Racial Residential Segregation and COVID-19 Mortality

Suresh Nath Neupane
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/geosciences_theses

Recommended Citation
doi: https://doi.org/10.57709/32375413

This Thesis is brought to you for free and open access by the Department of Geosciences at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Geosciences Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Racial Residential Segregation and COVID-19 Mortality

by

Suresh Nath Neupane

Under the Direction of Katherine Hankins, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2022
ABSTRACT

Studies have shown that the social and physical environments are strong predictors of the health of the urban population. This study investigates if racial residential segregation has any impact on the poor health outcome of residents in light of the COVID-19 pandemic. This is a cross-sectional study with 8,668,744 observations at the individual level. The hierarchical logistic regression conducted to investigate the association between race and residential segregation with COVID-19 mortality showed that a one unit increase in segregation is associated with a 1% increase in mortality. Furthermore, people from Black and Asian ethnic communities were more likely than Whites to die from COVID-19, but Hispanics were less likely to die. This study has limitations such as modifiable aerial unit problem, as county-level segregation indices were used for the analysis.

INDEX WORDS: Race, Residential segregation, COVID-19
Racial Residential Segregation and COVID-19 Mortality

by

Suresh Nath Neupane

Committee Chair: Katherine Hankins
Committee: Erin Ruel
Chetan Tiwari

Electronic Version Approved:

Office of Graduate Services
College of Arts and Sciences
Georgia State University
December 2022
DEDICATION

I dedicate this thesis work to my mother Bhagawati Neupane and my father Kashinath Neupane.
ACKNOWLEDGEMENTS

I take this opportunity to express sincere gratitude to my Chair Dr. Katherine Hankins, and committee members Dr. Erin Ruel, and Dr. Chetan Tiwari for their valuable support in the thesis process. The meticulous reviews and feedback that I received on the content have not only taken this work to a new level but also personally helped me understand research a lot more. I would also like to thank Dr. Ruiyan Luo for her valuable feedback. Lastly, I would like to thank my wife Rukumani Rimal for her encouragement every step of the way.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. V

LIST OF TABLES .................................................................................................................................. VIII

LIST OF FIGURES ................................................................................................................................ IX

1 INTRODUCTION ............................................................................................................................. 1

2 LITERATURE REVIEW ......................................................................................................................... 3

2.1 Roles of Underlying Health Conditions ....................................................................................... 5

2.2 Racial Residential Segregation ..................................................................................................... 7

2.3 Age and Other Factors Driving infections ..................................................................................... 11

3 METHODOLOGY ............................................................................................................................. 13

3.1 Data .............................................................................................................................................. 13

3.2 Constructs .................................................................................................................................... 15

3.3 Residential Segregation ................................................................................................................. 16

3.4 Control Variables .......................................................................................................................... 17

4 ANALYSIS ....................................................................................................................................... 19

4.1 Model ............................................................................................................................................ 19

5 RESULTS ......................................................................................................................................... 22

6 DISCUSSION .................................................................................................................................... 29

7 CONCLUSION .................................................................................................................................. 33

7.1 Threats to Validity and Reliability ............................................................................................... 34
7.2 Limitations and Significance of the Study ................................................................. 34

7.3 Future Research ........................................................................................................ 36

REFERENCES .................................................................................................................. 37
LIST OF TABLES

Table 1 Descriptive Statistics for Individual-Level and County-Level Variables

Table 2 Correlation Matrix of the Predictors

Table 3 Analysis of Variance Comparison Between Models

Table 4 Multicollinearity Diagnostics Table

Table 5 The Hierarchical Logistic Regression Models
LIST OF FIGURES

Figure 1 National Center for Health Statistics 2013 Urban-Rural Classification Scheme for Counties. Non-core counties were removed from the study. (National Center for Health Statistics) .......................................................... 15

Figure 2 The Predictors and COVID-19 death odds ratio ............ Error! Bookmark not defined.
1 INTRODUCTION

The coronavirus disease (COVID-19), which became a global pandemic in early March 2020, has shaken the economic, political, and social orders across the world. Public health agencies first detected 14 of the COVID-19 cases in the United States from January 21 through February 23, 2020, and all cases were related to travel outside the country (CDC, 2020). In nearly two years, the infection and death rates have skyrocketed, with over 865,310 deaths and growing, and more than 70 million confirmed cases in the United States as of January 22, 2022 (Johns Hopkins University, 2022).

Initial studies show counties with a significant Black population have higher COVID-19 infection and death rates, suggesting the disproportionate impact of the disease on racial minorities (Millett et al., 2020; McLaren, 2020; Tai et al., 2021). This disproportionate impact should be unsurprising, as historically, studies have shown that minority populations bear the greatest burden of the impact of a range of diseases (Centers for Disease Control and Prevention, 2013; Mainous et al., 2007). In fact, scholars have documented that socioeconomic status impacts an individual’s risk of contracting diseases (often termed as differential exposure), and also what consequences they face as a result of diseases (termed as differential vulnerability) (Diderichsen et al., 2019). Health inequalities measured using various health outcomes, such as infant mortality, cancer, and diabetes associated with the differential vulnerability and differential exposures to diseases, reflect how social determinants of health have contributed to existing health disparities among minorities (Holmes et al., 2020; Burstrom and Tao, 2020). Spatial settings, such as geographically defined places of residence, are also critical to examine, as they influence one’s life chances (Massey and Denton, 1993). For example, Subramanian et al. (2020) show an association between racial disparities in self-rated health and the geographic
variation in racial segregation, especially in metropolitan areas across the United States. The study found that African Americans were 1.5 times more likely to report poor health conditions than Whites (Subramanian et al., 2005). Since people from minority communities face a higher risk of having one or more chronic illnesses than that of White Americans, it is crucial to examine if the structural disparities, such as residential segregation, are making people of color more vulnerable to COVID-19 mortality (Egede & Walker, 2020). Yancy (2020) argues that although conditions such as hypertension, diabetes, obesity, and cardiovascular diseases prevalent among Black populations may have served as factors affecting COVID-19 infections, factors such as the geographic location of residences, housing density, crime rates, and access to healthy foods, among other structural conditions matter to predict the spread of COVID-19 (Yancy, 2020; CDC, 2020).

The socio-economic conditions in which people live and the social structures, such as policies and health care access, that shape the outcomes help predict the risk factors associated with COVID-19 (Thakur et al., 2020). Ecological studies across cities in the United States have shown associations between segregation and overall poor health outcomes in general (Cooper, 2001; Krivo et al., 2009; Niemesh & Shester, 2020; Hayanga et al., 2013; Hart et al., 1998; Louis-Jean et al., 2020; Kimm et al., 1996; Kramer and Hogue, 2009). Massey and Denton (1993) argue that despite segregation’s important role in creating socioeconomic inequalities among communities, conventional theories have ignored segregation as a predictive factor to such disparities. As such, this study will add to the literature by examining how racial residential segregation is linked to COVID-19 mortality. This study shows if existing structural disparities as manifested through racial residential segregation contribute to the racialized inequities of health with regard to COVID-19.
2 LITERATURE REVIEW

Inequity in health can be defined as disparities in health that are systematically associated with social advantages or disadvantages among different socioeconomic groups (Braveman & Gruskin, 2003; Whitehead & Dahlgren, 2006). Various physical environments, structural factors, as well as socioeconomic status, impact racial minorities for their poor overall health outcomes compared to the White population (See Geron et al., 2022; Javed et al., 2022; Yamoah et al., 2022). Race has thus been used predominantly as a predictor in a majority of public health research over the past few decades, Dressler et al., (2005). Reducing the existing health disparities requires addressing such structural inequities that are also tied to the socioeconomic status of individuals, (Zavala et al., 2021; Islami et al., 2022).

In their systemic review of cancer health disparities, Zavala et al., (2021) show racial-ethnic minorities bearing disproportionate impacts of various forms of cancers in the United States. For instance, the study shows that Black men have the highest lung cancer incidence rate of 71.2 per 100,000 people, compared to rates ranging from 35.1 to 65.3 among other racial/ethnic groups. In terms of overall cancer deaths, Black males and Black females have 19% and 12% higher rates respectively as compared to White counterparts (Islami et al., 2022).

Inequities in health are not only shaped by conditions integral to everyday human life activities, such as work, that impact a range of health and quality of life risks, but also by individuals’ inability to attain full health due to disadvantageous social positions or socially determined influences, such as wealth and educational attainment disparities (Marmot, 2005; Whitehead & Dahlgren, 2006; Braveman et al., 2011; Braveman, 2006). Thus, socioeconomic status (SES), associated with social determinants, shapes the health outcome of individuals (Rollston & Galea, 2020).
Health disparities make socially and economically underprivileged groups of populations such as racial or ethnic minorities more vulnerable to poor health outcomes (Braveman & Gruskin, 2003; Nelson, 2002). In the United States, people from minority populations, including Black, Hispanics/Latinos, American Indians, and Alaska Natives, and Asian Americans, Native Hawaiians, and other Pacific Islanders, face remarkable disparities in the burden of illness and mortality compared to that of the White population (NIH, National Heart, Lung, and Blood Institute 2014). And COVID-19 has been no exception.

COVID-19 made disproportionate impacts on racial minorities on several fronts. Compared to Whites, Black patients with COVID-19 experienced 1.4 times the risk of hospitalization, and 1.36 times the risk of an increased likelihood of dying of the COVID-19 infections than Whites (Poulson et al., 2021; Mackey et al., 2021). Ethnicity is an important predictor of COVID-19 infections, with higher rates of infections observed among Black and Hispanic populations, as well as in communities with higher densities of minority populations (Escobar et al., 2021; Mackey et al., 2021; Yang et al., 2020, Wiemers et al., 2020). Azar et al., (2020) show that the Black population with confirmed COVID-19 cases are 2.7 times more likely to be admitted to the hospital as compared to non-Hispanic Whites after various sociodemographic and clinical factors were controlled. Besides the higher number of hospitalizations, the Black population experienced more severe illnesses caused by COVID-19 than that experienced by other population groups (Azar et al., 2020; Poulson et al., 2021). This finding is consistent with the Selden and Berdahl (2020) study, which shows Black adults in every age group are more likely than White adults to experience health risks associated with
COVID-19 illness. The mortality rates are also higher among essential workers\(^1\) from minority communities than essential workers who are White (Rogers, et al., 2020; Chen et al., 2021).

The minority population is less privileged compared to that of Whites in terms of practicing preventive measures such as social distancing, which is considered the most effective strategy to lower the spread of COVID-19. In addition, studies have confirmed that some minority populations do not get or fully understand the public service messaging due to a lack of language proficiency (Mein, 2020; Milne & Xie, 2020; Thakur et al., 2020). Studies show social distancing interventions, as well as stay-at-home orders, were effective measures to slow the spread of COVID-19 (Delen et al., 2020; Abouk & Heydari, 2021).

In addition, due to the complex nature of the disease to explain to the general public, messaging to minority communities with low socioeconomic status and limited language skills became problematic (Thakur et al., 2020; Karaye and Horney, 2020). In addition to the residential segregation, and SES, other individual factors such as underlying biological conditions that are linked to the structural inequalities make the minority population vulnerable to growing COVID-19 infections.

2.1 Roles of Underlying Health Conditions

Higher prevalence of underlying conditions for COVID-19 such as diabetes and obesity among the hospitalized Black patients indicate disproportionate impacts of the disease on racial minorities (Krishnamoorthy et al., 2021; Clark et al., 2020). Improving basic social determinants of health, such as low SES, poor housing, overcrowding, poverty, limited health knowledge, access to quality and affordable health care, and proper sanitation and infrastructure development

---

\(^1\) The essential services belong to critical sectors such as healthcare, food, and agriculture, and transportation among others, whose operations and services are essential to continue despite the COVID-19 pandemic
pertaining to public health is considered crucial in mitigating challenges of underlying health conditions (Alcendor, 2020; Abrams & Szefler, 2020; Duarte et al., 2021).

Consistent with other studies in the past that show how the Black population is vulnerable to comorbidities, such as hypertension and diabetes compared to that of the White population, a clinical research study recently demonstrated that the Black population is more likely to experience adverse lung problems due to COVID-19 infections (Joseph et al., 2020; Kopel et al., 2020; CDC, 2019). Likewise, Azar et al., (2020) found that advanced age and underlying health conditions of individuals were likely to increase the risk of COVID-19 illness. There was a larger increase in the risk of Black mortality due to COVID-19 infection than that of Whites and Asians (Golestaneh et al., 2020; Poulson et al., 2021; Bryan et al., 2021; Rushovich et al., 2021). The increased mortality risk in COVID-19 cases has been driven by the pre-existing health conditions that are often linked with other social determinants of health, which makes people of color susceptible to poor health outcomes (Williamson et al., 2020; Anderson-Carpenter & Neal, 2022).

The longstanding health disparities such as diabetes, hypertension, chronic obstructive disease, and pulmonary disease that are significantly prevalent among minorities put them at risk for clinically severe COVID-19 (Alcendor, 2020). The compromised immune system is driven by the ever-lasting structural poverty and food insecurity among Blacks, compared to Whites (Holmes et al., 2020). Lower socioeconomic status (SES) leads the minority population to reduced access to health care, which results in multiple comorbidities, and poor diets, all combined factors weaken one’s immune system, thus making them vulnerable to COVID-19 (Kopel et al., 2020). Since underlying health conditions and COVID-19 mortality are strongly associated, reducing racial health disparities, thus necessitates targeting improvements in
socioeconomic conditions, not just at an individual but at the geographic level such as areas of racial residential segregation (Williams & Collins, 2001).

2.2 Racial Residential Segregation

Racial and socioeconomic disparities shape one’s access to health care and facilities that impact individual health conditions. The communities that have been structurally discriminated against or underprivileged bear the extra burden of COVID-19. The persistence of residential segregation in U.S. cities despite legislative progress made shows that it is crucial to understand its impact on the minority population, which has been bearing the disproportionate burden of not just economic marginalization but also poor health outcomes.

Racial residential segregation has long characterized urbanization patterns in the United States (Meyer, 2000). In the late nineteenth and early twentieth century, various migrations of Black Americans alongside discriminatory practices rooted in white supremacy shaped the residential distribution of races, thus contributing to growing ethnic segregation across cities in the United States (Meyer, 2000; Massey, 2009). The purposeful action of spatially, socially, and economically marginalizing the Black population forced Black residents to migrate to city cores, thus creating conditions for potential Black ghettoization during much of the twentieth century (Massey, 2015). Decades of overt discrimination practices, enforced through social or economic policies that favored middle-class Whites, have continued to make race a dominant feature for residential segregation of communities across the United States, (Jargowsky, 2020; Massey & Denton, 1993). The prevalence of exclusionary tactics such as denying full access to housing markets through discriminatory practices used by landlords and mortgage lenders, especially against the Black population, has contributed to the continued racial segregation of the urban population (Maly, 2011; Jargowsky et al., 2014; Massey & Denton, 1993). For instance,
Rothstein (2015) argues how discriminatory government policies such as racially explicit zoning that contributed to the production of the urban ghetto, restrictive covenants, which excluded the Black population from White areas, and segregated public housing projects, among others, implemented in the late 20th century, led to the persistence of racial segregation in St. Louis, as in many other cities.

Racial segregation rose slightly in metro areas across the United States during the early period of the 1970s, with the Black population highly likely to be in segregated neighborhoods than any other minority group such as Hispanics and Asians (Massey, 2001). Glaeser & Vigdor (2012) show the metro areas began to see a decline in segregation since the 1970s and by 2010 it was at the lowest level in nearly a century. For instance, the isolation indices in metropolitan areas such as Chicago, Cleveland, Detroit, Gary, New York, and Newark were 80 or more in 1990, suggesting that most Black people lived in communities that were over 80 percent Black (Massey, 2001). A study comparing the 50 largest cities, which included the top Metropolitan Statistical Areas by population in 1950 and 2018, shows that metros in the United States have changed from 90% White in 1950 to 60% White in 2018, suggesting that rapidly changing racial composition further helping produce segregated areas (Tracy Hadden Loh et al., 2020).

Residential segregation is associated with a larger system of inequality that has impacted minorities, who have been affected by a slow decrease in health and SES gaps, despite the passage of the civil rights legislation and Fair Housing Act of 1968, (Krysan & Crowder, 2017; Kramer and Hogue, 2009). Segregation is a cause of racial differences in socioeconomic status in that segregation determines access to education and employment opportunities; making residential segregation a primary cause of racial health disparities (Williams & Collins, 2001). Racial segregation limits access to opportunities, especially for people from minority
communities, to remain healthy (Schulz et al., 2002). Besides impeding minorities from accessing education, and job opportunities, racial segregation also shapes conditions such as exposure to poverty and poor housing, which adversely affect individual health, thus confining minorities to disadvantaged places in urban spaces (Williams & Collins, 2001; Massey & Denton, 1993).

2.3 Socioeconomic Status and Income Inequality

Inequalities in health begin early in life as one is exposed to such unfair conditions that are perpetuated through institutional policies that can be seen prevalent in discrimination in employment, education, and other economic opportunities (Fiscella, and Williams, 2004). The two intertwined concepts—health inequalities, referring to disparities or differences in health outcomes among groups such as race, and gender; and health inequities, which refers to structural inequalities stemming from prevalent injustice—shape the population health (Kawachi et al., 2002). Such inequalities and inequities are linked to SES and race that are used as crucial measures in predicting population health outcomes (Fiscella and Williams, 2004). While income is an important variable used to examine the association with individuals as well as area-level health, an ecological variable, income inequality, has been used as a measure to probe the impact on population health (Pickett & Wilkinson, 2015).

The fundamental reason for associating individuals’ socioeconomic status (SES) with causes of disease is because of the former’s role in building capability to access resources that can either be used to avoid potential risk factors or control the consequences of diseases (Link and Phelan, 1995). The level of economic inequality results in the production of poverty and thus undermines the social structures that support health (Raphael, 2000). A research study on health services and social and economic characteristics of particular urban spaces help understand the
particular cause of health problems and the prevalence of health inequalities and inequities among various race and ethnicity (Illsley, 1990; Pickett & Wilkinson, 2015; Williams & Collins, 2001).

The social class or SES is used to measure the individual health of such groups, thus linking individual characteristics to larger groups to examine the advantages or disadvantages of the social characteristics, (Honjo, 2004, also see Deaton & Lubotsky, 2003). The association between SES, often measured by income and poverty level among other determinants, and mortality rates have been used not only at the individual levels but also while measuring the geographic level such as counties or census tracts (Adler et al., 1993; McCord and Freeman, 1990; Deaton & Lubotsky, 2003).

As studies show mixed effects of areas income inequality on various outcomes of population health, it is crucial not to ignore income inequality as a predictor while considering potential confounding effects (See Matthew & Brodersen, 2018; Lynch et al., 2004; Fiscella & Franks, 1997; Lochner et al., 2001; Lobmayer and Wilkinson, 2002). This study uses income inequality as an independent variable to help explain the relationship between racial segregation and COVID-19 death among various racial-ethnic groups. The study measures the ratio between the median household income 80th percentile to that of the 20th percentile in a county as an index for income inequality. For instance, the 80th percentile is the level where 20% of the households have higher incomes, while the 20th percentile represents the 20% of the population. Higher the index values, the greater the gap between individuals living on the top and bottom end of the group. The data was based on the American Community Survey (ACS) 2015-2019 estimate. The positively skewed income inequality variable was log10-transformed.
2.4 Age and Other Factors Driving infections

Other contributing factors that this study will control for are age and population density. A majority of research supports that increased COVID-19 mortality risks are found among elderly population groups over 70 years of age (Li et al., 2020). While more data were available later during the pandemic, the initial data from San Francisco, California, in mid-April 2020 showed the young Latinx population was disproportionately impacted by COVID-19, with 42% of the cases among individuals below 40 years of age and that 24% of them were Hispanic or Latinx (Haynes and Cooper, 2020). Although studies show high mortality rates among the population of 65 or older age group, there are patterns of people of color dying at younger ages (Bassett et al., 2020). Having age as a control variable in the study would provide an unbiased examination of segregation on COVID-19 mortality. Age-related and place-based factors also contribute to placing older Black adults at risk for the infection and potential deaths caused by COVID-19 (Chatters et al., 2020).

Yehia et al., (2020) argue that in the absence of implementation of any social distancing measures, people’s living conditions such as densely populated areas, multigeneration households, underlying health conditions, and health issues that emerged due to lack of access to care, may also explain the disproportionate impacts of COVID-19 on individuals (Yehia et al., 2020). Unlike underlying health conditions and risk factors that health officials can clearly identify, predictors associated with economic, structural, and social disparities such as limited access to healthy foods, housing density, and high crime rates, among others need more exploration (Yancy, 2020; Haynes and Cooper, 2020; Watson et al., 2020).

This paper examines whether there is any disparity in the impact of the COVID-19 pandemic on racial groups in segregated areas. The research question is What is the effect of
racial residential segregation on the racial disparities in mortality caused by Covid-19? In order to examine the relationship between racial residential segregation and COVID-19 mortality, this paper proposes a hypothesis concerning the effects of structural inequalities on COVID-19 infections and death rates.

Hypothesis:

1. Black individuals are more likely to die of COVID-19 than are Whites and other races.

2. Racial residential segregation will explain and therefore attenuate the covid 19 mortality gap between Whites and Blacks.

3. Black individuals living in segregated counties are more likely to die of COVID-19 than Blacks living in less segregated counties.
3 METHODOLOGY

This is a cross-sectional study intended to measure if racial residential segregation is a predictor of mortality caused by COVID-19. The objective of this study is to examine the association of COVID-19 death as the outcome variable, and racial residential segregation, gender, age, population density, and income inequality as predictors.

3.1 Data

Individual-level data with demographic variables, such as race, gender, age, and COVID-19 death, were used for this study. The dataset was provided by the Center for Disease Control and Prevention (CDC)’s Case Surveillance Task Force and Surveillance Review and Response Group (SRRG). The restricted data were made available for limited use upon completion of the registration information and data use restrictions agreement (RIDURA). This SRRG public use dataset has variables including demographics, COVID-19, underlying conditions, outcomes such as death caused by COVID-19, and county and state of all COVID-19 cases.

The dataset starts from January 1, 2020, as the earliest date of the COVID-19 case reported, until April 15, 2021. The dataset may include some vaccinated individuals as the United States Food and Drug Administration (FDA) issued the first emergency use authorization for use of the Pfizer-BioNTech COVID-19 vaccine on December 11, 2020. While teachers, school staff, and childcare workers were deemed eligible for vaccines by early March 2021, by mid-April the eligibility was expanded to the general public (U.S. Department of Health & Human Services, 2020). The dataset was released on April 30, 2021. Access to the data was given via CDC’s GitHub portal for a limited period of time.

In order to avoid large central, fringe, and medium metropolitans with high-level of segregation masking over non-core regions, rural areas were removed from the dataset. A total of
1,341 nonmetropolitan areas with their respective FIPS code that are also known as non-core with a population less than 10,000, according to the National Center for Health Statistics classification scheme of 2013, were removed from the dataset (National Center for Health Statistics, 2019). Figure 1 shows the non-core rural areas classified as #6 on the NCHS dataset. The counties included in the analysis were large central metro, large fringe, medium metro, and small metro. The CDC dataset with 23,723,384 observations was used for the study. First, a total of 1,271,137 observations that belong to non-core rural areas, were removed from the original dataset. When rural counties were removed, 8.5% (a total of 663,946 observations) of the White population and 4.3% (a total of 72,652 observations) of Black population were removed from the original dataset. Similarly, 1.2% of Hispanics and 0.12% of Asians belonging to rural counties were removed from the sample. In terms of non-responses, Race variable had the most missing values. A total of 7,857,055 such missing observations were removed from the dataset. Omitting the missing values, White and Black race had a 49% reduction in the sample size each, followed by Asians at 37% and Hispanic at 36%, compared to the original dataset. After removing overall missing values, the final sample size was reduced to 8,668,744 for the analysis.

In the original CDC dataset, responses that had been left unanswered (blank) were re-classified to an “Unknown” value and such values were treated as “missing” and removed while conducting the final analysis. For instance, in the question “Did the patient die as a result of this illness?,” with responses “Yes,” “No,” or “Unknown,” the “Unknown” values were removed. Similarly, “Unknown” response values in other variables were also treated as missing values and removed.
To measure the impact of COVID-19 on individuals living in racially-segregated areas, this study used death caused by COVID-19 as a response variable while treating variables such as residential segregation, race and income inequality as predictors, controlling for age and population density across counties in the United States.

Dependent Variable: This study used death among confirmed COVID-19 cases as a dependent variable. The responses on the individual-level dataset were coded with Yes =1 and No=0 to represent COVID-19 infected individuals who died or survived respectively for all 8,668,744 observations. Of the total observations, 346,982 were entered as “Yes.”
3.3 Residential Segregation

Racial residential segregation is the independent variable that this study used to examine if segregation can explain the death caused by COVID-19 infections. The county-level segregation index was used from the publicly available Robert Wood Johnson Foundation’s (RWJF) web portal (Robert Wood Johnson Foundation, 2021). The residential segregation between the two groups—non-White and White races in the study is measured using the dissimilarity index, which shows greater segregation between the two groups with higher values. The county-level indices from RWJF’s dataset were assigned to the respective individual-level observations by the county-FIPS (Federal Information Processing System) codes information. Even though the segregation index was on a county-level, it was assigned to individual-level observations.

The dissimilarity index (D) is interpreted as the proportion of one particular race that would have to change their tract of residence to make an equal distribution of the different racial groups (Duncan & Duncan, 1955). This index measures if Whites or Blacks or Others, which included American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander—is distributed across census tracts in the county in a similar fashion. The D is measured at a county level and is based on the distribution of non-Whites across the counties as units of analysis. The value of D ranges from 0, a complete integration to 100, complete segregation. For instance, a score of 0 indicates both groups—Whites and non-Whites—are completely integrated across all census tracts in the county, whereas higher scores indicate a higher level of segregation among census tracts within that county. The segregation index variable was mean-centered before the analysis in order to improve the interpretation of model parameters.
The dissimilarity index below computes the proportion of minorities that need to move to achieve an even distribution:

\[ D = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{P_{1i}}{P_1} - \frac{P_{2i}}{P_2} \right| \]

*P_1 & P_2*: county-wide population Group 1 or Group 2

*P_{1i} & P_{2i}*: population of Group 1 or Group 2 in the neighborhood I out of n neighborhoods

Since focusing on cities rather than counties would limit the analysis within the city boundaries, county-level segregation was chosen to appropriately include rural areas. The dissimilarity index was calculated to compare the urban and rural counties across the United States. This study acknowledges the limitations as put forth by Winship (1977) that dissimilarity index does not address the concerns with respect to inaccurate comparison between random segregation, the pattern of choice of residence by households without any regard to racial composition, and the actual segregation in the county.

### 3.4 Control Variables

Age and population density serve as control variables for this study. To control for the race and age-specific impacts on mortality, age groups was used. Similarly, the population density for the county was used as a control variable.

The age was divided into two groups to examine if older age groups, defined as 60 and older, can explain high mortality to COVID-19. The cut-off age of 60 years old age was made based on the CDC data that stratified ages 10 years apart, starting from 0-9 years to 80+ years old. Having age as a control variable is important with respect to COVID-19 mortality because studies in the recent past have shown mixed results (see Haynes and Cooper, 2020; Selden and
The population density across counties in the United States was calculated based on an average number of people per square mile of the land area using the most recent data from the American Community Survey (ACS) (2014-2018) five-year estimate. The positively skewed population density variable was log10-transformed before the analysis.

Race is classified into five groups—Whites, Blacks, Hispanic, Asians, and Others (which include American Indian, Alaska Native, Asian, Multiple/Other, Native Hawaiian/Other Pacific Islander, and Hispanic/Latino). Dummy coded, each race is treated as a variable, and interaction terms were applied using different races and segregation index. The race and segregation index interaction terms act as explanatory variables to examine if the outcome variable of death caused by COVID-19 can be predicted by individuals belonging to particular racial groups living in segregated areas.
4 ANALYSIS

A hierarchical logistic regression (HLR) was conducted to examine the correlation between the race and segregation index with COVID-19 mortality or survival. The multilevel regression shows if residential segregation, a county-level variable, is associated with the individual-level outcome of mortality, adjusting for both county and individual-level confounding factors. This study used HLR with a binomial response variable, death caused by COVID-19. Interaction terms between race and segregation were used to examine if individuals from a particular racial group living in the segregated area are likely to survive or die of COVID-19.

The HLR model with random county effects is also used to satisfy the violation of assumptions of independence that are made in basic regression models that the health outcomes of individuals living in the same area (or county) may be correlated (Hubbard et al., 2010). I used the lme4 package in R statistical tool for fitting the random county Model

The model has a Bernoulli distribution:

\[ Y \sim Bi; (1, \pi) \]

where \( Y = \)COVID-19 death and \( \pi \) is the probability of death.

A random intercept model:

\[
\text{Logit}(\pi_{ij}) = \beta_{0ij} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij} + \beta_7 X_{7ij} \quad (1)
\]

where \( \beta_{0ij} \sim N(\beta_0, \sigma^2) \)

where \( i = 1, \ldots, n_j \) denotes individuals within counties and \( j = 1, 2, \ldots N_{ct} \), where \( N_{ct} \) is the number of counties. Death is an outcome variable (Y), and predictors are Whites (X1), Black (X2), Hispanic (X3), Asian (X4), and Others (X5), age below 60 years old (X6), and age over 60 years old (X7).
The level-two variables include segregation \((Z_j)\), White*segregation\((X_1*Z)_j\), Black*segregation \((X_2*Z)_j\), Hispanic*segregation \((X_3*Z)_j\), Asian*segregation \((X_4*Z)_j\), and Other race--others*segregation \((X_5*Z)_j\), and population density \((X_8)_j\) is represented in the county-level equation mentioned below:

\[
\text{Logit} (\pi_{ij}) = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij} + \beta_7 X_{7ij} + \beta_8 Z_j + \beta_9 X_1*Z_j + \beta_{10} X_2*Z_j + \beta_{11} X_3*Z_j + \beta_{12} X_4*Z_j + \beta_{13} X_5*Z_j + \beta_{14} X_{6j} + \beta_{15} X_{7j} + \beta_{16} X_{8j} \tag{2}
\]

where \(\beta_{0j} \sim N(\beta_0, \sigma^2)\)

The continuous measure of segregation index for county \(j\) would be the same for each individual \(i\) within the same county. The level-1 represents variables with individual-level data such as race and age, and level-2 represents the county-level segregation index and population density and interaction terms between race and segregation.

In the HLR model, the level-2 equations (2) treat the coefficients from level-1 equations as random variables and the values to be predicted from county-level characteristics of residential segregation.

Based on the two equations, three models—Model 1 with individual-level variables, gender, race and age, Model 2 with individuals and county-level variables that included race, age, segregation, income inequality, population density, and Model 3 with, in addition to variables from Model1 and 2, predictors such as race and segregation interaction terms were used to fit the generalized mixed-effects models for fixed and random effect models. The county FIPS code in the dataset was used to have a random effect in model 2 with the ‘binomial’ family function in the R package, \textit{lme4}.

The two models below show the final model fit using ‘gmler’ function from \textit{lme4} package in R:
Model 1 <- glmer(death ~ (1|county)+ gender+ black + hispanic + asian + others + over60, family = binomial("logit"), nAGQ=1, data=data1)

Model 2 <- glmer(death ~ (1|county)+ gender+ black + hispanic + asian + others + over60 + segregation + income_ineq+ popdensity, family = binomial("logit"), nAGQ=1, data = data1)

Model 3 <- glmer(death ~ (1|county)+ gender+ black + hispanic + asian + others + over60 + segregation + income_ineq+ popdensity + black_segregation + white_segregation + hispanic_segregation + asian_segregation + others_segregation, family = binomial("logit"), nAGQ=1, data = data1)

The model family=logit describes the logistic type of conditional distribution of the response given the random effects. The hierarchical logistic regression model will have a logit link function to show the relations between the predictors and the response variable. The parameter of interest in the model for the logistic regression is the probability between death (1) and survival (0). The model with the dependent variable death, a binary outcome, is predicted by the sum of explanatory variables on the right with (1|county) used to specify the county random effects. The random effects show how the explanatory variables have the same random county effects across the responses. The COVID-19 death as a response variable has a binomial distribution. The predictors used to predict mortality are segregation, race, age, interaction terms of race and segregation, and population density.
5 RESULTS

Descriptive statistics for all variables used in the study are shown in Table 1. The largest proportion of confirmed COVID-19 patients were White, representing 46%, Hispanic 33%, Black 10%, Asian 4%, and the rest are identified as ‘other races’ in the sample. While age ‘below 60 years’ old represents 79% of the total observations, 21% of the sample are aged over 60 years old. The death proportion for people over the age of 60 years was 0.17, whereas for people below the age of 60 years the proportion was 0.01. The data show Blacks with a slightly higher proportion of death, at 0.054, followed by Whites at 0.051. Asians’ proportion of death was 0.045, while Hispanics and Other race were at the lowest at 0.024. The data also shows that males have a higher proportion of COVID-19 deaths than females, with 54.6% of total deaths. Table 1 also shows the White population being the largest sample (46%), representing more than 58.35% of the total deaths, followed by Hispanic, Black, and Asian at 19.45%, 13.48%, and 4.05% respectively. Of the total sample size of 8,668, 744 confirmed COVID-19 cases, 346,982 died, making the overall death rate at 4%. Also, the mean segregation, measured by dissimilarity index (D) ranging 0-100, was at 0, after the variable was mean-centered. The standard deviation for the segregation index was 9.91. Similarly, the average population density of the counties represented in the sample population was 1444 per square mile with a standard deviation of 3243. The average index value for income inequality was 4.72 with 0.646 standard deviation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (Std. Dev./Percentage)</th>
<th>Death Proportion</th>
<th>% Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>0.04 (0.196)</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>47%</td>
<td>0.05</td>
<td>56.6</td>
</tr>
</tbody>
</table>

2 Hispanic/Latino
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>53%</td>
<td>0.03</td>
<td>45.4</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>46%</td>
<td>0.051</td>
<td>58.35</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>10%</td>
<td>0.054</td>
<td>13.48</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>33%</td>
<td>0.024</td>
<td>19.45</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>4%</td>
<td>0.045</td>
<td>4.05</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>8%</td>
<td>0.024</td>
<td>4.66</td>
<td></td>
</tr>
<tr>
<td>Over 60 yrs</td>
<td>21%</td>
<td>0.17</td>
<td>87.45</td>
<td></td>
</tr>
<tr>
<td>Below 60 yrs</td>
<td>79%</td>
<td>0.01</td>
<td>12.55</td>
<td></td>
</tr>
<tr>
<td>Pop. Density</td>
<td>1444</td>
<td>3243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segregation</td>
<td>0(9.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White*Segregation</td>
<td>0.082</td>
<td>6.995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black*Segregation</td>
<td>0.423</td>
<td>3.647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian*Segregation</td>
<td>-0.018</td>
<td>1.467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others*Segregation</td>
<td>0.004</td>
<td>2.762</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income Inequality</td>
<td>4.72</td>
<td>0.646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8668744</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Correlation Matrix of the Predictors

<table>
<thead>
<tr>
<th></th>
<th>Death</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Others</th>
<th>Over60</th>
<th>Below60</th>
<th>Pop Density</th>
<th>Segregation</th>
<th>White*Segregation</th>
<th>Black*Segregation</th>
<th>Asian*Segregation</th>
<th>Others*Segregation</th>
<th>Income_ineq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>1</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.06</td>
<td>-0.18</td>
<td>-0.02</td>
<td>0.34</td>
<td>-0.34</td>
<td>0.05</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.05</td>
</tr>
<tr>
<td>White</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop Density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White*Segregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black*Segregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian*Segregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others*Segregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income_ineq</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 8668744
Table 3 Multicollinearity Diagnostics Table

<table>
<thead>
<tr>
<th>Term</th>
<th>VIF</th>
<th>Increased SE</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>1.32</td>
<td>1.15</td>
<td>0.76</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.16</td>
<td>1.07</td>
<td>0.87</td>
</tr>
<tr>
<td>Asian</td>
<td>1.07</td>
<td>1.04</td>
<td>0.93</td>
</tr>
<tr>
<td>Others Race</td>
<td>1.13</td>
<td>1.06</td>
<td>0.89</td>
</tr>
<tr>
<td>Over 60 Yrs Old</td>
<td>1.02</td>
<td>1.01</td>
<td>0.98</td>
</tr>
<tr>
<td>Segregation</td>
<td>1.14</td>
<td>1.07</td>
<td>0.88</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>1.06</td>
<td>1.03</td>
<td>0.95</td>
</tr>
<tr>
<td>Pop. Density</td>
<td>1.11</td>
<td>1.05</td>
<td>0.90</td>
</tr>
<tr>
<td>Black*Segregation</td>
<td>1.31</td>
<td>1.15</td>
<td>0.76</td>
</tr>
<tr>
<td>Hispanic*Segregation</td>
<td>1.09</td>
<td>1.04</td>
<td>0.92</td>
</tr>
<tr>
<td>Asian*Segregation</td>
<td>1.03</td>
<td>1.02</td>
<td>0.97</td>
</tr>
<tr>
<td>Others*Segregation</td>
<td>1.12</td>
<td>1.06</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Of the three models, Model 1 and Model 2 are nested within Model 3. Model 1 includes Gender, Black, Hispanic, Asian, Others, age ‘Over 60 years’ variables. Model two adds Segregation, Income Inequality, and Population Density to the variables included in Model 1. Model 3 adds interaction terms, Black*Segregation, Hispanic*Segregation, Asian*Segregation, and Others*Segregation variables in addition to the variables in Model 2. White race being the largest group, and age ‘below 60 years’ were used as reference categories. Model 1 shows that the Black population is 20% more likely than Whites to die of COVID-19, controlling for other variables. Second to Black patients, Asians had a 15% higher chance of dying of the virus than Whites. The likelihood of females was consistent over all three models with 32% less likely than males to die of COVID-19. Hispanics were 16% less likely to die of COVID-19 than Whites.

In Model 2, in addition to demographic variables from Model1, variable of interest, segregation, and others such as income inequality, and control variable population density all
measured at a county-level were added. Model 2 saw some changes on demographic variables. The likelihood of COVID-19 death for Blacks and Hispanics increased by one percent points, while for Asians, the likelihood for mortality decreased by nine percent points. There was a slight one percent point increase in the likelihood of mortality for females in Model 2. There was a statistically significant relationship between residential segregation and COVID-19 mortality. The results showed that as racial residential segregation increased by one unit, the likelihood of dying of COVID-19 increased by 1% on average. No association between income inequality and COVID-19 death was found. The model shows that a one unit increase in population density is associated with a 38% increased probability of dying of COVID-19.

In Model 3 there were slight changes in the demographic variables. Black populations’ likelihood of dying from COVID-19 was 20% higher (95% CI: 1.16 – 1.20) than Whites, which is the same as in Model 1 but one percent point less than Model 2. Hispanics were 14% less likely (95% CI: 0.83 – 0.85) to die of COVID-19, compared to whites, and Asians had 7% greater likelihood (95% CI: 1.12 – 1.17) of dying of COVID-19 than Whites, controlling all the other variables. Likewise, for Others race category the odds ratio was 0.68 (95% CI: 0.60 – 0.63), making 32% less likelihood of dying of COVID-19 than Whites, controlling for other variables. The main effect of racial residential segregation on COVID-19 mortality suggests that a one unit increase in segregation is associated with a 1% increase in the probability of mortality for whites, controlling for other variables. The interaction terms show that the effect of segregation on COVID-19 mortality varies across the racial groups although that is true only for Asians. For Blacks, Hispanics, and other racial groups, the confidence intervals suggest that segregation operates similarly as for Whites. The results showed Black living in areas with high non-White-White segregation, the likelihood of dying of COVID-19 was equally likely for that
of Whites (O.R. 1, 95% C.I.: 1:00-1:00). The exact same results were observed for Hispanics as well with no group effects on segregation. Model 3 shows that as segregation increases by one unit, the probability of mortality declines for Asians by 1%.

Same as in Model 1, and 2, income inequality showed no association with COVID-19 mortality. The population density showed similar results as in the previous two models, with areas of high population density, the likelihood of COVID-19 mortality would increase by 38%. Similarly, consistent with a widely established risk factor, this model also showed a strong correlation between people over 60 years old, who were 30.87 times likely (O.R. 30.87, 95% CI: 29.73 – 30.4) to die of COVID-19 death (see Williamson et al., 2020; Woolf et al., 2021).

### Table 4 The Hierarchical Logistic Regression Models

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>CI</td>
<td>OR</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>4.592</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>-0.016</td>
<td>0.01</td>
<td>-0.016</td>
</tr>
<tr>
<td>Gender [F]</td>
<td>-0.386</td>
<td>0.68</td>
<td>-0.386</td>
</tr>
<tr>
<td></td>
<td>-0.004</td>
<td>0.67 – 0.69</td>
<td>-0.004</td>
</tr>
<tr>
<td>Black</td>
<td>0.181</td>
<td>1.2</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td>-0.007</td>
<td>1.18 – 1.21</td>
<td>-0.007</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.170</td>
<td>0.84</td>
<td>-0.170</td>
</tr>
<tr>
<td></td>
<td>-0.007</td>
<td>0.83 – 0.85</td>
<td>-0.006</td>
</tr>
<tr>
<td>Asian</td>
<td>0.138</td>
<td>1.15</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>-0.014</td>
<td>1.12 – 1.17</td>
<td>-0.011</td>
</tr>
<tr>
<td>Others</td>
<td>-0.463</td>
<td>0.63</td>
<td>-0.463</td>
</tr>
<tr>
<td></td>
<td>-0.011</td>
<td>0.62 – 0.64</td>
<td>-0.011</td>
</tr>
<tr>
<td>Over 60</td>
<td>3.403</td>
<td>3.403</td>
<td>29.73 – 30.4</td>
</tr>
</tbody>
</table>
Segregation 

Income Inequality 

Pop Density 

Black*Segregation 

Hispanic*Segregation 

Asian*Segregation 

Others*Segregation 

Random Effects 

\[
\begin{align*} 
\sigma^2 & \quad 3.29 & \quad 3.29 & \quad 3.29 \\
\tau_{\text{county}} & \quad 3.95 & \quad 3.87 & \quad 3.87 \\
\text{ICC} & \quad 0.55 & \quad 0.54 & \quad 0.54 \\
N & \quad 1475 & \quad 1475 & \quad 1475 \\
\text{Observations} & \quad 8668744 & \quad 8668744 & \quad 8668744 \\
\text{Marginal } R^2 & \quad 0.216 / 0.644 & \quad 0.223 / 0.643 & \quad 0.223 / 0.643 \\
\end{align*}
\]

Table 5 Analysis of Variance Comparison Between Models

<table>
<thead>
<tr>
<th>Model</th>
<th>np</th>
<th>ar</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>8</td>
<td></td>
<td>1819477</td>
<td>1819589</td>
<td>-909731</td>
<td>1819461</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>11</td>
<td></td>
<td>1819460</td>
<td>1819613</td>
<td>-909719</td>
<td>1819438</td>
<td>23.68</td>
<td>3</td>
<td>2.91E-05***</td>
</tr>
<tr>
<td>Model 3</td>
<td>15</td>
<td></td>
<td>1819245</td>
<td>1819455</td>
<td>-909607</td>
<td>1819215</td>
<td>222.82</td>
<td>4</td>
<td>&lt; 2.2e-16***</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Comparing the three models, Model 1 and 2 are nested with full model, Model 3. Since these models are nested, the lowest value of -2 Log-Likelihood, which is Model 3, is the best fit model at a statistically significant p-value of <0.0001. The best fit, Model 3 was used for the analysis. Table 2 also shows that all of the predictors have the variance inflation factor (VIF), which detects multicollinearity in regression analysis, values less than 10. Thus, it appears that the model has no multicollinearity problem.

*Figure 2 The Predictors and COVID-19 death odds ratio*
6 DISCUSSION

This study examined the relationship between racial residential segregation and COVID-19 death to contribute to the existing literature that suggests that racial residential segregation is an important factor in the prevalence of health inequalities (Williams & Collins, 2001; Schulz et al., 2002). A study of the COVID-19 pandemic’s impact on racial minorities living in segregated communities was important to examine if such disparities were actually manifested in communities. Since the only geographic unit of the dataset was a county, the segregation index was used at a county level. This study showed that racial residential segregation is significantly associated with higher COVID-19 deaths. The positive correlation between segregation, a county-level variable assigned to individuals with COVID-19 cases, and death caused by the COVID-19 shows how confinement of any disadvantaged racial group to residentially segregated communities impact the individual as well as population health and wellbeing (Massey & Denton, 1993).

Supporting Hypothesis 1, this study showed that the Black population was most likely to die of COVID-19 disease than Whites and other races. With a 20% more likelihood of dying of COVID-19 compared to Whites, the study found that the Black population faces the highest burden of COVID-19 fatalities when compared to White or other race categories. Also, compared to Whites, this study found that Hispanics had a 14% lower likelihood of dying of COVID-19. This result supports the literature on the so-called “Hispanic Paradox,” which suggests that despite their lower socioeconomic status, studies show Hispanic communities have
an advantage over various health outcomes compared to other minorities or even Whites (Markides & Coreil, 1986; Crimmins et al., 2007; Markides & Eschbach, 2005). Palloni and Asias, (2004) found that the ‘Hispanic paradox’, the advantages, was found only among foreign-born Mexicans not Cubans, Puerto Ricans and others. Further study is necessary to be more precise on this paradox, rather than just generalizing the whole community (Palloni & Arias, 2004). Second to Blacks, this study found that another racial minority, Asians were 7% more likely than Whites to die of the virus, controlling for other variables.

Hypothesis 2 states that segregation will reduce the COVID-19 mortality gap between Whites and Blacks. The hypothesis suggests that segregation increases the risk for all races with respect to COVID-19 mortality. For instance, are White populations living in segregated areas more likely than Whites living in less segregated or non-segregated areas to die of COVID-19? In other words, does the death gap remain slimmer between the White and Black population in segregated counties? The results did not support the hypothesis. Compared to Whites, the Black population living in segregated counties have no association with COVID-19 mortality as the odds ratio for interaction terms between Black and segregation is 1 (95% CI: 1.00-1.00). As Model 3, showed that a 1 unit increase in non-White-White segregation is associated with a 1% increase in COVID-19 mortality controlling for all other variables.

Paradoxically, this study showed that people belonging to a particular race such as Black and Asians, and living in a higher level of residential segregation were less likely to die of COVID-19, while the association between segregation and the death from the virus itself was positive. The result suggested that segregation has the same impact on Whites, Blacks, and Hispanics with respect to COVID-19 mortality, hence Hypothesis 3 is not supported. The study also showed that the Hispanic population and segregation had no effect on COVID-19 mortality.
For Asians, there was a 1% lower likelihood of dying of COVID-19 when living in segregated as opposed to non-segregated or less segregated areas. Segregation creates conditions that are unfavorable to population health and thus assumed to contribute to health disparities, but this study, which found no association between COVID-19 death and Black and Hispanic races living in segregated counties, suggests that more confounding factors are needed to be considered when studying such relationships (Williams & Collins, 2001).

Measuring the impact of segregation on urban or individual health is complex, especially, from the study design perspective, as it is critical not to ignore various confounding effects that determine one’s access to social, economic, and political resources (Schulz et al., 2002). Besides individual accounts of health, factors such as social structural inequalities as seen in residential settlements, education, and employment status have shown how minorities have been underprivileged from improving their life chances (Rogers, 1992; Pearl et al., 2001). There have been multiple studies in the past that show the health effects of racial residential or other forms of segregation but addressing the precise impact has remained challenging (see Kramer & Hogue, 2009; Greer et al., 2014).

Studies have shown disparities in COVID-19 health outcomes among racial minorities also resulted from limited testing of the disease owing to issues such as lack of access to health care and mistrust in the medical system (Andraska et al., 2021; Lieberman-Cribbin et al., 2020; Mody et al., 2021). Besides insufficient testing, underreporting and undetected cases of COVID-19 were also prevalent in many countries, including the United States (Lau et al., 2021; Tatar et al., 2021). For instance, Shen et al. (2021) show that 40% of COVID-19 deaths that occurred prior to the beginning of the federal reporting system in nursing homes in May 2020, were not reported (also see Perniciaro & Weinberger, 2021). Early studies suggested a large percentage of
excess deaths\(^3\) across the U.S. counties, with 87.5\% nationally, attributed to COVID-19 since March 2020 (Ackley et al., 2022; Shiels et al., 2021; Rossen et al., 2020; Center for Disease Control and Prevention, 2022).

\(^3\) Defined as the difference between the observed and the expected number of deaths in a given time period
7 CONCLUSION

Studies have shown that the social and physical environments are strong determinants of the health of the urban population (Tonnel et al., 2021; Ompad et al., 2008). This study examines if racial residential segregation has any impact on the poor health outcome of residents in light of the COVID-19 pandemic. This study found that racial residential segregation is a predictor of COVID-19 death and that racial minorities, with mixed results, bear the burden of the impact of the coronavirus that has claimed the lives of millions of people across the world.

There is a considerable number of studies examining the association between segregation and various health outcomes at the individual as well as groups levels (See Kramer and Hogue (2009). Adding to the literature, this study found that people living in highly segregated counties were 1% more likely to die of COVID-19 death than people living in less segregated or non-segregated counties. Among the racial minorities, the Black population faced the greatest burden of COVID-19 death compared to Whites. After Blacks, Asians were more likely to die of COVID-19 than Whites. However, as segregation increased in counties, COVID-19 deaths declined for Asians compared to whites. controlling for other variables such as age, income inequality, and population density. Hispanics were less likely to die from the COVID-19, compared to Whites when age, income inequality, and population density were controlled. This finding supports the existing literature on the narrative of the “Hispanic paradox,” which states that despite their minority status and lower SES, Hispanics paradoxically were less likely than Whites to die of COVID-19 at a statistically significant level. Further, a multidimensional approach, as Dannenberg et al. (2003) suggest, which incorporates factors such as environmental, socioeconomic, and health care access and behavioral dimensions is crucial to study population health.
7.1 **Threats to Validity and Reliability**

Examining the association between diseases and race is challenging work. There could be many confounding factors that influence measures across various racial groups. Salgado et al., (2020) emphasize a multidimensional approach, incorporating factors such as environmental, socioeconomic, and health care access and behavioral dimensions deemed crucial to study population health. While a wide range of studies has shown the effects of underlying health conditions on COVID-19, there remains a possibility of misconstrued analysis because of the impacts of confounders on infection and death rates caused by the disease. This is a cross-sectional study thus a one-time measurement may be difficult to draw a causal relationship, which makes threats to internal validity imminent.

Since cities or counties across the states have varying demographic information along with varying rules imposed on lockdowns and social distancing measures, the correlation between variables and COVID-19 mortality may not have an equal effect on other geographical locations.

7.2 **Limitations and Significance of the Study**

Factors related to human behaviors such as mobility and social activities contribute to the risk of spreading of COVID-19 across counties with varying lockdowns and restrictions. This study does not include analysis pertaining to such human behaviors that may have caused or exacerbated the infections, which resulted in hospitalizations as well as mortality.

The dataset used in this study has a large number of missing data. As discussed above, the original dataset contained 23,723,384 observations, which was reduced to 22,452,247 when
rural counties were removed from the sample. Race/ethnicity one of the primary variables had 7,857,055 missing values. Removing individual non-response observations, the final dataset lost 63.44% of the observations.

Examining the relationship between race and racial residential segregation, and COVID-19 mortality at a county level to make an inference on health disparities also makes the study subject to ecological fallacy. Given the scarcity of individual-level data with geographic units smaller than county-level on the COVID-19 pandemic, using the available county-level data to measure the health inequalities is the best-suited approach. This study also has a modifiable aerial unit problem or MAUP, potentially bound to biases due to the use of spatially aggregated data such as the segregation index. Acevedo-Garcia et al., (2002) argue that although most of the studies examining the relationship between residential segregation and mortality or health outcome have methodological weakness because a study showing such association is ‘atheoretical’ and the study of the relationship has recently become an area of exploration (Acevedo-Garcia et al., 2002). Ecological studies, using counties or smaller units of analysis such as census tracts, as a strategy to examine the area effects on health, have been used to investigate mortality rates to narrate the variability to characteristics of a particular area (Diez Roux, 2001). From analytical perspectives, social-ecological analyses of health have expanded the traditional biomedical or biopsychological concepts, which incorporate the transdisciplinary models of health over the past century (Stokols, 2017).

It is crucial to identify the phenomena of the spread of the disease and how it has impacted people living in racially segregated neighborhoods. While the COVID-19 pandemic may come to an end or infection and mortality rates may significantly decline eventually, the general trend of higher morbidity and mortality among ethnic minorities will likely remain in the
future with persistent disparities in health outcomes, (El-Khatib et al., 2020). As more research studies are coming in on COVID-19, it would be incomplete to argue that one particular factor plays a role in health outcomes. More studies are needed to explain the disparities in COVID-19 infections, hospitalization, and mortality.

7.3 Future Research

With the availability of individual-level data, future studies on COVID-19 could reveal more phenomena on health outcomes among minority populations on an individual level. Another interesting future research on the issue would be to measure the economic impacts of the COVID-19 pandemic on ethnic communities.
REFERENCES


https://doi.org/10.1016/j.ribaf.2020.101281


https://doi.org/10.1001/jamasurgery.2013.408

https://doi.org/10.1161/CIRCULATIONAHA.120.048126


https://doi.org/10.3390/ijerph17124322


https://doi.org/10.3389/fpubh.2020.00418


https://doi.org/10.1086/597285


https://doi.org/10.1080/13557858.2020.1830036
