

# Predicting Health Disparities in Regions at Risk of Severe Illness to inform Healthcare Resource Allocations during Pandemics

Tara Fusillo

*John F. Kennedy High School*

*3000 Bellmore Avenue*

*Bellmore, NY 11710*

[tfushiro@gmail.com](mailto:tfushiro@gmail.com)

**Abstract**— Pandemics including COVID-19 have disproportionately affected socioeconomically vulnerable populations. To create a repeatable modelling process to identify regional population centers with pandemic vulnerability, readily available COVID-19 and socioeconomic variable datasets were compiled, and linear regression models were built during the early days of the COVID-19 pandemic. The models were validated later in the pandemic timeline using actual COVID-19 mortality rates in states with high population densities, with New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan and Pennsylvania showing the strongest predictive results. Our models have been shared with the Department of Health Commissioners of each of these states as input into a much needed “pandemic playbook” for local healthcare agencies in allocating medical testing and treatment resources.

**Keywords**— Coronavirus; SARS-CoV-2; Pandemic; Socioeconomic status; Predictive model; Healthcare resource allocation

## I. INTRODUCTION

Socioeconomic vulnerability can directly influence the severity of pandemics and their impact on mortalities in ways including access to healthcare, residence in crowded housing units and

comorbidities. Prior studies of swine flu (H1N1) have pointed to these factors as contributors to the spread and severity of that pandemic<sup>1</sup>. Other studies have identified national level correlations that are helpful, but not actionable at a local level where actual healthcare resource allocation decisions are made<sup>2</sup>.

Early and accurate decisioning for healthcare resource allocations are particularly critical in geographic locations with high population density. This research created a repeatable modelling process that uses readily available data sources to identify the top counties in densely populated states with high likelihoods of disproportionate COVID-19 mortality rates. These models have been shared with the Department of Health Commissioners of New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan and Pennsylvania where the models performed best. All Commissioners responded positively which validates our hypothesis that neither other researchers nor the Departments of Health themselves have developed a similar modelling process as a means to allocate scarce healthcare resources such as testing, treatment, education and communication.

## II. METHODS

Exploratory data research at a national level was performed using county level data (Federal Information Processing Standards for county

identification; also known as FIPS). COVID-19 mortality datasets (deaths per 100,000 of population) were created using the John’s Hopkins Dataset (<https://coronavirus.jhu.edu/map.html>)<sup>3</sup> and data from the US Census Bureau<sup>4</sup>. Socioeconomic vulnerability datasets at a county level were created using sub-components of the CDC Social Vulnerability Index (<https://svi.cdc.gov/data-and-tools-download.html>)<sup>5</sup>.

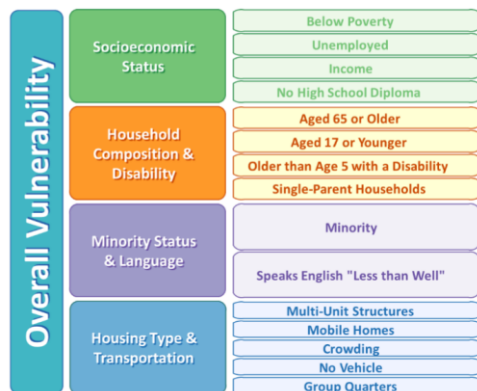


Fig. 1 Full list of sub-component variables from the CDC (<https://svi.cdc.gov/data-and-tools-download.html>)

Scatterplots and trendlines were used to identify those variables most correlated with COVID-19 mortalities. Few, if any, social vulnerability variables correlated across all of the 3,142 FIPS counties, but minority status correlated strongly in certain regions, particularly those with high mortality rates.

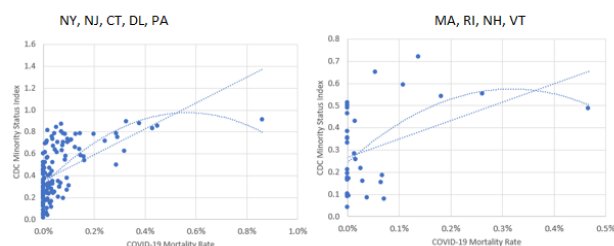


Fig. 2 Examples of regional scatterplots and trendlines

These findings focused the next steps on state level rather than national level correlations. County-level dependent variable datasets were created using COVID-19 mortality and population data from coronadatascraper.com (data service that scrapes county level COVID-19 data on a daily basis)<sup>6</sup> as

well as usafacts.org<sup>7</sup> for cumulative mortalities as of April 8, 2020 and May 8, 2020.

County-level independent variable datasets were created using socioeconomic data from countyhealthrankings.org, a collaboration dataset created by the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute<sup>8</sup>. Given the year-to-year stability of most of these socioeconomic variables, the latest data available in the countyhealthrankings.org was used and no attempt was made to augment the dataset to try to match the time-series to either April 8, 2020 or May 8, 2020.

### 2020 Ranked Measures & Data Sources

Measure	Source	Years of Data	
<b>HEALTH OUTCOMES</b>			
Length of Life	Premature death (50%)	National Center for Health Statistics – Mortality files	2016-2018
Quality of Life	Poor or fair health (10%)	Behavioral Risk Factor Surveillance System	2017
	Poor physical health days (10%)	Behavioral Risk Factor Surveillance System	2017
	Poor mental health days (10%)	Behavioral Risk Factor Surveillance System	2017
	Low birthweight (20%)	National Center for Health Statistics – Natality files	2012-2018
<b>HEALTH FACTORS</b>			
<b>HEALTH BEHAVIORS</b>			
Tobacco Use	Adult smoking (10%)	Behavioral Risk Factor Surveillance System	2017
Diet and Exercise	Adult obesity (5%)	United States Diabetes Surveillance System	2016
	Food environment index (2%)	USDA Food Environment Atlas, Map the Meal Gap	2015 & 2017
	Physical inactivity (2%)	United States Diabetes Surveillance System	2016
	Access to exercise opportunities (1%)	Business Analyst, Delorme map data, ESRI, & US Census Tigerline Files	2010 & 2019
Alcohol and Drug Use	Excessive drinking (2.5%)	Behavioral Risk Factor Surveillance System	2017
	Alcohol-impaired driving deaths (2.5%)	Fatality Analysis Reporting System	2014-2018
Sexual Activity	Sexually transmitted infections (2.5%)	National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention	2017
	Teen births (2.5%)	National Center for Health Statistics – Natality files	2012-2018
<b>CLINICAL CARE</b>			
Access to Care	Uninsured (5%)	Small Area Health Insurance Estimates	2017
	Primary care physicians (3%)	Area Health Resource File/American Medical Association	2017
	Dentists (1%)	Area Health Resource File/National Provider Identification file	2018
	Mental health providers (1%)	CMS, National Provider Identification file	2019
Quality of Care	Preventable hospital stays (5%)	Mapping Medicare Disparities Tool	2017
	Mammography screening (2.5%)	Mapping Medicare Disparities Tool	2017
	Flu vaccinations (2.5%)	Mapping Medicare Disparities Tool	2017
<b>SOCIAL AND ECONOMIC FACTORS</b>			
Education	High school graduation (5%)	EDFacts & State-specific sources	Varies
	Some college (5%)	American Community Survey	2014-2018
Employment	Unemployment (10%)	Bureau of Labor Statistics	2018
Income	Children in poverty (7.5%)	Small Area Income and Poverty Estimates	2018
	Income inequality (2.5%)	American Community Survey	2014-2018
Family and Social Support	Children in single-parent households (2.5%)	American Community Survey	2014-2018
	Social associations (2.5%)	County Business Patterns	2017
Community Safety	Violent crime (2.5%)	Uniform Crime Reporting – FBI	2014 & 2016
	Injury deaths (2.5%)	National Center for Health Statistics – Mortality files	2014-2018
<b>PHYSICAL ENVIRONMENT</b>			
Air and Water Quality	Air pollution – particulate matter <sup>1</sup> (2.5%)	Environmental Public Health Tracking Network	2014
	Drinking water violations (2.5%)	Safe Drinking Water Information System	2018
Housing and Transit	Severe housing problems (2%)	Comprehensive Housing Affordability Strategy (CHAS) data	2012-2016
	Driving alone to work (2%)	American Community Survey	2014-2018
	Lone commute – drive alone (1%)	American Community Survey	2014-2018

Fig. 3 Full variable list from countyhealthrankings.org

Cumulative COVID-19 specific mortality data (deaths per 100,000 of population) by county for high mortality rate states (NY, NJ, CT, MA, LA, MI, PA) as of May 8th was used as the dependent variable.

**Death rates from coronavirus (COVID-19) in the United States**  
(per 100,000 people)

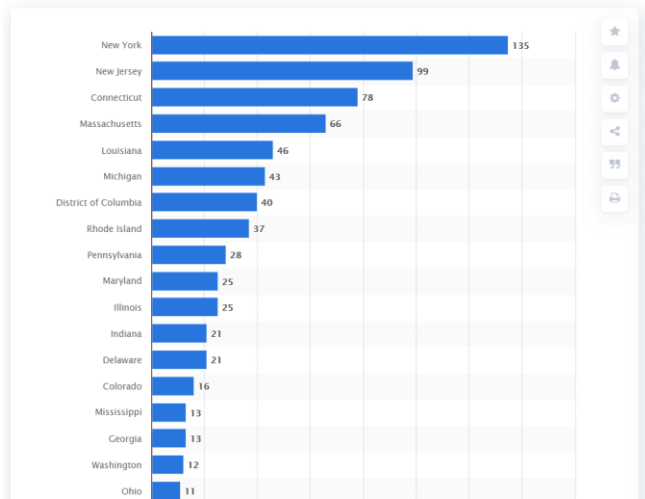


Fig. 4 COVID-19 mortality rates by state as of May 8, 2020 from statista.com<sup>9</sup>

The May 8, 2020 date was used to ensure that the dependent variable would be tuned to a timeframe at or around the peak in daily mortalities when healthcare resources (testing, treatment, tracing) are typically most needed. The mortality curves below provide support for May 8th as the overall date for mortality predictions as shown in the Institute of Health Metrics and Evaluation dataset (<https://covid19.healthdata.org/>)<sup>10</sup>.

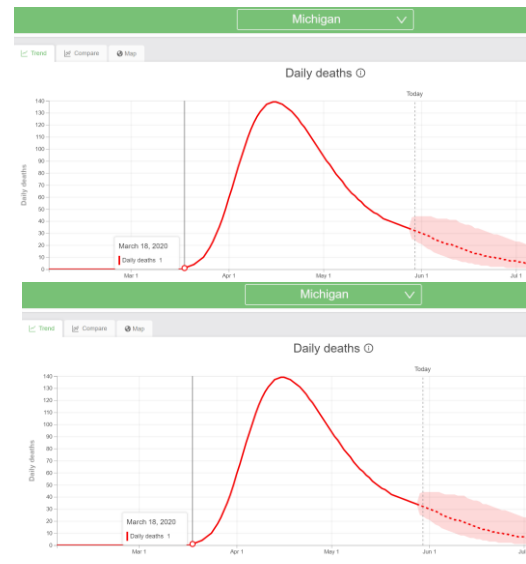
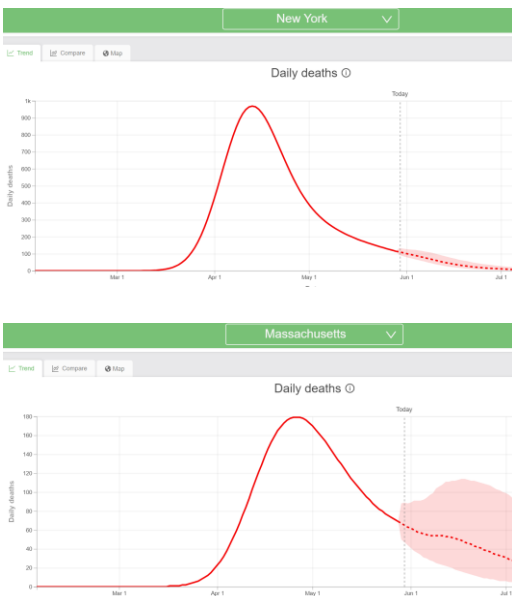


Fig. 5 COVID-19 mortality curves examples from Institute of Health Metrics and Evaluation website (<https://covid19.healthdata.org/>)

A stepwise linear regression technique was used to build each state level model. All relevant independent variables (Figure 3) were initially used in the model (i.e., include severe housing problems, but exclude violent deaths). Next, the variable with the lowest T-statistic was removed and the linear regression was re-run. This process was repeated until all T-statistics for the remaining independent variables were near a value of 2 or greater.

### III. RESULTS

Predictive models were completed for New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan and Pennsylvania with statistically significant results.



Dependent Variable	Top States by COVID-19 Mortality						
	NY	NJ	CT	MA	LA	MI	PA
Deaths per 100,000 Population	X	X	X	X	X	X	X
Cases per 100,000 Population							
Independent Variable (T-Statistic)							
Demographics							
% Black	2.1	1.3	3.1	2.5	5.0	7.7	
% Hispanic	2.1				2.8		
% Asian							
% 65 and over							
% White							
Socioeconomics							
Median Household Income	2.2				1.6	3.8	
% Severe Housing Problems (multi-family)	4.0	2.9					4.8
% Uninsured							
Average Daily PM2.5 (Pollution)				4.8			
% Unemployed							
Health							
% Adults with Obesity							
Primary Care Access (Z-Score)					1.9		
Life Expectancy (comorbidity)	1.6	5.1	1.9		1.9		
% Limited Access to Healthy Foods							
Model Strength							
F-Statistic	0.000%	0.008%	2.380%	0.093%	0.008%	0.000%	0.000%
R-Squared	89%	71%	78%	72%	33%	44%	48%

Fig. 6 Final model variables and statistical significance

Model validation comparing predicted to actual county rankings by cumulative mortality rate (deaths per 100,000 of population) through May 8<sup>th</sup> and data visualizations using state level maps were completed. These validations were used to share model methods and results with the Commissioners of each state's Department of Health.

New York Counties	Deaths per 100,000 (ACTUALS)	Deaths per 100,000 (MODEL)
Bronx	198	197
Kings County, NY	167	155
Queens County, NY	188	148
Westchester	115	118
New York	104	118
Rockland	129	114
Nassau	134	102
Suffolk	87	91
Richmond County, NY	128	88
Orange	68	83

New Jersey Counties	Deaths per 100,000 (ACTUALS)	Deaths per 100,000 (MODEL)
Hudson County	139.8	138.6
Passaic County	142.5	125.9
Bergen County	142.6	121.4
Union County	151.7	116.2
Essex County	175.0	100.9
Middlesex County	90.7	93.4
Hunterdon County	35.4	92.8
Somerset County	99.7	88.4
Morris County	102.9	87.7

Fig. 7 Top counties by mortality rate for New York and New Jersey (actuals vs. model prediction). Red indicates high mortality rates, yellow indicates medium and green indicates low relative to all other counties within a particular state

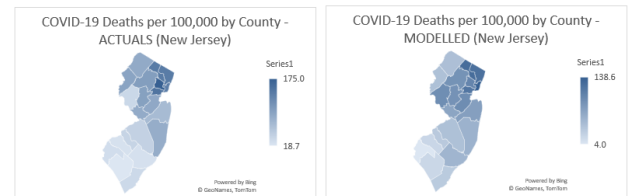
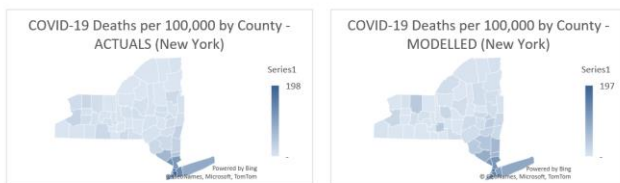


Fig. 8 COVID-19 mortality rate heat maps for New York and New Jersey (actuals vs. model prediction)<sup>11</sup>

Two further validations were completed as well. The first validation was to check model performance using COVID-19 mortality data on April 8, 2020 instead of May 8, 2020. This validation tested whether models using data available early in the pandemic would have been sufficient to build the models. The same variables were used, but coefficients were recalibrated with the April 8 dataset. April 8 and May 8 model outputs were compared to test for stability in the top counties predicted for high COVID-19 mortality rates. For New York, New Jersey and Connecticut, the models were stable. For Massachusetts, the 4/8 model performance was not strong, but this was easily corrected by using case data in place of mortality data. This is an important finding as it validates the predictive power contained in early case data which is more readily available at the start of a pandemic.

New York Counties	Deaths per 100,000 (5/8)	Deaths per 100,000 (MODEL)
Bronx	198	197
Kings County, NY	167	155
Queens County, NY	188	148
Westchester	115	118
New York	104	118
Rockland	129	114
Nassau	134	102
Suffolk	87	91
Richmond County, NY	128	88
Orange	68	83

New York Counties	Deaths per 100,000 (5/8)	Deaths per 100,000 (4/8 MODEL)
Bronx	198	68
Queens County, NY	188	47
Kings County, NY	167	46
New York	104	38
Westchester	115	37
Rockland	129	34
Nassau	134	28
Richmond County, NY	128	26
Suffolk	87	26
Orange	68	25

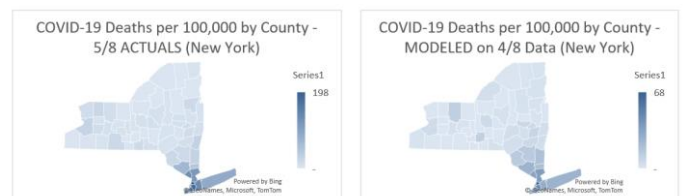


Fig. 9 New York model validation results example using April 8 mortality data

Massachusetts County	Deaths per 100,000 (ACTUALS)	Deaths per 100,000 (MODEL)	Massachusetts County	Deaths per 100,000 (ACTUALS - 5/8)	Deaths per 100,000 (MODEL - 4/8)
Suffolk County	75.8	83.3	Middlesex County	63.9	10.4
Middlesex County	63.9	73.4	Hampshire County	29.2	10.1
Hampden County	88.6	70.9	Franklin County	52.4	10.0
Norfolk County	81.8	66.6	Essex County	66.9	9.6
Worcester County	39.9	55.8	Berkshire County	28.6	9.4
Hampshire County	29.2	54.4	Worcester County	39.9	9.4

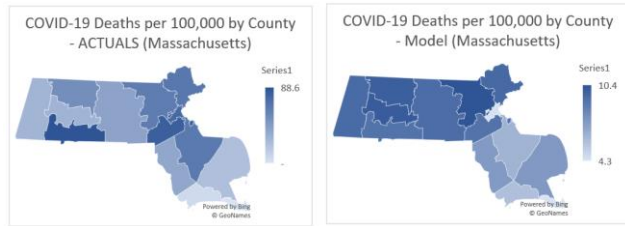


Fig. 10 Massachusetts model validation results example using April 8 mortality data

Massachusetts County	Deaths per 100,000 (ACTUALS)	Deaths per 100,000 (MODEL)	Massachusetts County	Deaths per 100,000 (ACTUALS - 5/8)	Cases per 100,000 (MODEL - 4/8)
Suffolk County	75.8	83.3	Suffolk County	75.8	406.9
Middlesex County	63.9	73.4	Hampden County	88.6	267.7
Hampden County	88.6	70.9	Norfolk County	81.8	252.1
Norfolk County	81.8	66.6	Middlesex County	63.9	251.7
Worcester County	39.9	55.8	Plymouth County	65.8	245.2
Hampshire County	29.2	54.4	Worcester County	39.9	206.5

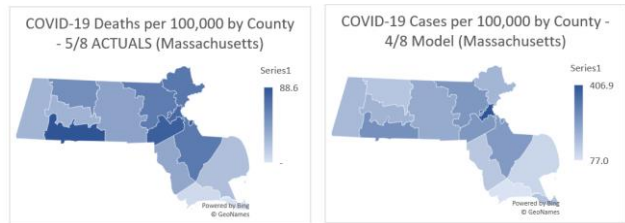


Fig. 11 Massachusetts model validation results example using April 8 case data

The second validation was to check the independent variables for multi-collinearity, with a specific focus on the correlations between Black and variables such as severe housing and uninsured. Strong multi-collinearity was seen in Connecticut, Massachusetts, Louisiana and Michigan partially explaining why these variables did not remain in the model after the stepwise regression process.

	Top States by COVID-19 Mortality			
	CT	MA	LA	MI
Correlation	Black	Black	Black	Black
Severe Housing	75%	56%	58%	45%
Uninsured	51%			
Primary Care		33%		
Unemployed			44%	
Obesity				

Fig. 12 Multi-collinearity results for states without severe housing and other socioeconomically vulnerable variables

With models and validations completed, the Departments of Health for New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan and Pennsylvania were contacted. Additionally, agents of the New York Department of Health (Northwell Health and CORE) and a third-party statistical modelling firm for Connecticut (COVIDACTNOW) were contacted. Response from these contacts were positive and, in some cases, occurred within an hour of outreach (Northwell Health). This response indicates the strong appetite and need for this type of healthcare resource allocation tool for pandemics and other health crises. In fact, the Pennsylvania Department of Health indicated this tool's importance in any second waves of COVID-19.

Department of Health or Partner	Models Shared	Receipt Acknowledged	Zoom/Webex Session
New York DOH			
New York Northwell Health			
New York CORE			
Jew Jersey DOH			
Connecticut DOH			
Connecticut COVIDACTNOW			
Michigan DOH			
Louisiana DOH			
Pennsylvania DOH			

Fig. 14 State Department of Health contact summary

#### IV. DISCUSSION

The research described in this manuscript has shown the extent to which Black Americans, people living in crowded housing units, and households with less access to healthcare are at higher risk of severe illness during a pandemic. The results of other studies provide several possible explanations for these results. Lower income households are more likely to live in crowded housing units and multi-family homes<sup>12</sup>. Lower income households are also structurally disadvantaged in their access to medical insurance and healthcare<sup>13</sup>. Some studies have also pointed to a concept called “weathering” within the Black American community. Arline Geronimus, a professor of public health at the University of Michigan, showed through her research that among black communities, coping with financial strain, discrimination and barriers to good education elevates the stress response

contributing to obesity, diabetes, hypertension and heart disease<sup>14</sup>.

The CDC conducted research on the co-occurrence of COVID-19 and certain ethnicities using a sample of 580 patients with lab-confirmed data<sup>15</sup>. Their results showed that more hospitalizations in Black patients than White patients. The researchers cited underlying medical conditions, work circumstance and living conditions to be major factors in COVID-19 mortalities in their sample. For example, members of racial and ethnic minorities are more likely to live in densely populated areas, making it more difficult to practice social distancing and being more susceptible to contracting and spreading COVID-19<sup>15</sup>. These members also lived farther away from grocery stores and medical facilities, thus being less able to receive necessary resources and medical attention<sup>15</sup>. Other examples cited included Hispanic and Black American workers employed in higher risk industries and often lacking paid sick leave. The researchers hypothesized that these types of workers were more likely to continue working despite being sick, thus exposing other workers to the disease<sup>15</sup>. The CDC recommended at the conclusion of this research that public health officials communicate to different population groups about COVID-19 and provide more healthcare services to ethnic minority groups<sup>15</sup>.

While studies such as these are insightful, we are not aware of any research that translates these impacts into predictive models that can be used locally to direct healthcare resources to the communities most likely to need them, thereby reducing mortalities caused by an ongoing set of institutional inequities.

That said, some organizations have attempted to create healthcare resource allocation methods using descriptive statistics. The Centers for Disease Control and Prevention (CDC) created the social vulnerability index (SVI), allowing the healthcare community to see which factors contribute to socioeconomic vulnerability. The CDC SVI factors are grouped into 4 groups: Socioeconomic status, Household Composition &

Disability, Minority Status & Language, and Housing Type & Transportation. Although these factors are crucial inputs in identifying specific vulnerable communities, these four categories of factors alone are not sufficient to create the types of predictive models that state and local healthcare agencies can use. One example of this insufficiency is a recent study at Emory University where researchers identified correlations between COVID-19 mortalities and the SVI at a national level in the US<sup>2</sup>. While this study is helpful, it stopped short of recommending methods or processes to effectively distribute healthcare resources to specific counties in the US, particularly in the early days of a pandemic. Another study from the Surgo Foundation states that COVID-19 created new challenges for many communities with a dependence on health and structural factors, which are not completely captured by the CDC SVI<sup>16</sup>. In addition to the four socioeconomic factors provided by the CDC, the Surgo Foundation added two more factors: Epidemiologic Factors and Healthcare System Factors. Underlying health conditions in addition to healthcare system factors have been proven to greatly enhance a community's vulnerability during a pandemic. The Surgo Foundation combined these two factors with the CDC SVI to create the COVID-19 Community Vulnerability Index (CCVI). The Surgo Foundation has created heatmaps to show retrospectively which counties were most vulnerable as measured by their CCVI. Similar to the Emory University study however, this methodology did not create a predictive model to identify where the mortalities would be highest at peak periods in a pandemic. We also compared the Surgo Foundation heatmap to our own predictive model rankings and confirmed that our projections of the top counties by per capita mortalities were far closer to reality.

Pennsylvania County	Deaths per 100,000 (ACTUAL) - May 8th	Deaths per 100,000 (MODEL)	CCVI (Surgo)
<a href="#">Philadelphia</a>	4.0	4.4	0.80
<a href="#">Chester</a>	2.9	4.3	0.15
<a href="#">Monroe</a>	3.5	4.0	0.68
<a href="#">Bucks</a>	4.5	4.0	0.14
<a href="#">Montgomery</a>	5.3	3.7	0.18
<a href="#">Delaware</a>	5.6	3.5	0.36
<a href="#">Lehigh</a>	2.6	3.2	0.27

New York Counties	Deaths per 100,000 (5/8)	Deaths per 100,000 (5/8 MODEL)	CCVI (Surgo)
Bronx	198	197	0.99
Kings County, NY	167	155	0.91
Queens County, NY	188	148	0.83
Westchester	115	118	0.52
New York	104	118	0.32
Rockland	129	114	0.56
Nassau	134	102	0.17
Suffolk	87	91	0.28
Richmond County, NY	128	88	0.58
Orange	68	83	0.60

County	Deaths per 100,000 (ACTUALS 5/8)	Deaths per 100,000 (MODEL 5/8)	CCVI (Surgo)
Hudson County	139.8	138.6	0.64
Passaic County	142.5	125.9	0.40
Bergen County	142.6	121.4	0.08
Union County	151.7	116.2	0.45
Essex County	175.0	100.9	0.52
Middlesex County	90.7	93.4	0.38
Hartford County	35.4	92.8	0.06
Somerset County	99.7	88.4	0.08
Morris County	102.9	87.7	0.10

Massachusetts County	Deaths per 100,000 (ACTUALS)	Deaths per 100,000 (MODEL)	CCVI (Surgo)
Suffolk County	75.8	83.3	0.31
Middlesex County	63.9	73.4	0.30
Hampden County	88.6	70.9	0.38
Norfolk County	81.8	66.6	0.16
Worcester County	39.9	55.8	0.33
Hampshire County	29.2	54.4	0.10

Fig. 14 Comparison to Surgo Foundation Heatmap CCVI rankings

During the early phases of the research described in this manuscript, we used the CDC's SVI, similar to the Emory University study. We explored *all* of the sub-category factors in the SVI, but none showed strong correlations at a FIPS county level across the US. We then grouped states together by region and found strong correlations in the most densely populated regions, particularly with the minority status sub-factors. Similar to the Surgo Foundation study, we posited that the CDC's SVI subfactors alone would be insufficient to build predictive models, so a far more complete independent variable dataset of socioeconomic and healthcare data was sourced from countyhealthrankings.org as discussed earlier. The key differentiation of our work are the predictive models *for each state* given the local differences in how each of the factors act as predictors of peak COVID-19 per capita mortalities. Our experience in meeting with the Pennsylvania Department of Health in early June 2020 confirms the uniqueness and usefulness of the approach given that they will be using the predictive modelling process for healthcare resource allocations in the event of a potential second wave of COVID-19 in the Fall of 2020. Other state level departments of health were similarly intrigued by our predictive approach including those in New York, New Jersey and Connecticut.

### Limitations

The models employed in the analyses are reliant on the accuracy of the datasets compiled. COVID-19 mortality data in particular has been notoriously difficult for states to report accurately at a county level throughout the pandemic for reasons including mortality cause classification errors at local coroner's offices<sup>17</sup>.

## V. CONCLUSIONS

Our modelling process can be used to identify early in future pandemics or health crises, the communities most in need of healthcare resources. The COVID-19 models and the overall methodology have been received with enthusiasm by the New York, New Jersey, Connecticut, Massachusetts, Louisiana, Michigan and Pennsylvania Departments of Health.

### ACKNOWLEDGMENT

The author wishes to acknowledge Dr. Fathima Wakeel, Associate Professor at Lehigh University College of Health, for her mentorship throughout this project and for her review of this manuscript. Dr. Wakeel's guidance was instrumental, particularly in directing the research toward a state level and in driving to actionable recommendations for state departments of health. The author also wishes to acknowledge Iwao Fusillo, Global Head of Data & Analytics for the National Football League (NFL) and Ms. Fusillo's father, for helping her learn various techniques in dataset creation, statistical analysis and data visualization.

### DISCLOSURES

The author has no conflicts of interest to disclose.

## REFERENCES

- [1] Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *American Journal of Public Health*. 2011;101(2):285-293.
- [2] Social Vulnerability and the COVID-19 Pandemic in U.S. ; Aditi Nayak, MD; Department of Medicine, Division of Cardiology, Emory University School of Medicine, Atlanta, GA, *medRxiv*
- [3] “COVID-19 Map.” *Johns Hopkins Coronavirus Resource Center*, [coronavirus.jhu.edu/map.html](https://coronavirus.jhu.edu/map.html).
- [4] US Census Bureau, <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html>
- [5] “CDC’s Social Vulnerability Index (SVI).” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, [svi.cdc.gov/data-and-tools-download.html](https://svi.cdc.gov/data-and-tools-download.html).
- [6] Davis, Larry. *Corona Data Scraper*, [coronadatascraper.com/](https://coronadatascraper.com/).
- [7] “Coronavirus Live Map: US Coronavirus Cases by County.” *USA Facts*, USA Facts, 23 June 2020, [usafacts.org/visualizations/coronavirus-covid-19-spread-map/](https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/).
- [8] County Health Rankings, <https://www.countyhealthrankings.org/>
- [9] Elflein, John. “Trailing Seven Day Average Number of COVID 19 Deaths Select Countries Worldwide 2020.” *Statista*, 24 June 2020, [www.statista.com/statistics/1111867/trailing-seven-day-average-number-of-covid-19-deaths-select-countries-worldwide/](https://www.statista.com/statistics/1111867/trailing-seven-day-average-number-of-covid-19-deaths-select-countries-worldwide/).
- [10] “IHME: COVID-19 Projections.” *Institute for Health Metrics and Evaluation*, [covid19.healthdata.org/united-states-of-america](https://covid19.healthdata.org/united-states-of-america).
- [11] Bing & TomTom data from the Excel mapping tools
- [12] Estimating the Gap in Affordable and Available Rental Units for Families APRIL 2019 | WHITNEY AIRGOOD-OBRYCKI & JENNIFER MOLINSKY  
[https://www.jchs.harvard.edu/sites/default/files/harvard\\_jchs\\_family\\_sized\\_rental\\_housing\\_2019.pdf](https://www.jchs.harvard.edu/sites/default/files/harvard_jchs_family_sized_rental_housing_2019.pdf)
- [13] Health Care Inequality in America, KIMBERLY AMADEO, May 11, 2020
- [14] Adamy, Janet. “Coronavirus, Economic Toll Threaten to Worsen Black Mortality Rates.” *The Wall Street Journal*, Dow Jones & Company, 13 June 2020, [www.wsj.com/articles/coronavirus-economic-toll-threaten-to-worsen-black-mortality-rates-11592046000?emailToken=865c35fefb52e208b8fb39a2daa49e71UJO5dRbOdXLCnrfSk7YDNR2mmkOfFnCWn38RjAX7hD+wJGoYU110bKesAk7gIpjqmxQi2ZTaWWo0fgXpOQ+k8TaXNAMN4%2FE%2FBhLJB5YOn4KAuApr1yuL59UPETY3LbB](https://www.wsj.com/articles/coronavirus-economic-toll-threaten-to-worsen-black-mortality-rates-11592046000?emailToken=865c35fefb52e208b8fb39a2daa49e71UJO5dRbOdXLCnrfSk7YDNR2mmkOfFnCWn38RjAX7hD+wJGoYU110bKesAk7gIpjqmxQi2ZTaWWo0fgXpOQ+k8TaXNAMN4%2FE%2FBhLJB5YOn4KAuApr1yuL59UPETY3LbB).
- [15] “COVID-19 in Racial and Ethnic Minority Groups.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 25 June 2020, [www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html).
- [16] *Bringing Greater Precision to the COVID-19 Response*, [precisionforcovid.org/ccvi](https://precisionforcovid.org/ccvi).
- [17] ‘Probable’ COVID-19 Death Reporting Varies by State, Amanda D’Ambrosio, April 23, 2020, *MedPage Today*

### Other datasets referenced in this research.

- Tableau Dashboard on Vulnerability: <https://c19hcc.org/resource/vulnerable-population>
- World Health Organization (WHO) dashboard: <https://who.sprinklr.com/>
- Full list of datasets: <https://cgdv.github.io/challenges/COVID-19/datasource/>
- Uninsured data: [https://www.census.gov/data-tools/demo/sahie/#/data\\_2013](https://www.census.gov/data-tools/demo/sahie/#/data_2013)
- COVID-19 deaths by ethnicity: <https://www.apmresearchlab.org/covid/deaths-by-race>
- FIPS to Zip code mapping: <https://data.world/niccolley/us-zipcode-to->



[county-state/workspace/file?filename=ZIP-COUNTY-FIPS\\_2018-03.csv](#)

- Nursing Home (Medicare) by Zip code: <https://data.medicare.gov/data/nursing-home-compare>
- Johns Hopkins tracking of states that currently provide COVID data by ethnicity ( <https://coronavirus.jhu.edu/data/racial-data-transparency> )

Other research papers evaluated during the course of this research.

- Research paper based on China data (complicated paper, but some good ideas): <https://doi.org/10.1101/2020.03.13.20035238doi>
- Research paper on socio-economic status and cardio-vascular disease (not directly applicable to COVID-19, but interesting): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1694190/>
- FastCompany – excellent article with many ideas for hypotheses to test: <https://www.fastcompany.com/90479231/9-maps-that-show-which-areas-could-be-more-vulnerable-to-the-covid-19-pandemic>
- Fivethirtyeight article: <https://fivethirtyeight.com/features/the-young-americans-most-vulnerable-to-covid-19-are-people-of-color-and-the-working-class/>
- The Conversation article: <https://theconversation.com/covid-19-is-hitting-black-and-poor-communities-the-hardest-underscoring-fault-lines-in-access-and-care-for-those-on-margins-135615>
- HealthAffairs article: <https://www.healthaffairs.org/doi/10.1377/hllog20200319.757883/full/>
- Vox article: <https://www.vox.com/identities/2020/4/7/21211849/coronavirus-black-americans>
- FastCompany article on interesting data tool for Urban Footprint: <https://www.fastcompany.com/90481120/this-tool-is-helping-cities-find-the-neighborhoods-most-vulnerable-to-coronavirus>
- Persistent geographic variations in availability and quality of nursing home care in the United States: 1996 to 2016 | BMC Geriatrics: <https://bmcgeriatr.biomedcentral.com/article/s/10.1186/s12877-019-1117-z>
- Do Lockdowns Save Many Lives? In Most Places, the Data Say No: [https://www.wsj.com/articles/do-lockdowns-save-many-lives-is-most-places-the-data-say-no-11587930911?emailToken=f098d16b875951b5c8f988a12e639f1eOHpwB2N8i9ta8xko7k29ejymFwfBHRJxh+pWaTu1FkXQvn85QSXeuxBX0JI/nDe1DCUBUTnL12Lah6/ggNGXoFIA205HsvxNzx0X69Dkif50N1DISdruzjxIT/8BzYe&reflink=article\\_email\\_share](https://www.wsj.com/articles/do-lockdowns-save-many-lives-is-most-places-the-data-say-no-11587930911?emailToken=f098d16b875951b5c8f988a12e639f1eOHpwB2N8i9ta8xko7k29ejymFwfBHRJxh+pWaTu1FkXQvn85QSXeuxBX0JI/nDe1DCUBUTnL12Lah6/ggNGXoFIA205HsvxNzx0X69Dkif50N1DISdruzjxIT/8BzYe&reflink=article_email_share)
- NYC’s Deaths Mirror Patterns Elsewhere: [https://www.wsj.com/articles/new-york-citys-coronavirus-deaths-match-demographics-in-other-hot-spots-11587214800?emailToken=431a69ce96c62d7963b263078ba230737i1M5JE3rrp14zRmD2CWM3JsOVaJyWmS5o0CyI69f0ZsVl8dolZ1yLPaoZtPrXy1OTsEJJYygy4rPaqPY3w2uK45f2aAFtPA007//tC6A3ZlzBkmToXjwuLlllX2LEAjTb2Pf9cH54YINUPzTwSG2Q%3D%3D&reflink=article\\_email\\_share](https://www.wsj.com/articles/new-york-citys-coronavirus-deaths-match-demographics-in-other-hot-spots-11587214800?emailToken=431a69ce96c62d7963b263078ba230737i1M5JE3rrp14zRmD2CWM3JsOVaJyWmS5o0CyI69f0ZsVl8dolZ1yLPaoZtPrXy1OTsEJJYygy4rPaqPY3w2uK45f2aAFtPA007//tC6A3ZlzBkmToXjwuLlllX2LEAjTb2Pf9cH54YINUPzTwSG2Q%3D%3D&reflink=article_email_share)
- New COVID-19 Community Vulnerability Map Uses Social Determinants of Health to Identify Populations At Greater Risk: <https://hitconsultant.net/2020/03/24/covid-19-community-vulnerability-map/#.XqS3EzNKhPY>