

# Segregation Predicts COVID-19 Fatalities in Less Densely Populated Counties

Becky Li<sup>1</sup>, Ryan J. Quinn<sup>2</sup>, Salimah Meghani<sup>3</sup>, Jesse L. Chittams<sup>2</sup>, Vijay Rajput<sup>4</sup>

Received 01/03/2022

Review began 01/06/2022

Review ended 01/13/2022

Published 01/17/2022

© Copyright 2022

Li et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. School of Medicine, Nova Southeastern University Dr. Kiran C. Patel College of Allopathic Medicine, Fort Lauderdale, USA 2. Biostatistics, University of Pennsylvania School of Nursing, Philadelphia, USA 3. Biobehavioral Health Sciences, University of Pennsylvania School of Nursing, Philadelphia, USA 4. Medical Education, Nova Southeastern University Dr. Kiran C. Patel College of Allopathic Medicine, Fort Lauderdale, USA

Corresponding author: Becky Li, bl896@mynsu.nova.edu

---

## Abstract

### Aim

It is well known that social determinants of health (SDoH) have affected COVID-19 outcomes, but these determinants are broad and complex. Identifying essential determinants is a prerequisite to address widening health disparities during the evolving COVID-19 pandemic.

### Methods

County-specific COVID-19 fatality data from California, Illinois, and New York, three US states with the highest county-level COVID-19 fatalities as of June 15, 2020, were analyzed. Twenty-three county-level SDoH, collected from County Health Rankings & Roadmaps (CHRR), were considered. A median split on the population-adjusted COVID-19 fatality rate created an indicator for high or low fatality. The decision tree method, which employs machine learning techniques, analyzed and visualized associations between SDoH and high COVID-19 fatality rate at the county level.

### Results

Of the 23 county-level SDoH considered, population density, residential segregation (between white and non-white populations), and preventable hospitalization rates were key predictors of COVID-19 fatalities. Segregation was an important predictor of COVID-19 fatalities in counties of low population density. The model area under the curve (AUC) was 0.79, with a sensitivity of 74% and specificity of 76%.

### Conclusion

Our findings, using a novel analytical lens, suggest that COVID-19 fatality is high in areas of high population density. While population density correlates to COVID-19 fatality, our study also finds that segregation predicts COVID-19 fatality in less densely populated counties. These findings have implications for COVID-19 resource planning and require appropriate attention.

---

**Categories:** Infectious Disease, Public Health, Epidemiology/Public Health

**Keywords:** barriers to healthcare, residential segregation, covid-19, health disparities, social determinants of health

## Introduction

The COVID-19 pandemic is predicted to widen the health gap for minorities with pre-existing inequalities, such as Black, Hispanic, and Native American populations [1,2]. Minorities and low-income populations are vulnerable due to inadequate healthcare access, fewer opportunities to clarify misinformation (due to reduced access to high-quality information channels), and susceptibility to comorbidities [1,3].

The disproportionate consequences of COVID-19 are exacerbated by the inability of minorities and low-income families to maintain adequate social distancing, as they constitute a great portion of the frontline and essential workforce and reside in densely populated homes [1]. Frontline occupations include nurses, delivery workers, and others who could not work from home. In the United States (US), 41.2% of frontline workers identify as a person of color, and more than a third of the frontline workers are supporting low-income families [4]. We, therefore, investigated social determinants of health (SDoH) as they are associated with disparities in COVID-19 transmission and outcomes [5].

Notably, the current literature investigating SDoH and COVID-19 together at the county level is limited, particularly in methods of analysis. Previous studies have identified county-level risk factors affecting COVID-19 susceptibility and mortality utilizing bivariate or regression analysis [6-10]. Only one study has used the tree-based machine-learning analytical method but focused on county-level COVID-19 incidence in a single state [11]. Our research is centered on contextualizing county-level data in COVID-19 outcomes

### How to cite this article

Li B, Quinn R J, Meghani S, et al. (January 17, 2022) Segregation Predicts COVID-19 Fatalities in Less Densely Populated Counties. Cureus 14(1): e21319. DOI 10.7759/cureus.21319

rather than susceptibilities. Employing 23 county-level SDoH and a unique tree-based analytical lens, our study aimed to identify the key county-level social determinants of COVID-19 fatality in the midst of the first COVID wave in three US states: New York, Illinois, and California.

## Materials And Methods

### Study population and methods

We selected three US states, namely California, Illinois, and New York, containing counties with the highest absolute COVID-19 fatalities as of June 15, 2020 [6]. Total cumulative COVID-19 cases and fatalities were gathered from state department of health websites in June 2020 [12-14]. County attributes were simultaneously collected from County Health Rankings & Roadmaps (CHRR) which assembled widely used county-level data from publicly available datasets [15-18]. Population density was measured in persons per square mile [19,20]. Residential segregation was studied as an index of dissimilarity between white and non-white populations on a scale of 0 (integration) to 100 (segregation) [21]. Preventable hospital stays were defined as the rate of hospital stays for conditions that can be treated as outpatient, per 100,000 Medicare enrollees. This study was granted exemption by the Nova Southeastern University Institutional Review Board (IRB) for not involving human subjects per the federal regulations (IRB #2020-229). Data analysis was conducted in August and September 2020.

### Statistical analysis

Twenty-three county-level predictors of interest were analyzed through summary statistics. The outcome measure was COVID-19 fatality rate calculated as deaths per 100,000 county population. Inclusion criteria was implemented to only include counties with a population density between 1.5 to 50,000 persons per square mile (n=198). Bivariate associations between SDoH and continuous COVID-19 fatality rate were assessed with Spearman correlations. A median split performed on the population-adjusted COVID-19 fatality rate created a high/low fatality indicator used in subsequent predictive modeling using decision tree analysis.

The decision tree was generated using the HPSPLIT procedure in SAS (Statistical Analysis System; SAS Institute, Cary, USA) to visualize associations among SDoH and to explore the profile of counties most at risk for high COVID-19 fatality. We selected decision tree analysis for its robustness, ease of interpretation, and simplification of complex relationships seen in SDoH [22]. Tree analyses are built from known data and subsequently utilized to predict future outcomes [23]. In certain analyses, decision trees have been superior to logistic regression in predicting case outcomes, particularly outcomes that behave in a non-linear fashion [24].

The decision tree method employs machine learning techniques to determine a parsimonious model, defining profiles which best classify counties by outcome status [22]. Tree building starts at the root node, which contains all the data (county fatalities), and is partitioned recursively into child nodes until it reaches its terminal nodes. The split is based on selecting the predictor that best discriminates between high and low fatality counties [23]. Ten-fold cross-validation was employed for pruning and validation of the final tree [25]. Model accuracy was evaluated based on sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Sensitivity analyses confirmed variable selection results. In a follow-up analysis, Kruskal-Wallis tests were applied to analyze how counties in certain model-defined profiles differed from other counties based on predictors of interest.

## Results

Table 1 presents summary statistics on variables of interest. The median fatality rate was 4.5 deaths per 100,000 county population. Spearman correlations identified seven county-level SDoH exhibiting a moderate association ( $r>0.3$ ) with the continuous COVID-19 fatality rate (Table 2). Upon further analysis, the decision tree model identified population density, residential segregation, and preventable hospitalizations to be key predictors of counties with high COVID-19 fatality rates (Figure 1). The tree had four total terminal nodes reflecting county profiles with a color characterization of low COVID fatality (blue) or high COVID fatality (pink). The model area under the curve (AUC) was 0.79, with 74% sensitivity, 76% specificity, and 25% misclassification rate. The average sensitivity, specificity, and misclassification rates among cross-validation subsamples were 64%, 54%, and 40%, respectively.

Measure	Overall median (IQR)	CA median (IQR)	IL median (IQR)	NY median (IQR)
COVID-19-related outcomes				
COVID-positive cases <sup>a</sup> (# total positive cases as defined by each state department of health)	131 (854)	254 (1941.5)	45 (211)	253.5 (1766.5)

COVID-19 fatalities <sup>a</sup> (# fatalities/100,000 county population)	4.5 (33)	4 (43)	1.5 (15)	21 (90.5)
Sociodemographic factors				
Population density <sup>b</sup> (# persons per square mile estimated from 2019 population and 2010 land area)	87.9 (200.17)	111.81 (380.88)	56.64 (86.8)	120.91 (376.86)
Elderly <sup>c</sup> (% adults ages 65 and older)	18.8 (4.5)	15.8 (7.45)	19.3 (2.9)	18.45 (2.55)
Black (% non-Hispanic Black or African American)	2.55 (5.8)	2 (3)	3.55 (7.1)	4 (7.1)
Non-Hispanic White <sup>c</sup> (%)	82.9 (26.9)	48.55 (37.55)	90 (12.8)	86.7 (17.9)
Female <sup>c</sup> (%)	50.3 (1.1)	50.2 (1.3)	50.35 (1)	50.4 (1.3)
Rural <sup>c</sup> (%)	40.85 (47)	15.15 (35.75)	44.4 (36.3)	46.75 (41.65)
High school graduation rate <sup>c</sup> (% ninth-graders who graduate in 4 years)	86 (7)	85 (6)	88.5 (8)	85 (4.5)
Some college <sup>c</sup> (% adults ages 25-44 with some post-secondary education)	62.5 (11)	60 (17)	63 (8)	62 (9.5)
Unemployment rate <sup>c</sup> (% people age ≥ 16 unemployed but seeking work)	4.6 (1.3)	4.65 (2.8)	4.8 (0.9)	4.35 (1)
Income inequality <sup>c</sup> (ratio of 80th percentile income to 20th percentile income)	4.5 (0.6)	4.7 (0.5)	4.35 (0.7)	4.5 (0.6)
Residential segregation <sup>c</sup> (index of dissimilarity ranging 0-100, used in the American Community Survey where higher values indicate greater residential segregation between non-White and White county residents)	33 (18)	25 (10.5)	36.5 (16)	39 (15.5)
Food insecurity <sup>c</sup> (%)	11 (3)	12.5 (3.5)	11 (3)	11 (2)
Limited access to healthy foods <sup>c</sup> (% low-income population who do not live close to grocery store)	5 (4)	5 (5)	5.5 (4)	4 (3)
Physical environment				
Severe housing problems <sup>c</sup> (% households at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities)	15 (10)	23 (4)	11 (4)	15 (4)
Homeowner percentage <sup>c</sup> (% occupied housing units that are owned)	71 (12)	61 (9)	75 (7)	71 (6.5)
Severe housing cost burden <sup>c</sup> (% households that spend ≥ 50% household income on housing)	13 (7)	18 (3)	10 (4)	14 (4)
Drive alone to work <sup>c</sup> (%)	80.5 (7)	76 (7.5)	83 (4)	80 (6)
Average traffic <sup>c</sup> (traffic volume per meter)	152 (258)	215 (439.5)	89 (100)	296 (594.5)
Health behaviors				
Physical inactivity <sup>c</sup> (% adults aged ≥ 20 who report no leisure-time physical activity)	25 (6)	21 (9)	26 (6)	26 (5)
Clinical care				
Uninsured <sup>c</sup> (% persons under age 65 without health insurance)	6 (2)	8 (3)	6 (1)	5 (1)
Primary care physicians (ratio of county population to PCP)	1933 (1376)	1433 (1091.5)	2271.5 (1320)	1816 (1329)
Preventable hospital stays <sup>c</sup> (rate of hospital stays for ambulatory-care sensitive conditions per 100,000 Medicare enrollees in a year)	4440.5 (1649)	3307.5 (1077)	5011.5 (1821)	4477.5 (1098)
Flu vaccinations <sup>c</sup> (% fee-for-service [FFS] Medicare enrollees that had an annual flu				50.5

vaccination)

46 (10) 41 (8.5) 44 (9) (4.5)

**TABLE 1: Descriptive statistics on county attributes and COVID-19 outcomes**

CA - California; IL - Illinois; NY -- New York; IQR - interquartile range; PCP - primary care physicians

<sup>a</sup>Data was retrieved on June 15, 2020 from the California, Illinois, and New York Department of Health [12-14]<sup>b</sup>Data was retrieved on June 15, 2020 from US Census Bureau [19,20]<sup>c</sup>Data was retrieved on June 15, 2020 from County Health Rankings & Roadmaps [16-18]

Social determinant	R value <sup>a</sup>
Population density <sup>b</sup> (# persons per square mile)	0.54486
Average traffic <sup>c</sup> (traffic volume per meter)	0.46935
Black (% non-Hispanic Black or African American)	0.44544
Residential segregation <sup>c</sup> (index of dissimilarity ranging 0-100, used in the American Community Survey where higher values indicate greater residential segregation between non-White and White county residents)	0.43384
Flu vaccinations <sup>c</sup> (% fee-for-service [FFS] Medicare enrollees that had an annual flu vaccination)	0.42261
Rural <sup>c</sup> (% population)	- 0.36288
Elderly <sup>c</sup> (% adults ages 65 and older)	- 0.34550
Food insecurity <sup>c</sup> (%)	- 0.26115
Primary care physicians <sup>c</sup> (ratio of county population to PCP)	- 0.24212
Unemployment rate <sup>c</sup> (% people age ≥ 16 unemployed but seeking work)	- 0.23788
Severe housing cost burden <sup>c</sup> (% households that spend ≥ 50% household income on housing)	0.23565
Female <sup>c</sup> (%)	0.22523
Non-Hispanic White <sup>c</sup> (%)	- 0.22470
Some college <sup>c</sup> (% adults ages 25-44 with some post-secondary education)	0.21744
Severe housing problems <sup>c</sup> (% households at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities)	0.18491
Income inequality <sup>c</sup> (ratio of 80th percentile income to 20th percentile income)	0.18212
Limited access to healthy foods <sup>c</sup> (% low-income population who do not live close to grocery store)	- 0.12510
Uninsured <sup>c</sup> (% persons under age 65 without health insurance)	0.09888
Preventable hospital stays <sup>c</sup> (rate of hospital stays for ambulatory-care sensitive conditions per 100,000 Medicare enrollees in a year)	0.09763
Homeowner percentage <sup>c</sup> (% occupied housing units that are owned)	- 0.09586

Drive alone to work <sup>b</sup> (%)	-
	0.08880
Physical inactivity <sup>c</sup> (% adults aged ≥ 20 who report no leisure-time physical activity)	-
	0.06643
High school graduation rate <sup>c</sup> (%)	0.03359

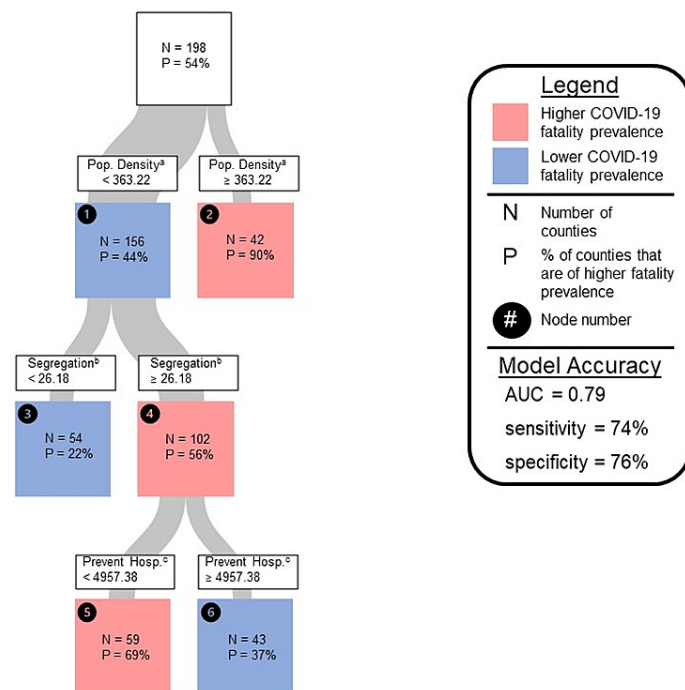
**TABLE 2: Strength of association between social determinants of health and COVID-19 mortality**

PCP - primary care physicians

<sup>a</sup>R-values as reported from Spearman's Correlation test

<sup>b</sup>Data was retrieved on June 15, 2020 from US Census Bureau [19,20]

<sup>c</sup>Data was retrieved on June 15, 2020 from County Health Rankings & Roadmaps [16-18]



**FIGURE 1: Decision tree analysis of social determinants of health and COVID-19 fatalities**

Pop. - population; Hosp. - hospitalizations; N - number of counties; P - % of counties that are of higher fatality prevalence; AUC - area under the receiver operating characteristic curve

<sup>a</sup>Population density in persons per square mile. Data retrieved on June 15, 2020 from US Census Bureau [19,20].

<sup>b</sup>Residential segregation as an index of dissimilarity (between white vs. non-white). Data retrieved on June 15, 2020 from County Health Rankings [16-18].

<sup>c</sup>Preventable hospital stays as hospitalization rate of outpatient-sensitive conditions per 100,000 Medicare enrollees. Data retrieved June on 15, 2020 from County Health Rankings [16-18].

Counties with population density ≥363.220 persons per square mile (p/mile<sup>2</sup>) were classified as high fatality. Of note, among counties with population density <363.220 p/mile<sup>2</sup>, those with residential segregation <26.180 were classified as low fatality. Interestingly, among counties with population density <363.220 p/mile<sup>2</sup> and segregation ≥26.180, those with a lower preventable hospitalization rate per 100,000 Medicare enrollees (<4957.38) were classified as high COVID-19 fatality rate, as seen in terminal node number

five (Figure 1). In a follow-up analysis, counties in node five exhibited significantly higher residential segregation, higher elderly percentage, lower severe housing issue percentage, and lower preventable hospitalization rate compared to other counties analyzed ( $p < 0.05$ ; see Table 3). Counties in node six, with the profile of population density  $< 363,220$  p/mile<sup>2</sup>, residential segregation  $\geq 26.180$ , and preventable hospitalizations  $\geq 4957.380$  hospitalizations per 100,000 Medicare enrollees were predicted to have lower fatality rates (Figure 1).

Measure	Node 5 median	Other counties (excl. Node 5) median	Node 6 median	Node 5 vs. other counties p-value	Node 5 vs. Node 6 p-value
<b>Sociodemographic factors</b>					
Population density <sup>a</sup> (# persons per square mile estimated from 2019 population and 2010 land area)	79.23	94.08	58.96	0.2884	0.0775
Elderly <sup>b</sup> (% adults ages 65 and older)	19.1	18	19	0.0413*	0.7194
Black (% non-Hispanic Black or African American)	2.5	2.7	2.8	0.6341	0.9379
Non-Hispanic White <sup>b</sup> (%)	86.6	79.6	90.4	0.0514	0.0148*
Female <sup>b</sup> (%)	50.3	50.3	50.2	0.7788	0.5082
Rural <sup>b</sup> (%)	43.3	38.9	58	0.1268	0.0339*
High school graduation rate <sup>b</sup> (% ninth graders who graduate in 4 years)	86	87	89	0.8930	0.0820
Some college <sup>b</sup> (% adults ages 25-44 with some post-secondary education)	63	62	60	0.3963	0.0130*
Unemployment rate <sup>b</sup> (% people age $\geq 16$ unemployed but seeking work)	4.5	4.7	4.9	0.4729	0.0091*
Income inequality <sup>b</sup> (ratio of 80th percentile income to 20th percentile income)	4.5	4.5	4.4	0.6285	0.2752
Residential segregation <sup>b</sup> (index of dissimilarity ranging 0-100, used in the American Community Survey where higher values indicate greater residential segregation between non-White and White county residents)	34	31	38	0.0087*	0.0180*
Food insecurity <sup>b</sup> (%)	11	12	12	0.0964	0.0116*
Limited access to healthy foods <sup>b</sup> (% low-income population who do not live close to grocery store)	5	5	5	0.2851	1.0000
<b>Physical environment</b>					
Severe housing problems <sup>b</sup> (% households at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities)	14	15	12	0.0279*	0.0503
Homeowner percentage <sup>b</sup> (% occupied housing units that are owned)	72	71	73	0.5901	0.0703
Severe housing cost burden <sup>b</sup> (% households that spend $\geq 50\%$ household income on housing)	12	13	11	0.0973	0.0310*
Drive alone to work <sup>b</sup> (%)	81	80	83	0.1043	0.0209*
Average traffic <sup>b</sup> (traffic volume per meter)	152	152	99	0.5204	0.0758
<b>Health behaviors</b>					
Physical inactivity <sup>b</sup> (% adults aged $\geq 20$ who report no leisure-time physical activity)	25	25	28	0.7546	0.0005*
<b>Clinical care</b>					
Uninsured <sup>b</sup> (% persons under age 65 without health insurance)	6	6	6	0.5261	0.4584

Primary care physicians <sup>b</sup> (ratio of county population to PCP)	1931	1935	2438	0.8293	0.0342*
Preventable hospital stays <sup>b</sup> (rate of hospital stays for ambulatory-care sensitive conditions per 100,000 Medicare enrollees in a year)	4163	4711	5795	0.0006*	< .0001>
Flu vaccinations <sup>b</sup> (% fee-for-service [FFS] Medicare enrollees that had an annual flu vaccination)	46	45	45	0.1776	0.2218

**TABLE 3: Descriptive statistics and Kruskal-Wallis analyses of Node 5 versus Node 6 and other counties**

excl. - excluding; PCP - primary care physicians

<sup>a</sup>Data was retrieved on June 15, 2020 from US Census Bureau [19,26]

<sup>b</sup>Data was retrieved on June 15, 2020 from County Health Rankings & Roadmaps [16-18]

\*p<0.05

## Discussion

This study sought to use the innovative decision tree model approach to identify relevant SDoH affecting COVID-19 fatalities in the US. While earlier studies have identified SDoH to be influential in COVID-19 related mortality, the key finding of our study is that even in counties of low population density, higher levels of segregation are substantially associated with high county-level COVID-19 related deaths.

Consistent with the Spearman correlations tested, population density and residential segregation remained important in the decision tree modeling. The population density was identified to be the most important county-level predictor of COVID-19 fatalities. This can be explained by denser areas having greater transmission rates, augmenting fatal outcomes. Despite population density being the most important factor, counties of low population density still exhibited relatively high COVID-19 fatality if there was a high degree of residential segregation. Residential segregation was measured in this study by the dissimilarity index, the most widely used measure of evenness comparing spatial distributions of different groups [26]. Measured as an index (0-100), it describes the percentage of white or non-white populations that would have to move to match the population distribution of the metropolitan (larger) area [21]. This study's data-driven decision tree approach determined the segregation index value of 26.180 to be the optimal threshold to be operationalized for predicting high county-level COVID-19 mortality rates. To our knowledge, there is no national standard that would otherwise define the parameters for "high" segregation.

Residential segregation was the second most important county-level predictor, introducing the second split in our decision tree. Although legally banned since 1968, racial residential segregation has intact structures that continue to cause health disparities today [27]. Segregation affects other SDoH, such as socioeconomic (SES) status and poverty, resulting in poorer health outcomes for minorities. Segregation impairs SES status through diminishing access and resources for high-quality education and concentrating higher-pay jobs in areas outside of minority communities [27]. By creating areas of concentrated poverty, segregation has exposed disenfranchised populations' health to harms such as pollution, poor-quality infrastructure, and psychosocial stressors [28]. These harms result in disparities in income, life expectancy, and other SDoH that contribute to poor health outcomes [29]. Therefore, because elements of segregation are pervasive in driving social determinants of health, it is important to study the area of residence when investigating the mechanism of disease onset and progression. Previous studies have shown a positive association between segregation and county-level COVID-19 infection rates [8,10,30]. Our results show that segregation also predicts high COVID-19 fatality rates at the county level. Thus, it is essential to address segregation at the local level in addition to state and national interventions.

The potential relationship between population density and segregation may explain segregation's influence on COVID-19 outcomes in specifically less densely populated counties. In urban areas, population density has an inverse relationship with segregation. Anti-density zoning has restricted private property rights and kept population density low in targeted neighborhoods, increasing residential segregation [31]. Segregation conversely declines with urban population growth, diminishing segregation's influence on health outcomes in populated areas [32]. The observed negative relationship between population density and segregation in urban areas calls forth investigation of possible similar patterns in less dense areas.

Preventable hospital stays are defined as the rate of hospital stays for conditions that can be treated as outpatient per 100,000 Medicare enrollees. It is an indicator of unsatisfactory outpatient care or a pattern of excessively seeking urgent/emergency care [33]. Interestingly, our findings suggest that lower preventable



hospital stays can predict COVID-19 fatalities in less dense but highly segregated counties. This finding is counterintuitive as it would mean that counties with higher COVID-19 fatalities are less likely to have Medicare enrollees excessively using urgent/emergency care for less urgent issues. This finding can potentially be explained by racial disparities, as this group of counties contained a significantly higher proportion of minorities who experience disproportionate barriers to accessing healthcare. Minorities have disproportionately faced higher COVID-19 mortality, which is related to their heightened exposure risk, illness severity, and barriers to testing [34]. Further analysis also showed that this group had a significantly greater housing cost burden, which also is recognized as a barrier to healthcare access [35, 36]. Further investigation is needed into other factors related to preventable hospital stays that can possibly explain this relationship.

## Limitations

The study had the following limitations. First, we used data from three states (New York, Illinois, and California) chosen based on the highest county-level COVID-19 fatalities as of June 15, 2020. This contextualized studying the pertinent factors in the midst of the first COVID wave in three states that had the highest cumulative county-level COVID-19 fatalities up to June 15. However, the analysis did not adjust for COVID-19 testing capacities for each state. Fatality data was also recorded differently per state (confirmed versus presumed deaths) and may not account for Americans who died of COVID-19 before being tested. Although this is a limitation, this applies to other major national datasets, making our data the best available at the time. County attributed data is restricted to the latest published appropriate year. While AUC and model validation results exhibit room for improvement using tree modeling, these results may be influenced by the limited sample size of the present analysis. Our limited sample size also explains our population density inclusion criteria, as we did not have adequate representation of densely populated counties of less than 1.5 persons or greater than 50,000 persons per square mile.

## Conclusions

To our knowledge, this is the first study to employ a novel decision tree method, which utilized machine-learning techniques to study associations between SDoH and COVID-19 fatalities. The findings support the influence of SDoH on health outcomes, including COVID-19 outcomes, and display the success of decision tree analysis used in this context. Our key finding was that pockets of segregation exist among less densely populated counties and are suggested to predispose those residents to disproportionate COVID-19 outcomes. These areas should be targeted for county-level attention/intervention on multiple levels, including resource planning and allocations. Suggested interventions include increasing educational and employment opportunities as well as community financial resources in these counties identified to be at high risk of COVID-19 fatality. Further research should be directed towards examining the current infrastructures that allow segregation to continue and effective interventions.

## Additional Information

### Disclosures

**Human subjects:** Consent was obtained or waived by all participants in this study. Nova Southeastern University's Institutional Review Board issued approval 2020-229. Based on the information provided, your protocol does not require IRB review or approval because its procedures do not fall within the IRB's jurisdiction based on 45 CFR 46.102. Therefore, your protocol has been classified as "Research outside the purview of the IRB" for IRB purposes; your study may still be classified as "research" for academic purposes or for other regulations, such as regulations pertaining to educational records (FERPA) and/or protected health information (HIPAA). This protocol does not involve "human subjects research" for one of the following reasons: (a) The study does not meet the definition of "research", as per federal regulations: "research" means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. (b) The study does not involve "human subjects," per federal regulations. "Human subject" means a living individual about whom an investigator conducting research obtains: (1) Data through intervention or interaction with the individual, or (2) Identifiable private information. (c) Other: Please retain a copy of this memorandum for your records as it indicates that this submission was reviewed by Nova Southeastern University's Institutional Review Board. The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed by Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## References

1. Dorn AV, Cooney RE, Sabin ML: COVID-19 exacerbating inequalities in the US. *Lancet*. 2020, 395:1243-4.



- [10.1016/S0140-6736\(20\)30893-X](https://doi.org/10.1016/S0140-6736(20)30893-X)
2. Maness SB, Merrell L, Thompson EL, Griner SB, Kline N, Wheldon C: Social determinants of health and health disparities: COVID-19 exposures and mortality among African American people in the United States. *Public Health Rep.* 2021, 136:18-22. [10.1177/0033354920969169](https://doi.org/10.1177/0033354920969169)
  3. Ahmed F, Ahmed N, Pissarides C, Stiglitz J: Why inequality could spread COVID-19. *Lancet Public Health.* 2020, 5:E240. [10.1016/S2468-2667\(20\)30085-2](https://doi.org/10.1016/S2468-2667(20)30085-2)
  4. A basic demographic profile of workers in frontline industries . (2020). Accessed: May 1, 2020: <https://cepr.net/wp-content/uploads/2020/04/2020-04-Frontline-Workers.pdf>.
  5. Abrams EM, Szeffler SJ: COVID-19 and the impact of social determinants of health . *Lancet Respir Med.* 2020, 8:659-61. [10.1016/S2213-2600\(20\)30234-4](https://doi.org/10.1016/S2213-2600(20)30234-4)
  6. Dong E, Du H, Gardner L: An interactive web-based dashboard to track COVID-19 in real time . *Lancet Infect Dis.* 2020, 20:533-4. [10.1016/S1473-3099\(20\)50120-1](https://doi.org/10.1016/S1473-3099(20)50120-1)
  7. Fielding-Miller RK, Sundaram ME, Brouwer K: Social determinants of COVID-19 mortality at the county level. *PLoS One.* 2020, 15:e0240151. [10.1371/journal.pone.0240151](https://doi.org/10.1371/journal.pone.0240151)
  8. Yang TC, Emily Choi SW, Sun F: COVID-19 cases in US counties: roles of racial/ethnic density and residential segregation. *Ethn Health.* 2021, 26:11-21. [10.1080/13557858.2020.1830036](https://doi.org/10.1080/13557858.2020.1830036)
  9. Dalsania AK, Fastiggi MJ, Kahlam A, et al.: The relationship between social determinants of health and racial disparities in COVID-19 mortality. *J Racial Ethn Health Disparities.* 2021, [10.1007/s40615-020-00952-y](https://doi.org/10.1007/s40615-020-00952-y)
  10. Paul R, Arif A, Pokhrel K, Ghosh S: The association of social determinants of health with COVID-19 mortality in rural and urban counties. *J Rural Health.* 2021, 37:278-86. [10.1111/jrh.12557](https://doi.org/10.1111/jrh.12557)
  11. Wylezinski LS, Harris CR, Heiser CN, Gray JD, Spurlock CF: Influence of social determinants of health and county vaccination rates on machine learning models to predict COVID-19 case growth in Tennessee. *BMJ Health Care Inform.* 2021, 28:e100439. [10.1136/bmjhci-2021-100439](https://doi.org/10.1136/bmjhci-2021-100439)
  12. COVID-19 statewide case statistics . (2020). Accessed: June 15, 2020: [https://public.tableau.com/views/COVID-19PublicDashboard/Covid-19Public?:embed=y&:display\\_count=no&:showVizHome=no](https://public.tableau.com/views/COVID-19PublicDashboard/Covid-19Public?:embed=y&:display_count=no&:showVizHome=no).
  13. COVID-19 statistics. (2020). Accessed: June 15, 2020: <https://www.dph.illinois.gov/covid19/covid19-statistics>.
  14. NYSDOH COVID-19 tracker. (2020). Accessed: June 15, 2020: <https://covid19tracker.health.ny.gov/>.
  15. Remington PL, Catlin BB, Gennuso KP: The County Health Rankings: rationale and methods . *Popul Health Metr.* 2015, 13:11. [10.1186/s12963-015-0044-2](https://doi.org/10.1186/s12963-015-0044-2)
  16. California rankings data. (2020). Accessed: June 15, 2020: <https://www.countyhealthrankings.org/app/california/2020/downloads>.
  17. Illinois rankings data. (2020). Accessed: June 15, 2020: <https://www.countyhealthrankings.org/app/illinois/2020/downloads>.
  18. New York rankings data. (2020). Accessed: June 15, 2020: <https://www.countyhealthrankings.org/app/new-york/2020/downloads>.
  19. County population totals: 2010-2019. (2020). Accessed: June 15, 2020: <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>
  20. USA counties: 2011. (2011). Accessed: June 15, 2020: <https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html>.
  21. Residential segregation - non-white/white . (2020). Accessed: September 2, 2020: <https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health...>
  22. Banerjee M, Reynolds E, Andersson HB, Nallamotheu BK: Tree-Based Analysis. *Circ Cardiovasc Qual Outcomes.* 2019, 12:e004879. [10.1161/CIRCOUTCOMES.118.004879](https://doi.org/10.1161/CIRCOUTCOMES.118.004879)
  23. SAS Institute Inc.: SAS/STAT® 14.1 User's Guide. The HPSPLIT Procedure . SAS/STAT® 141 User's Guide. SAS Institute Inc., Cary, NC; 2015. 4576-658.
  24. Morris KL, Perna FM: Decision tree model vs traditional measures to identify patterns of sun-protective behaviors and sun sensitivity associated with sunburn. *JAMA Dermatol.* 2018, 154:897-902. [10.1001/jamadermatol.2018.1646](https://doi.org/10.1001/jamadermatol.2018.1646)
  25. A gentle introduction to k-fold cross-validation . (2018). Accessed: January 23, 2021: <https://machinelearningmastery.com/k-fold-cross-validation/>.
  26. Appendix B: measures of residential segregation. (2016). Accessed: November 23, 2020: <https://www.census.gov/topics/housing/housing-patterns/guidance/appendix-b.html>.
  27. Williams DR, Lawrence JA, Davis BA: Racism and health: evidence and needed research. *Annu Rev Public Health.* 2019, 40:105-25. [10.1146/annurev-publhealth-040218-043750](https://doi.org/10.1146/annurev-publhealth-040218-043750)
  28. Williams DR, Collins C: Racial residential segregation: a fundamental cause of racial disparities in health . *Public Health Rep.* 2001, 116:404-16. [10.1093/phr/116.5.404](https://doi.org/10.1093/phr/116.5.404)
  29. The gaps between White and Black America, in charts . (2020). Accessed: June 19, 2020: <https://www.nytimes.com/interactive/2020/06/19/opinion/politics/opportunity-gaps-race-inequality.html>.
  30. Millett GA, Honermann B, Jones A, Lankiewicz E, Sherwood J, Blumenthal S, Sayas A: White counties stand apart: the primacy of residential segregation in COVID-19 and HIV diagnoses. *AIDS Patient Care STDS.* 2020, 34:417-24. [10.1089/apc.2020.0155](https://doi.org/10.1089/apc.2020.0155)
  31. Rothwell J, Massey DS: The effect of density zoning on racial segregation in U.S. urban areas . *Urban Aff Rev Thousand Oaks Calif.* 2009, 44:779-806. [10.1177/1078087409334163](https://doi.org/10.1177/1078087409334163)
  32. Bellman B, Spielman SE, Franklin RS: Local population change and variations in racial integration in the United States, 2000-2010. *Int Reg Sci Rev.* 2018, 41:233-55. [10.1177/0160017616665669](https://doi.org/10.1177/0160017616665669)
  33. Preventable hospital stays. (2020). Accessed: August 27, 2020: <https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health...>
  34. COVID-19 racial disparities in testing, infection, hospitalization, and death: analysis of epic patient data . (2020). Accessed: September 16, 2020: <https://www.kff.org/report-section/covid-19-racial-disparities-in-testing-infection-hospitalization-and-death-analysi...>

35. Kushel MB, Gupta R, Gee L, Haas JS: Housing instability and food insecurity as barriers to health care among low-income Americans. *J Gen Intern Med.* 2006, 21:71-7. [10.1111/j.1525-1497.2005.00278.x](https://doi.org/10.1111/j.1525-1497.2005.00278.x)
36. Dunn JR: Housing and health inequalities: review and prospects for research. *Hous Stud.* 2000, 15:341-366. [10.1080/02673030050009221](https://doi.org/10.1080/02673030050009221)