	-20.12** (8.893)	-5.294 (5.368)
17: Architecture and Engineering Occupations	-17.85 (13.31)	-9.942 (8.136)
19: Life, Physical, and Social Science Occupations	-40.35** (15.75)	-2.052 (10.92)
21: Community and Social Service Occupations	-65.45*** (16.14)	-24.93** (10.16)
23: Legal Occupations	-23.08 (25.27)	20.99 (13.08)
25: Education, Training, and Library Occupations	-16.56* (8.563)	0.2743 (5.062)
27: Arts, Design, Entertainment, Sports, and Media Occupations	26.20 (31.94)	16.55 (13.17)
29: Healthcare Practitioners and Technical Occupations	0.9475 (6.522)	-2.438 (3.380)
31: Healthcare Support Occupations	61.87*** (17.22)	34.02*** (7.731)
33: Protective Service Occupations	-7.755 (8.215)	2.785 (5.001)
35: Food Preparation and Serving Related Occupations	-12.47* (7.552)	-20.54*** (4.656)
37: Building and Grounds Cleaning and Maintenance Occupations	-4.059 (8.013)	3.274 (4.387)
39: Personal Care and Service Occupations	50.66** (20.48)	28.21*** (10.36)
41: Sales and Related Occupations	-14.30** (5.567)	-2.890 (3.139)
43: Office and Administrative Support Occupations	-14.67** (6.784)	-0.3502 (3.615)
45: Farming, Fishing, and Forestry Occupations	-6.966 (5.084)	-5.667** (2.795)
47: Construction and Extraction Occupations	4.580 (6.794)	-0.2213 (3.481)
49: Installation, Maintenance, and Repair Occupations	-12.03** (5.729)	-19.22*** (4.671)
51: Production Occupations	-4.488* (2.551)	0.7983 (1.549)
53: Transportation and Material Moving Occupations	6.103* (3.182)	4.684 (3.323)
Observations.	12,560	
R^2	0.7117	