Association Between Race and COVID-19 Death Rates

Carly McAdam

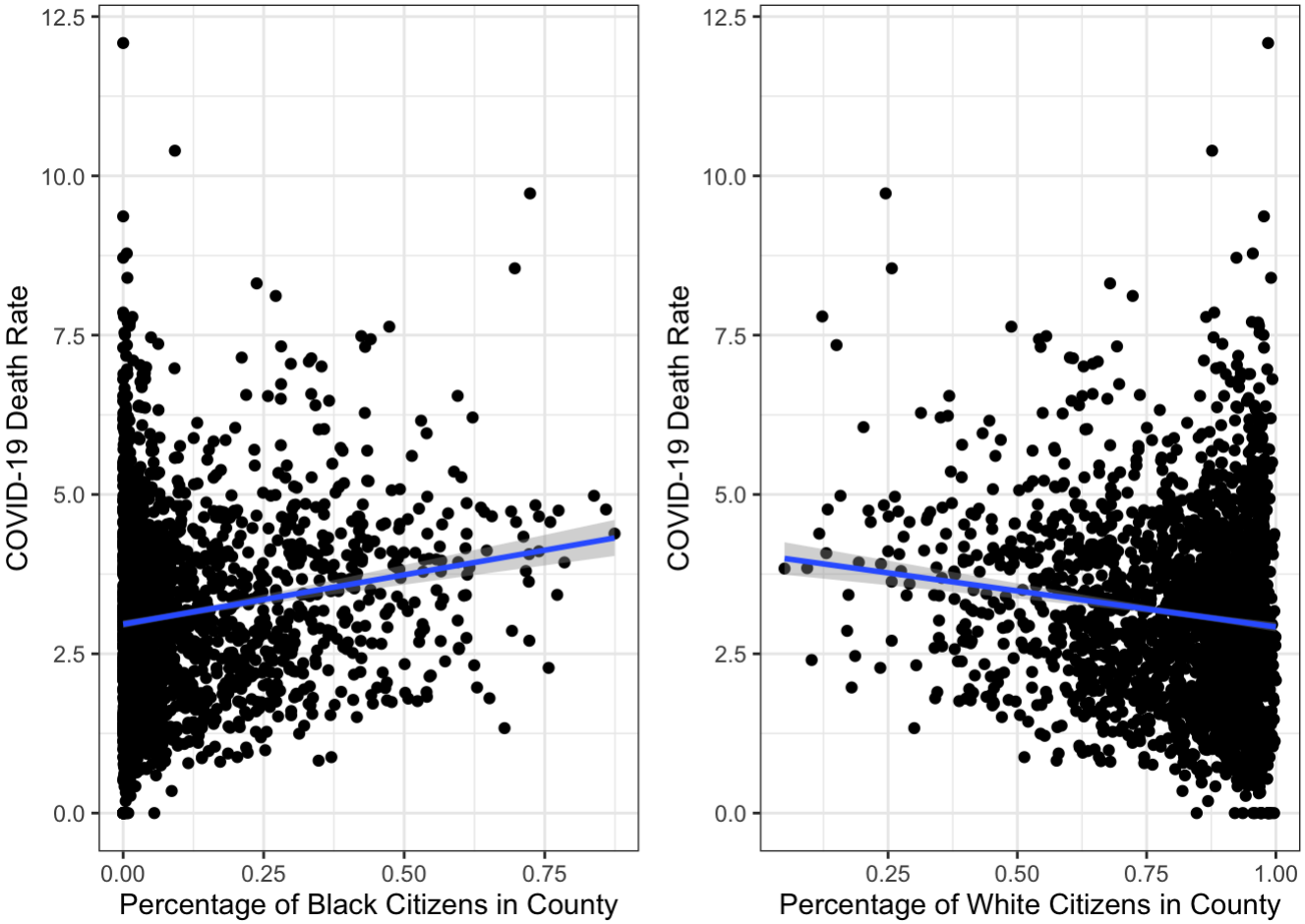
ECON 0210: Economic Statistics

Professor Rao

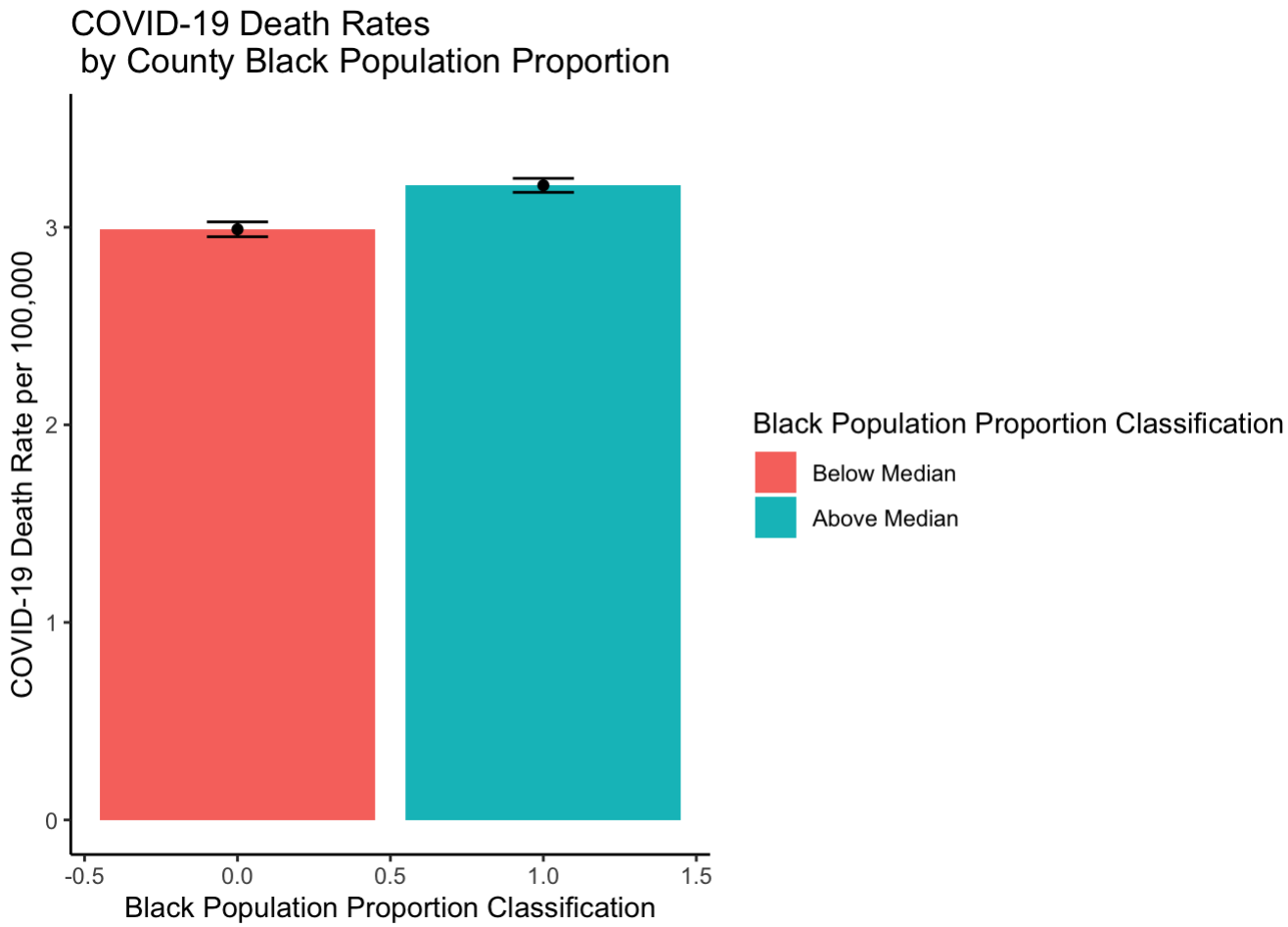
May 2022

The COVID-19 pandemic shed light on many inequalities in the United States, such as income inequality and inequality in access to education and healthcare. Not only did COVID-19 highlight these inequalities, but the trajectory of the pandemic was impacted by them. Even as early as April 2020, states were already reporting statistics suggesting that black people were dying from COVID-19 at disproportionately high rates (New York Times 2020). One of the biggest inequalities in the United States, which is often correlated with income, education, and healthcare access, is racial inequality. I wanted to look at the relationship between race and COVID-19 death rates in the United States. Specifically, the question I am looking to answer is whether counties with higher black population proportions have higher death rates from COVID-19.

I decided to look only at the year 2021, which was the first full year of the pandemic. The county-level demographic data I used comes from the 2020 American Community Survey 5-year estimate (ACS). ACS is an ongoing survey that surveys over 3.5 million households each year and has data for all counties in the United States. My data on COVID-19 comes from the New York Times (NYT) COVID-19 database. NYT has an extensive archive of data on COVID-19 that includes case rates, deaths, vaccination rates, and hospitalization rates for almost every county in the United States. My data on vaccination rates came from the Center for Disease Control and Prevention (CDC). I considered someone “vaccinated” if they had received two shots of the Pfizer or Moderna vaccine or one shot of the Johnson&Johnson vaccine by the end of 2021. After merging data from all of these sources, I had a complete set of data that includes COVID-19 death rate in 2021, vaccination rate at the end of 2021, median income, and population proportions of black and white people for 2,812 counties (93.5% of total counties) in the United States (see appendix for details). Before I get into the specifics of the data, these charts give an overview of trends in the data for death rates and racial makeup of a county. From these graphs, it looks like there could be a relationship between a county having a higher population of black people and a higher COVID-19 death rate.



I also used a t test to look at the difference in death rates for counties with above average black populations and counties with lower than average black populations. Again, it looks like counties with higher black population proportions have higher COVID-19 death rates than those with lower black proportion populations. There is a significant difference in means for the two groups because the p-value is less than 0.05 (see appendix for details). Although we can see that there is a statistically significant difference in means for the two groups, this does not necessarily imply causation. There are many other factors that could lead counties with high black population proportions to have high COVID-19 death rates.



I performed regressions on the relationship between black population proportion and death rate (see appendix for details). I controlled for vaccination rate in each county in both regressions and added a control for income on the second one. The first regression showed that for each 1% increase in black population proportion, the COVID-19 death rate increases by 1.15 deaths per 100,000. Even though my analysis shows a relationship between COVID-19 deaths and race, race is not necessarily a causal variable and there are likely other factors involved in the relationship. Socioeconomic factors, such as median household income, could be a factor in the association between COVID-19 deaths and race. I found a negative correlation between income and race that suggested counties with larger black populations have lower median incomes (see appendix for details). Similarly, I found a negative correlation between income and COVID-19 deaths that suggested counties with higher COVID-19 death rates have lower median incomes (see appendix for details). In my second regression, I controlled for income and found that a portion of the relationship between COVID-19 deaths and black population proportion was due to a difference in median income between counties with high black population proportions and low black population proportions (see appendix for details). This could be because lower income communities have been found to have more limited access to healthcare and live farther away from specialty hospitals (Nguyen et al. 2019). In these counties, someone diagnosed with COVID-19 is less likely to be able to access the care that they need to survive, so death rates are higher. The association between black population proportion and COVID-19 death rate is still positive and statistically significant even with the control for income, however, suggesting that there are likely more factors impacting this relationship.

The correlation that I found between race and COVID-19 death rate is not necessarily and accurate representation of the actual magnitude of the association between the two, due to omitted variables that I do not have data on. There are many variables which I didn't look at that likely have an impact on the relationship between race and COVID-19 death rate, such as health insurance coverage. If health insurance coverage is an omitted variable, it would cause my estimate of the magnitude of the relationship between race and COVID-19 deaths to be too high.

This is because health insurance coverage is most likely negatively correlated with both COVID-19 death rates and race: areas with more health insurance coverage are likely to have lower COVID-19 death rates and areas with higher black population proportions are likely to have less health insurance coverage.

Although race has an important role in causal studies, race is not necessarily a causal variable (Holland 2003). If a black person is found to be at higher risk from COVID-19, it is not simply because they are black that their risk is increased. What is likely happening in the case of COVID-19 deaths and race is that race is interacting with other causal variables that impact COVID-19 death rates. For example, a black person may be more likely to live in an area with less health insurance coverage. Black populations have also been found to have higher rates of comorbid diseases that can put them at higher risk of death from COVID-19 (Chin-Hong et al. 2020). Hypertension, a cardiovascular condition which is associated with worse COVID-19 outcomes, disproportionately affects Black individuals (Chin-Hong et al. 2020). Black people also have a 30-60% higher chance of developing Thromboembolism, a complication from COVID-19 that can lead to death (Chin-Hong et al. 2020). This higher instance of comorbidities and complications could also be related to the relationship between race and access to healthcare.

There are clearly many variables at play here that have an impact on COVID-19 death rates, and many of them may also have a relationship with race. It is challenging to make a definite causal statement about race and COVID-19 death rates, but my analysis shows that there is evidence suggesting that counties with high black population proportions are facing higher rates of COVID-19 deaths and that this may be due to these counties having lower median incomes. Other potential variables that I didn't look at which could be impacting the association are comorbidity rates and healthcare coverage rates.

Statistical Appendix

Correlation between Race and Income, COVID-19 Deaths and Income

```
## [1] -0.2391339
```

```
## [1] -0.4382583
```

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
death_rate	2,812	3.101	1.372	0.000	12.085
vax_rate	2,812	46.936	12.570	0.000	95.000
median_income	2,812	51,419.270	13,174.680	20,188	136,268
white_pct	2,812	0.841	0.160	0.050	1.000
black_pct	2,812	0.088	0.142	0.000	0.874

Table 2: Univariate Regression Results:

(Death Rate and Black Population Proportion)

	Dependent variable:
	death_rate
black_pct	1.548***
	(0.180)
Constant	2.964***
	(0.030)
Observations	2,812

R^2	0.026
Adjusted R^2	0.025
Residual Std. Error	1.355 (df = 2810)
F Statistic	73.718*** (df = 1; 2810)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

Table 3: Bivariate Regression Results

(Death Rate and Black Population Proportion with Vaccine Rate Controlled For)

	Dependent variable:
	death_rate
black_pct	1.151***
	(0.172)
vax_rate	-0.035***
	(0.002)
Constant	4.620***
	(0.097)
Observations	2,812
R^2	0.124
Adjusted R^2	0.123
Residual Std. Error	1.285 (df = 2809)
F Statistic	198.776*** (df = 2; 2809)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

Table 4: Bivariate Regression Results

(Death Rate and Black Population Proportion with Vaccine Rate, Median Income Controlled For)

	Dependent variable:
	death_rate
black_pct	0.473***
	(0.165)
vax_rate	-0.022***
	(0.002)
median_income	-0.00004***
	(0.00000)
Constant	5.991***
	(0.115)
Observations	2,812
R^2	0.230
Adjusted R^2	0.229
Residual Std. Error	1.205 (df = 2808)
F Statistic	279.931*** (df = 3; 2808)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

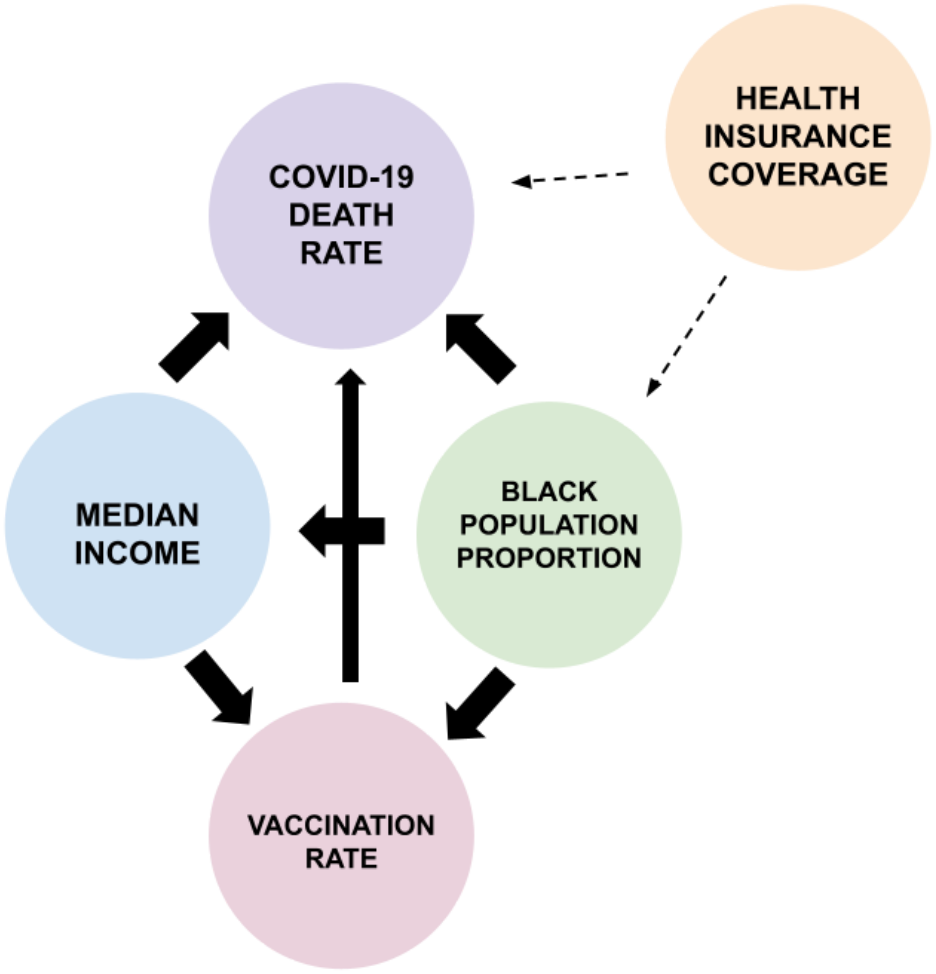
Data Description

1. County demographics (Source: American Community Survey) • This dataset contains demographic data for each county in the United States. There are 3,124 observations and variables include total population,

median income, education measures, black and white population proportions, and poverty levels. The data is a 5-year estimate using ongoing surveys from 2015-2020.

2. COVID-19 Data (Source: The New York Times) • This datasets contains cumulative COVID-19 cases and deaths for the year 2021 for each county in the United States. There is data for each day in 2021 for each county, creating 1,185,373 total observations. NYT gets their data on case rates and deaths from laboratory tests and hospital reports. I calculated the 2021 death rate for each county per 100,000 by dividing the total COVID-19 cases in 2021 by the total COVID-19 deaths in 2021 to get the death rate per 100 and multiplying the result by 1,000.
3. Vaccination Data (Source: Centers for Disease Control and Prevention) • This dataset contains COVID-19 vaccination data for each county in the United States. The data is cumulative and has vaccination data for each day in 2021 for each county, creating 1,680,000 observations. There are 66 variables including data on partially vaccinated individuals, vaccination rates for each vaccine, vaccination data separated by demographic and data on booster doses. I calculated the vaccination rate for each county by dividing the number of people fully vaccinated at the end of 2021 by the total population of the county, which I obtained from the ACS data.

Causal Diagram



How the Data was Merged

```
ACS <- read.csv("acs2018_county_small.csv") #read ACS data
nyt_covid <- read.csv("us-counties-2021.csv") #read NYT data
county_vaccination <- read.csv("COVID-19_Vaccinations_in_the_United_States_County.csv")
#read CDC data
county_vax_clean <- county_vaccination %>%
   filter(Date == "12/31/2021") %>% #take the values for last day of 2021 (data is cumul
ative)
   select(FIPS, Series_Complete_Pop_Pct)
county_vax_clean <- county_vax_clean %>%
   dplyr::rename(vax_rate = Series_Complete_Pop_Pct) %>% #rename variables for merging
   dplyr::rename(fips = FIPS)
clean_covid <- nyt_covid %>%
   group_by(fips) %>%
   filter(date == "2021-12-31") %>% #take the values for last day of 2021 (data is cumul
ative)
   dplyr::rename(total_cases = cases) %>% #rename variables for merging
   dplyr::rename(total_deaths = deaths) %>%
   select(fips, total_cases, total_deaths)
ACS_Covid <- merge(ACS, clean_covid, by = "fips", all = TRUE) #merge
ACS_Covid_Vax <- merge(ACS_Covid, county_vax_clean, by = "fips") %>%
   mutate(death_rate = (total_deaths/pop_total)*1000) %>% #calculate death rate per 100,
000
   select(fips, death_rate, vax_rate, median_income, white_pct, black_pct)
 write.csv(ACS_Covid_Vax, file = "ACS_Covid_Vax.csv") #write a new file for the final da
taset
```

The final merged dataset has 2,812 observations. The unit of observation is county and there are 5 variables: ‘death_rate’, ‘vax_rate’, ‘median_income’, ‘white_pct’, and ‘black_pct’.

Bibliography

Chin-Hong, Peter, Kevin M. Alexander, Norrisa Haynes, Michelle A. Albert, and The Association of Black Cardiologists. “Pulling at the Heart: COVID-19, Race/Ethnicity and Ongoing Disparities.” Nature Reviews Cardiology 17 (September 2020): 533-35.

Eligon, John, Audra D.S Burch, Dionne Searcey, and Richard A. Oppel, Jr. “Black Americans Face Alarming Rates of Coronavirus Infection in Some States.” The New York Times (New York, NY), April 14, 2020. https://www.nytimes.com/2020/04/07/us/coronavirus-race.html (https://www.nytimes.com/2020/04/07/us/coronavirus-race.html).

Golestaneh, Ladan, Joel Neugarten, Molly Fisher, Henny H. Billett, Morayma Reyes Gil, Tanya Johns, Milagros Yunes, Michele H. Mokrzycki, Maria Coco, Keith C. Norris, Hector R. Perez, ShaniScott, Ryung S. Kim, and Eran Bellin. “The Association of Race and COVID-19 Mortality.”EClinicalMedicine 25 (August 2020): 1-7. https://doi.org/10.1016/j.eclinm.2020.100455 (https://doi.org/10.1016/j.eclinm.2020.100455).

Nguyen, Christina A., Michael E. Chernow, Isabel Ostrer, and Nancy D. Beaulieu. “Comparison ofHealthcare Delivery Systems in Low- and High-Income Communities.” The American Journal of Accountable Care 7, no. 4 (December 2019): 11-18. https://www.ajmc.com/view/comparison-of-healthcare-delivery-systems-in-low-and-

highincome-communities (https://www.ajmc.com/view/comparison-of-healthcare-delivery-systems-in-low-and-highincome-communities).