Title: Assessing capacity to social distance and neighborhood-level health disparities during the COVID-19 pandemic

Authors: Daniel Carrión[1], Elena Colicino[1], Nicolo Foppa Pedretti[1], Johnathan Rush[1], Nicholas DeFelice[1], Allan C. Just[1,2]

Affiliations:
[1] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY
[2] Institute for Exposomic Research, Icahn School of Medicine at Mount Sinai, New York, NY

Abstract:
The COVID-19 pandemic has yielded disproportionate impacts on communities of color in New York City (NYC). Researchers have noted that social disadvantage may result in limited capacity to socially distance, and consequent disparities. Here, we investigate the role of neighborhood social disadvantage on the ability to socially distance, infections, and mortality. We combine Census Bureau and NYC open data with SARS-CoV-2 testing data using supervised dimensionality-reduction with Bayesian Weighted Quantile Sums regression. The result is a ZIP code-level index with relative weights for social factors facilitating infection risk. We find a positive association between neighborhood social disadvantage and infections, adjusting for the number of tests administered. Neighborhood infection risk is also associated with capacity to socially isolate, as measured by NYC subway data. Finally, infection risk is associated with COVID-19-related mortality. These analyses support that differences in capacity to socially isolate is a credible pathway between disadvantage and COVID-19 disparities.

Introduction

The 2019 novel coronavirus (SARS-CoV-2) emerged in Wuhan, China, and has since become a worldwide pandemic. In the United States, given the nature of this novel infectious disease, it was believed that anyone exposed to the pathogen was susceptible to infection, there were no proven pharmacologic treatments, and testing capacity was low. Pre-existing conditions are known risk factors of disease severity, and mortality increases sharply with age (Onder et al., 2020). Consequently, the United States federal, state, and local governments have principally relied on non-pharmaceutical interventions such as social distancing and mask-wearing. It is possible that non-pharmaceutical interventions are not equally observable across the population. We examine the role of social factors, such as employment, occupation, and poverty, in infection risk by way of mobility and transportation data.

It has been widely noted in popular media and emerging scientific evidence that COVID-19 is not being experienced equally across the population (Chowkwanyun & Reed, 2020; Dorn et al., 2020; Gold, 2020; Webb Hooper et al., 2020; Yancy, 2020). For example, in Chicago, Blacks comprise 70% of COVID-related deaths, but only 30% of the population. In New York City (NYC), Hispanics/Latinx and Blacks are disproportionately impacted, representing 34% and 29% of the deaths, but 28% and 22% of the population, respectively (Yancy, 2020). While differences in disease severity are likely attributed to higher levels of preexisting conditions, i.e. health disparities, this does not explain differences in disease incidence. A survey of laboratory-confirmed hospitalized cases across 14 states found, where race was reported, that 33.1% of hospitalized patients were non-Hispanic Black (Garg, 2020). In New York City (NYC), as of May 13, 2020, the cumulative rate of non-hospitalized positive cases were 798.2, 684.8, and 616.0 per 100,000 for Blacks/African Americans, Hispanic/Latinx, and Whites respectively (NYC DOHMH, 2020).

A body of literature on the social determinants of health suggest that there are numerous inequities that provide the scaffolding for increased COVID-19 infection rates in communities of color. Racism operates on both the individual and structural levels, the latter explaining the societal mechanisms that reinforce inequality, including through housing, employment, earnings, benefits, health care, criminal justice, etc. (Krieger, 2014). Those structural forms of social disadvantage are implicated in many of the health disparities we observe in communities of color (Bailey et al., 2017). Some manifestations relevant to the current coronavirus pandemic might include that people of color (POC) are more represented amongst low-wage jobs (Acs & Loprest, 2009), many of which are now deemed essential (Berchick et al., 2019). When they get home from work, they are more likely to return to densely populated homes and neighborhoods (Murray et al., 2006). Further, multigenerational housing is more common in communities of color (Guzman, 2019), making social distancing between least susceptible (healthy children) and most susceptible (elderly adults with chronic conditions) difficult. POC often live further from supermarkets and sources of nutritious foods, necessitating further travel for groceries (R. E. Walker et al., 2010). These factors, among others, may explain some of the many ways that social distancing is more difficult for communities of color (Yancy, 2020), offering some insight into the social mechanisms that facilitate viral spread, and consequently the preconditions of racial disparities in infection risk (Acevedo-Garcia, 2000).

In this study, we use socioeconomic data on neighborhood characteristics to understand differences in infection rates between neighborhoods. We examine the relative contribution of these measures of social disadvantage and if a proxy of social isolation, NYC subway utilization, helps us to understand these differences. We create a ZIP code level infection risk index for NYC and show how this index explain racial/ethnic disparities in cases, thus reflecting structural forms of disadvantage. Finally, we examine the relationship between neighborhood infection risk and neighborhood-level COVID-19 mortality. Ultimately, we create a tool that identifies social

factors that facilitate viral spread, and therefore, may be useful throughout the US to pinpoint potential areas for targeted public health intervention.

Methods

Data sources and cleaning

SARS-CoV-2 testing and COVID-19 mortality data. The New York City Department of Health and Mental Hygiene (NYC DOHMH) has been releasing daily testing data (positive and total tests) at the ZIP Code Tabulation Area (ZCTA) level since April 1, 2020, and more recently COVID-19 related mortality data since mid-May, both available on GitHub (NYC DOHMH, 2020). We leveraged these data to estimate SARS-CoV-2 infection and related mortality risk in the population. Our analyses relied on pre-pandemic demographic data to describe variation in neighborhood-level disease burden after much of the community had potential for exposure. Since spatiotemporal patterns in infection risk were highly variable at the beginning of the pandemic in relation to many independent viral introductions within NYC (Gonzalez-Reiche et al., 2020), we chose to model cumulative infections on May 7, 2020, four weeks after NYC's peak infection period. We rounded up the mean expected time from symptom onset to death to 16 days from the Centers for Disease Control and Prevention Preliminary COVID-19 estimates (CDC, 2020). Therefore, we chose May 23, 2020 for our cumulative COVID-19 mortality analysis. Although our study team has a research protocol for the study of COVID-19 healthcare demand approved by the IRB of the Mount Sinai Health System, this analysis only used public datasets without identifiable information and thus is not human subjects research.

Census data. We downloaded the Census Bureau's 2018 American Community Survey (ACS) data from the *tidycensus* package in R (K. Walker, 2020). Data were collected for the 177 ZCTAs in New York City. Variables included: the total population, number of households, median income, median rent, health insurance status, unemployment, individuals at or below 150% of the federal poverty level, race and ethnicity, industry of employment, and mode of

transportation to work. A full list of variables are provided in ***Supplemental Table 1***. Industry of employment was used to estimate a proxy for the proportion of the population in occupations likely to be essential workers. This was a sum of those who reported employment in the agricultural, construction, wholesale trade, transportation and utilities, and education/healthcare industries, divided by the total working-age population. To account for teachers being mostly home, and healthcare workers being essential, we included only half of the education/healthcare industry respondents. From these data we also derived an average household size measure by dividing the total population by the number of households.

Residential buildings and food access data. We used two sources of information on residential buildings in our analyses. First we downloaded a New York City building footprints dataset (NYC Open Data, 2020) and merged it with the Primary Land Use Tax Lot Output (PLUTO) dataset (NYC Department of Planning, 2020). These two datasets allowed us to calculate the volume (cubic footage) of residential space in a ZCTA and, when divided by the total population, a residential population density metric. Food access was calculated using data from New York State's Open Data portal for Retail Food Stores (NYS Food Safety, 2019). We restricted to businesses including J, A, and C establishment code designations in order to identify those most likely to provide fresh foods and produce, and then manually removed any business names that indicated being a corner store or pharmacy, or primarily selling alcohol/tobacco. We spatially joined the point locations to our ZCTA shapefile and used Census data to calculate a 'grocers per 1,000 people' variable as a proxy for food access.

Mobility and transit data. The Metropolitan Transit Authority (MTA) of New York City releases subway utilization data on a weekly basis (NYC MTA, 2020). These data include the number of entrances and exits per station, in each four-hour time span. To account for expected usage of the subway accounting for month and day of the week, we divided the total turnstile count for each day in 2020 with the median of the same day of the week within the same month in 2015-2019 to construct a metric of relative subway use for each day for a given station or

area. For a given area, the relative subway usage was calculated for the sum of all entrances and exits for all stations falling within that geographic area.

Quantitative Analyses

Cross-sectional Neighborhood Infection Risk Index. We develop a weighted combination of how a set of socioeconomic variables relates to the population burden of infection. Socioeconomic variables are known to be closely correlated with one another, which is a challenge to model fitting and interpretation of the underlying latent relationship. To address these challenges, we leveraged Bayesian Weighted Quantile Sums Regression (BWQS) (Colicino et al., 2020). In short, BWQS incorporates the quantiles of correlated independent variables into a composite weighted index, with weights showing the relative contribution of each independent variable to the index, and models the overall association (β_1) of this weighted index against the outcome. All results are computed with a Hamiltonian-Monte Carlo algorithm. We incorporated a link function for a negative binomial distribution for the dependent variable: the cumulative number of positive SARS-CoV-2 tests per 100,000 as of May 7, 2020. We included a large candidate list of socioeconomic variables in the BWQS that could represent some of the underlying infection dynamics attributable to socioeconomic disadvantage. All socioeconomic variables were ranked in deciles to limit the effect of outliers, to ease interpretation and comparability of ZCTA-level infection risk index, which is out of ten. They included selected demographic variables collected from the 2018 ACS, as well as derived variables such as population density (persons per square foot of the ZCTA) and residential population density (persons per cubic foot of ZCTA residential volume). Our final list of variables was based on an iterative process according to: 1) maximizing model fit, measured by the WAIC, 2) removing variables that were highly correlated with each other ($|\tau| \geq 0.9$), and 3) our understanding of underlying social processes in relation to infectious disease. The model also

includes covariate adjustment for the population-adjusted total number of tests administered per ZCTA to account for variation in disease surveillance.

Capacity to Social Distance. Our BWQS model uses cross-sectional administrative data to create an infection risk index, but we wanted to assess the degree to which those differences in infection were explained by inability to socially isolate/distance leading up to May 7, 2020. We utilized MTA transit data as a longitudinal proxy for social distancing where mobility via public transit may serve as a proxy for the conditions that contribute to greater exposure risks. Subway stations are represented in a fraction of NYC ZCTAs, so we aggregated subway utilization and our BWQS index to the United Hospital Fund (UHF) neighborhood level. UHF polygons are composed of adjacent ZCTAs approximating community districts. Aberrantly low utilization observations (<10%) in February and early March 2020 were removed when explained by planned weekend service changes - specifically those in low subway density areas. We computed a population-weighted BWQS index per UHF. Relative subway utilization is a proportion, therefore the transition from business-as-usual to social distancing roughly followed a sigmoidal decay. A mean nonlinear response can be modeled by nonlinear least squares when a functional form is specified, as implemented by the *drc* package in R (Ritz et al., 2015). We utilized a Weibull distribution formula, which took the following functional form:

$$relative\ use = \ c + (d - c)(1 - exp(b(log(time\ index) - log(e))))) \quad ,$$

where *c* is the lower asymptote, *d* is the upper asymptote, *b* is the slope, *time index* is transformation of the date as an integer, *e* is the midpoint *time index* between *c* and *d,* and *relative use* is the proportion of subway ridership. The model accommodates curve fitting with interaction terms to identify differences in model fit per group. For ease of interpretation and visualization, this was utilized to assess differences for high (above the median) versus low (below the median) BWQS index neighborhoods. We sought to identify differences in slope (*b*) and the lower asymptote (*c*) as indicative of differences in the ability to socially isolate. An F-test

was used to compare a naive model (without considering the BWQS index) to a model with interaction by high versus low BWQS index.

Neighborhood infection risk and mortality. Given high COVID-related mortality in disadvantaged communities, we wanted to assess if our measure of neighborhood infection risk was also associated with cumulative COVID mortality as of May 23, 2020, a period after our infection risk index. To do so, we employed a negative binomial model, regressing ZIP code-level COVID mortality on the BWQS infection risk index, adjusting for the proportion of the population that was greater than or equal to 65 years old. In order to adjust for spatial autocorrelation, and thus unmeasured spatial confounding, we employed a spatial filtering approach whereby we identify the eigenvector associated with spatial autocorrelation (as measured by Moran's I), and explicitly include those fitted values in the negative binomial regression (Griffith & Peres-Neto, 2006). The goal, then, is to "filter out" spatial autocorrelation from the residuals. Negative binomial models were implemented with the *MASS* package, supplemented with spatial functions from the *spdep* and *spatialreg* packages (Bivand et al., 2013; Bivand & Piras, 2015).

Mapping and coding

Geoprocessing and visualization of spatial data were conducted with the *sf* package in R (Pebesma, 2018). All analyses were conducted in R version 3.6.2 (R Core Team, 2019). Code for the analysis provided in ***Supplemental File 1***, with reference to required GitHub repositories when relevant.

Results

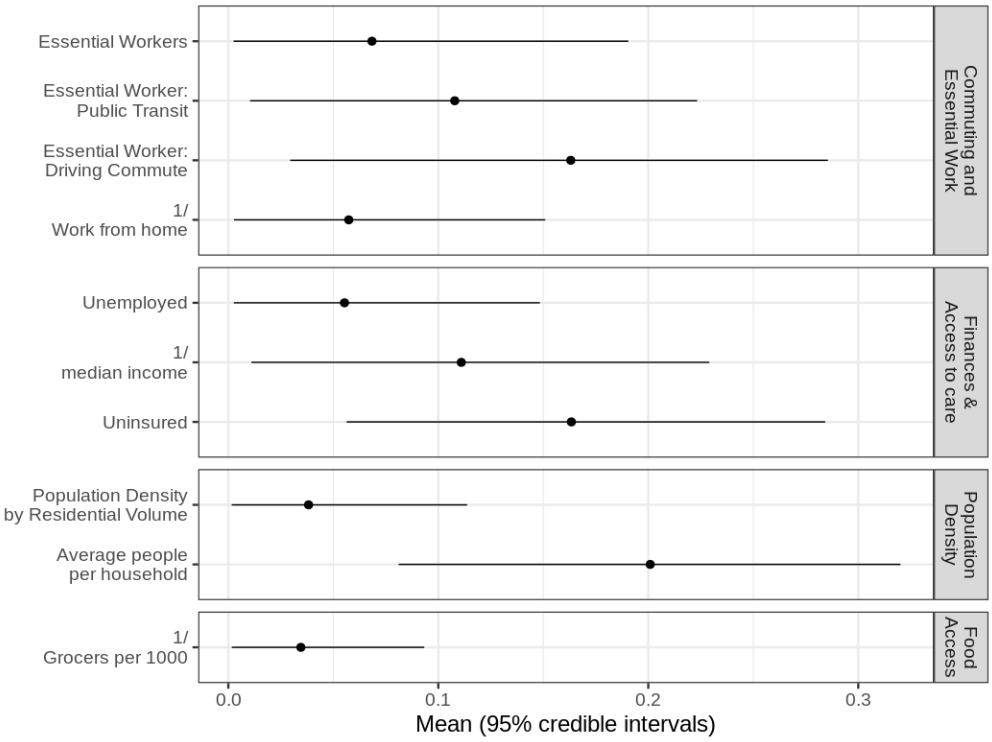
Cross-sectional neighborhood infection risk index

We wanted to identify an association between a neighborhood social disadvantage composite index and ZCTA-level COVID-19 infections. We performed bivariate Kendall's tau correlation tests for each variable and the number of infections per 100,000 (**Supplemental**

Table 2). An assumption of the BWQS regression is that the direction of the effect for each variable is the same as the overall effect. Given the a priori hypothesis that increased disadvantage yields higher infections, we used the inverse of variables that had negative effects.

The BWQS regression analysis identified evidence of an association between our composite variable of ZCTA-level social disadvantage and the number of infections per 100,000. We found that each decile increase in social disadvantage is associated with a 10% increase in infections per capita (Risk Ratio: 1.10; 95% Credible Interval: 1.08, 1.11). While all included variables contributed to this composite, they do not all contribute equally (**Figure 1**). We found that the average number of people in a household is the single largest contributor, followed by the proportion of the population who are essential workers and rely on personal vehicles or public transit to commute. Proportion of uninsured and the median income are also relatively informative compared to the other variables.

Figure 1: Estimated contribution of social variables to the BWQS index, with 95% credible intervals.



The spatial distribution of BWQS index (**Figure 2**) largely mirrors that of infections in NYC (**Supplemental Figure 1**). We examined the population demographics of neighborhoods at various quantiles of our BWQS according to reported race/ethnicity from the ACS (**Figure 3**). Our results demonstrate that White populations are overrepresented in ZCTAs in the lower quartile of the infection risk index (<25th percentile) and underrepresented in the upper quartile of infection risk (>75th percentile) ZCTAs. While Whites comprise approximately 32% of NYC's population, they only make up 11% of high infection risk ZCTAs. Conversely, Blacks and Hispanic/Latinx are 22% and 29% of NYC's population and 31% and 42% of high risk areas respectively.

Figure 2: Map of COVID-19 Infection Risk Index (BWQS index ranging from 0-9) by ZIP Code Tabulation Area.

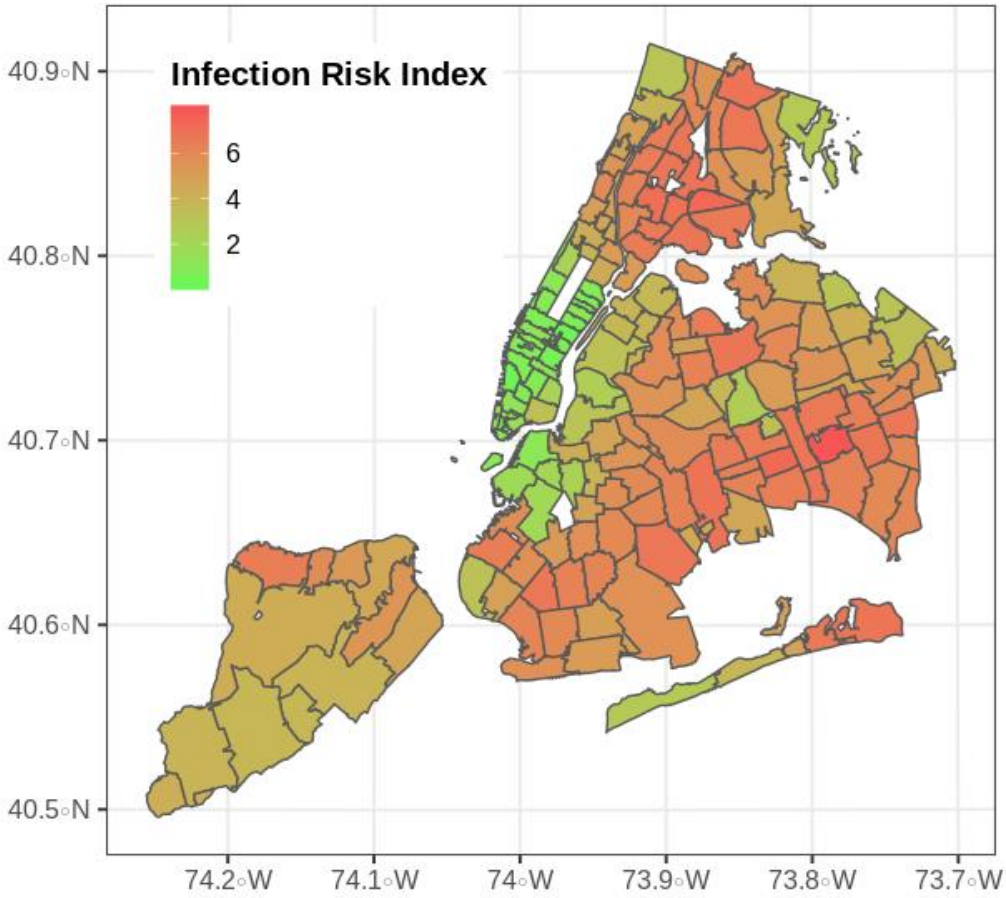
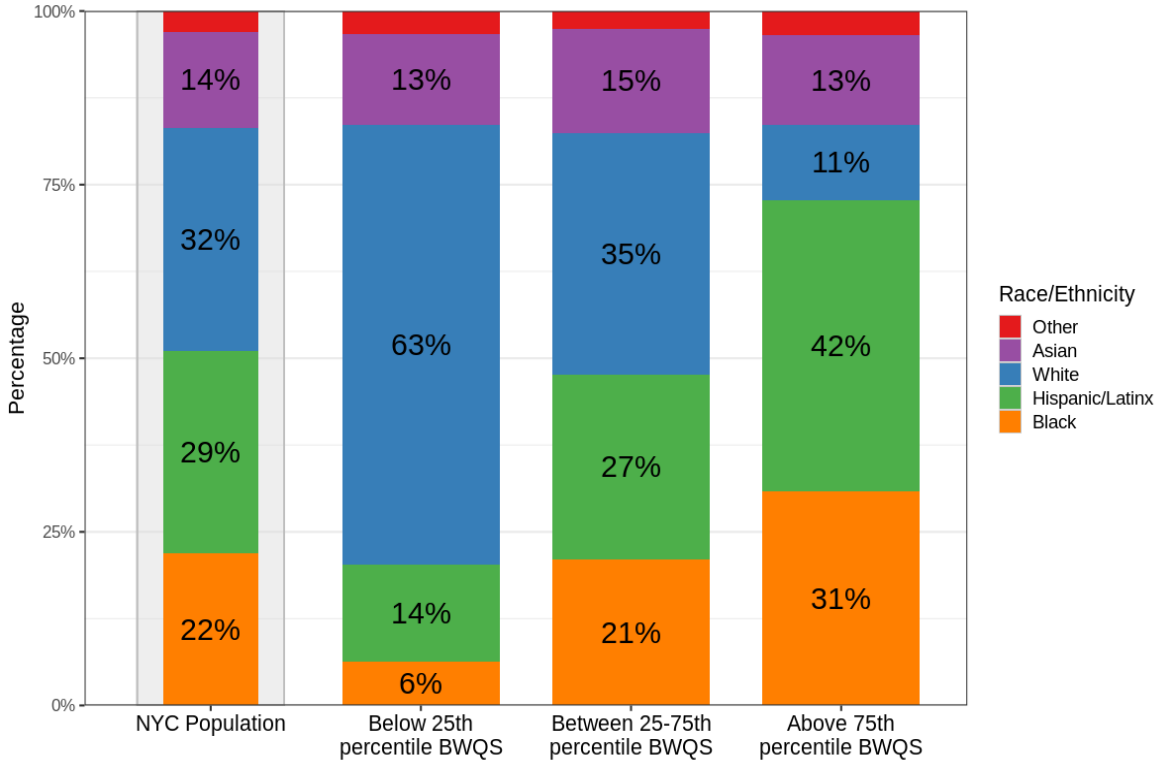


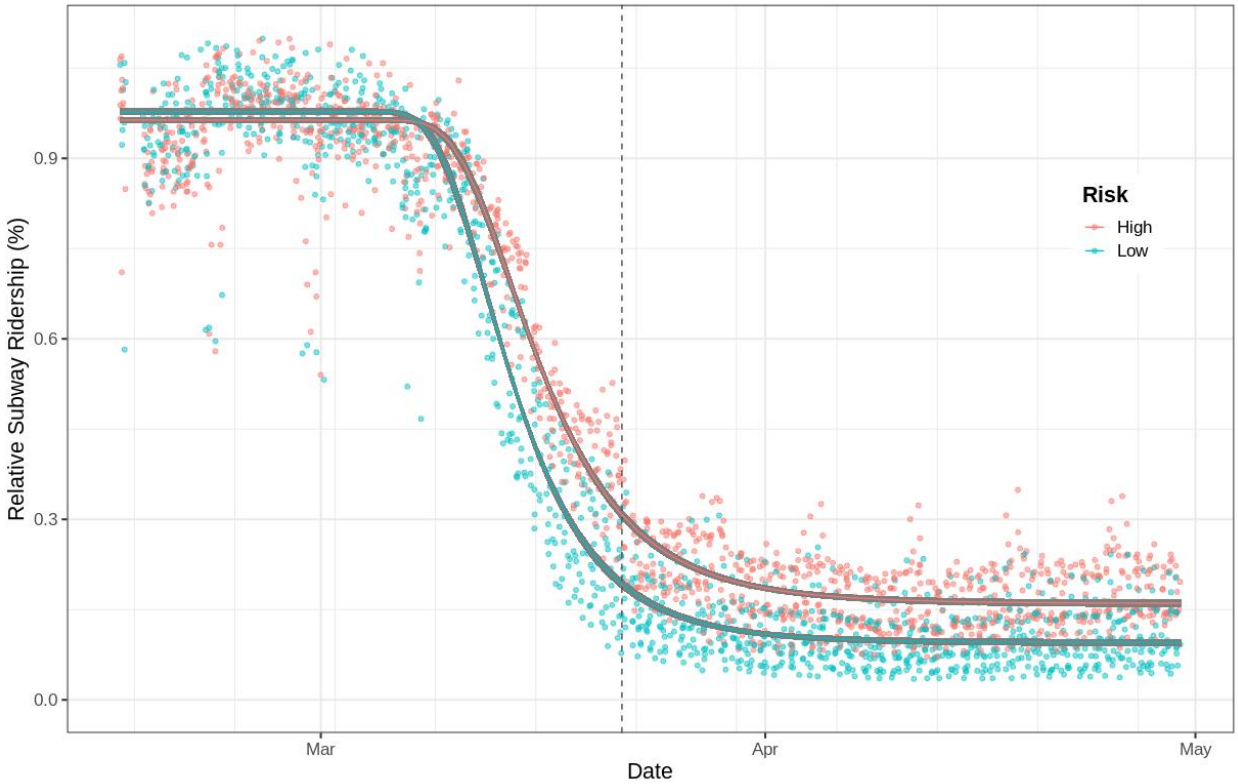
Figure 3: Race/ethnic composition of areas that fall within various quantiles of the ZCTA-level COVID-19 infection risk index. NYC total demographic breakdown provided as reference.



Capacity to social distance

We found that capacity to social distance appears lower in higher neighborhood infection risk areas, as indicated by the most important variables in our neighborhood infection risk analysis. To assess whether or not this was true using longitudinal data, we decided to model differences in subway utilization by UHFs in NYC. We only included UHFs with the most consistent data quality and/or that had subways present (**Supplemental Figure 2**). In order to identify the proper functional form of our Weibull equation, we fit it on the mean sigmoidal decay of subway utilization across all of NYC (**Supplemental Figure 3**). We then compared this model to a stratified model for UHF-level population-weighted BWQS index (**Figure 4**). An F test demonstrated that a model with an interaction term for BWQS index (above versus below the median) was a significantly better fit than one without the interaction term ($p<0.0001$).

Figure 4: Predicted response from nonlinear model with Weibull fit stratified by median BWQS infection risk index. Dashed line represents the beginning of enforced New York State social distancing policies.



The stratified model indicates that there is little difference between slopes for the high (-5.6% per day; 95% CI: -5.9, -5.3%) versus low (-6.3% per day; 95% CI: -6.7, -5.9%) infection risk areas (**Table 1**). However, the lower asymptote of subway utilization under social distancing policies is higher for high infection risk (16%; 95% CI: 15.3, 16.7%) areas compared to low risk infection risk areas (9.6%; 95% CI: 8.8, 10.1%).

Table 1: Coefficients from nonlinear regression.

Curve (Infection risk)	Term	Coefficient	95% CI
High	Slope	-5.6% per day	-5.9, -5.3%
	Lower asymptote	16%	15.3, 16.7%
Low	Slope	-6.3% per day	-6.7, -5.9%
	Lower asymptote	9.6%	8.8, 10.1%

Mortality related to neighborhood infection risk index

Results from a negative binomial model show an association between the BWQS ZCTA infection risk index and cumulative COVID mortality by May 23, 2020 (**Table 2**). This regression model employed a spatial filtering approach to account for potential spatial autocorrelation at the ZCTA level. We found that each point (decile) increase in the BWQS index is associated with a 20% increased risk of COVID-related mortality (Relative Risk: 1.20; 95% CI: 1.16, 1.24) when adjusting for the number of individuals 65+ and accounting for spatial dependence. There was still some evidence of residual spatial autocorrelation in the residuals (Moran's I p value = 0.044).

Table 2: Results from a negative binomial regression of BWQS neighborhood risk index and COVID-19 mortality, using a spatial filtering method.

Term	Relative Risk (95% CI)	P value
BWQS Index	1.20 (1.16, 1.24)	<0.0001
65 plus (z-score)	1.05 (0.99, 1.12)	0.137
Eigenvector of spatial element	3.7	0.0011

Discussion

We conducted a study using publicly available data to identify the role of neighborhood social disadvantage on cumulative COVID-19 infections and COVID-19-related mortality. The neighborhood infection risk index was also used to understand differences in social distancing, as measured by subway ridership. In creating our neighborhood infection risk index, we found that a combination of social variables, indicative of social disadvantage, is associated with cumulative infections and mortality. Black and Hispanic/Latinx communities are overrepresented in high infection risk neighborhoods, and Whites are overrepresented in low infection risk neighborhoods, which may represent structural forms of racism. When examining differences in capacity to socially isolate, we found that high risk neighborhoods had higher subway ridership during NYS-mandated social distancing. Finally, our neighborhood infection risk index is also

associated with cumulative COVID-19 mortality at the ZIP code level. This implies that the same social factors that inform increased disease risk are also associated with severe outcomes, either directly or through intermediates.

A growing body of literature is examining the greater impact of COVID-19 on communities of color. However, as some have noted, COVID-19 is not creating new health disparities, but exacerbating those that already exist (Dorn et al., 2020). A recent investigation found that county and ZCTA area-based socioeconomic measures were useful in identifying higher COVID-19 infections and mortality, specifically using crowding, percent POC, and a measure of racialized economic segregation in Illinois and New York (C. Chen Jarvis T. & Krieger, 2020). Work on COVID-19 mortality in Massachusetts has found excess death rates for areas of higher poverty, crowding, proportion POC, and racialized economic segregation (J. Chen et al., 2020). Similarly, researchers have begun to identify counties that are particularly susceptible to severe COVID-19 outcomes using a combination of biological, demographic, and socioeconomic variables to assess vulnerability (Chin et al., 2020). They identify areas with high population density, low rates of health insurance, and high poverty as particularly at risk. However, a stated limitation of this work is that many of these variables are interrelated.

Our study has many strengths. First, we acknowledge and address the strong interrelation of social variables by using a data-driven method for modeling mixtures of exposures: BWQS. By using this framework, we create a composite index (with a ten unit range) that captures the combined effect of the constituent variables. This process is also supervised, meaning that the variables are not weighted equally in the composite index, but instead the approach empirically learns their individual contributions to explaining the outcome. Others have addressed the multicollinearity of social determinants with the use of dimensionality reduction techniques such as principal components analysis (PCA) in the case of the neighborhood deprivation index (Messer et al., 2006). However, a more parsimonious PCA representation of a multi-faceted construct may not be suited to explain a related phenomenon

without supervised training. Although we did not evaluate our BWQS index outside of NYC, our approach which largely relies on ACS data available across the USA may allow for the identification of other communities particularly vulnerable to future outbreaks or even other novel respiratory viruses. We explicitly excluded race and ethnicity from the creation of the index because we were more interested in identifying social processes that may facilitate infection risk, rather than those that may imply biological or behavioral explanations to health disparities (Chowkwanyun & Reed, 2020). To demonstrate this, we employed the index to understand neighborhood differences in capacity to social distance. This finding provides additional evidence that low-income communities and communities of color may be less able to socially distance, thus representing a pathway for racial infection disparities (Yancy, 2020). Finally, our spatial analysis of COVID-19-mortality shows that the BWQS risk index may not only be useful in identifying infection risk, but also risk of severe outcomes.

This study also has notable limitations. First, we were unable to identify a measure of multigenerational housing at the ZCTA level, which may represent a pathway for infection, and potentially severe disease. Second, by not including race in our models, we may be missing an opportunity to tune these models to the impacts of individual and structural forms of racism (Geronimus et al., 2006). Third, early testing data in NYC was largely limited to hospitalized individuals, therefore those with more severe disease (NYC DOHMH, 2020). Consequently, ZCTA infection data may be confounded by the distribution of factors that drive disease severity. This is among the reasons we adjusted our BWQS regression for the amount of overall testing per ZCTA. Relatedly, for our spatial analysis of COVID-mortality, we were unable to access a ZCTA-level measure of chronic diseases. Since communities of color have higher rates of chronic disease at younger ages (Gee et al., 2019), and chronic diseases increase the likelihood of severe COVID-19 outcomes, this is an important challenge. However, because social disparities are a major contributor to differences in the chronic conditions that increase the likelihood of severe disease, we did not want to adjust for a causal intermediate. Instead, we

adjust for spatial autocorrelation to account for residual risk factors that are more similar in nearby neighborhoods. Fourth our analysis of public transit only utilized data from subway turnstiles, but not bus ridership. Although buses are an important form of transit in NYC, especially in the outer parts of the boroughs, the MTA does not provide time-varying ridership data. Further, buses were made free during the pandemic, so accurate ridership data are likely unavailable to the NYC government as well (Guse, 2020). Finally, an unfortunate potential consequence of creating a neighborhood risk index is the possibility of stigmatization of neighborhoods with high risk index values (Chowkwanyun & Reed, 2020). This is not our intention, and hopefully not the effect, as our goal is to identify social factors that facilitate viral spread, and demonstrate that current public health guidance is not equally observable by all populations. Therefore, it is up to policymakers and practitioners to identify those populations and design/implement interventions accordingly.

Conclusion

In this study, we created a neighborhood measure of social disadvantage that is specifically tuned to the impacts of COVID-19 infections and mortality and we show that this measure is associated with the capacity to socially distance, which may represent an important pathway for COVID-related health disparities. This is an important area of investigation given the large toll that COVID-19 has had, and will likely continue to have, on disadvantaged communities of color in NYC and elsewhere.

Acknowledgement: This work was supported by grant UL1TR001433 and P30ES023515. DC is funded by NIH T32HD049311. Thanks to Kodi Arfer for developing the procedures and indices for relative subway utilization and for programming support.

Contributions: DC and ACJ conceptualized the study and DC drafted the manuscript. DC conducted all analyses with statistical support from EC and NFP. ACJ, EC, and ND provided feedback on design and analysis. The Bayesian Weighted Quantile Sum regression was designed and implemented by EC and NFP, with the log link function implemented by NFP. JR ingested DOHMH data. All authors reviewed and approved the manuscript.

Competing interests: The authors have no conflicts of interest to report.

Data availability: All data are publicly available, and their sources are referenced.

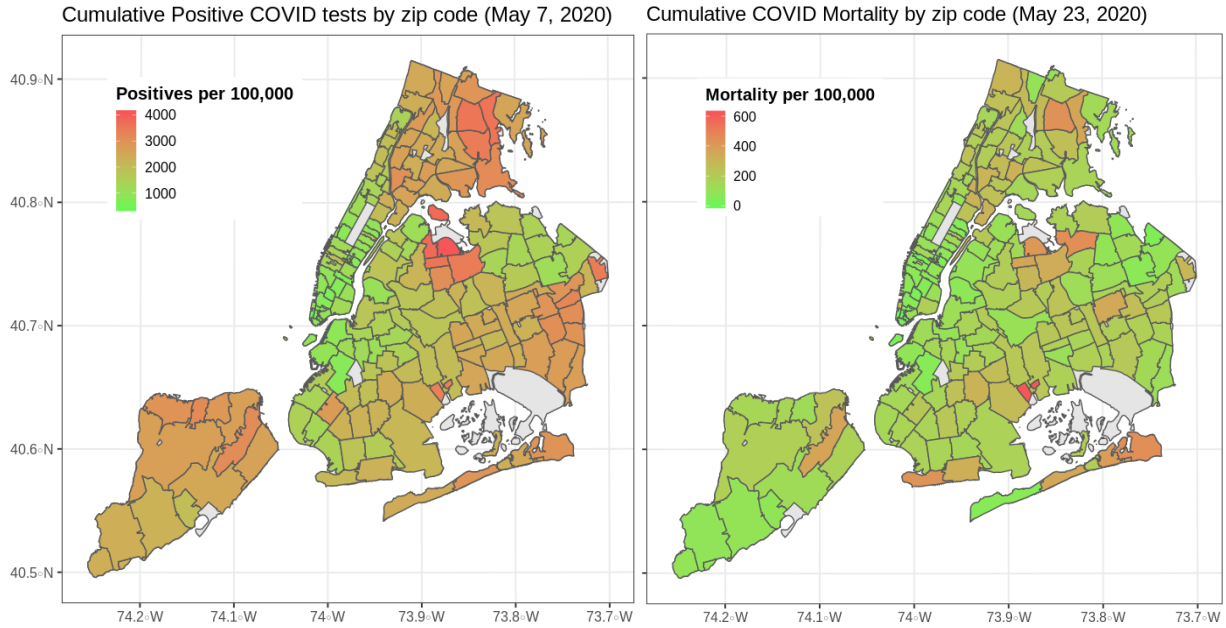
Supplemental materials:

Supplemental Table 1: 2018 American Community Survey (ACS) 5-year estimate variables collected via *tidycensus* R package for ZIP code tabulation area (ZCTA) units

ACS Variable Name	Short description	Explanation
B19013_001	Median income	Measure of neighborhood-level access to financial/economic resources
B06012_002	Federal Poverty Level (below 100%)	
B06012_003	Federal Poverty Level (between 100-150%)	
B23025_005	Unemployed	
B22003_002	Households with supplemental nutrition assistance (SNAP)	Financial/economic resources and food access
B01003_001	Total population per ZCTA	For adjustment and crude population rate calculations
B22003_001	Total households per ZCTA	To create household-based proportions
B27010_017	Uninsured (Under 19 years old)	Access to care and medical information/warnings
B27010_033	Uninsured (19-34 years old)	
B27010_050	Uninsured (35-64 years old)	
B27010_066	Uninsured (65+ years old)	
B08301_021	Work from home	Capacity to socially distance in early stages of pandemic
C24050_002	Employed by: Agricultural industry	Industries that are likely to employ essential workers
C24050_003	Employed by: Construction industry	
C24050_005	Employed by: Wholesale trade	
C24050_007	Employed by: Transportation industry	
C24050_011	Employed by: Education and Healthcare industry	
B08126_017	Employed by: Agriculture, Commute: personal car	Mode of transportation for industries employing essential workers

B08126_032	Employed by: Agriculture, Commute: carpool	
B08126_047	Employed by: Agriculture, Commute: Public Transit	
B08126_018	Employed by: Construction, Commute: personal car	
B08126_033	Employed by: Construction, Commute: carpool	
B08126_048	Employed by: Construction, Commute: Public Transit	
B08126_020	Employed by: Wholesale trade, Commute: personal car	
B08126_035	Employed by: Wholesale trade, Commute: carpool	
B08126_050	Employed by: Wholesale trade, Commute: Public Transit	
B08126_022	Employed by: Transportation, Commute: personal car	
B08126_037	Employed by: Transportation, Commute: carpool	
B08126_052	Employed by: Transportation, Commute: Public Transit	
B08126_026	Employed by: Education/Healthcare, Commute: personal car	
B08126_041	Employed by: Education/Healthcare, Commute: carpool	
B08126_056	Employed by: Education/Healthcare, Commute: Public Transit	

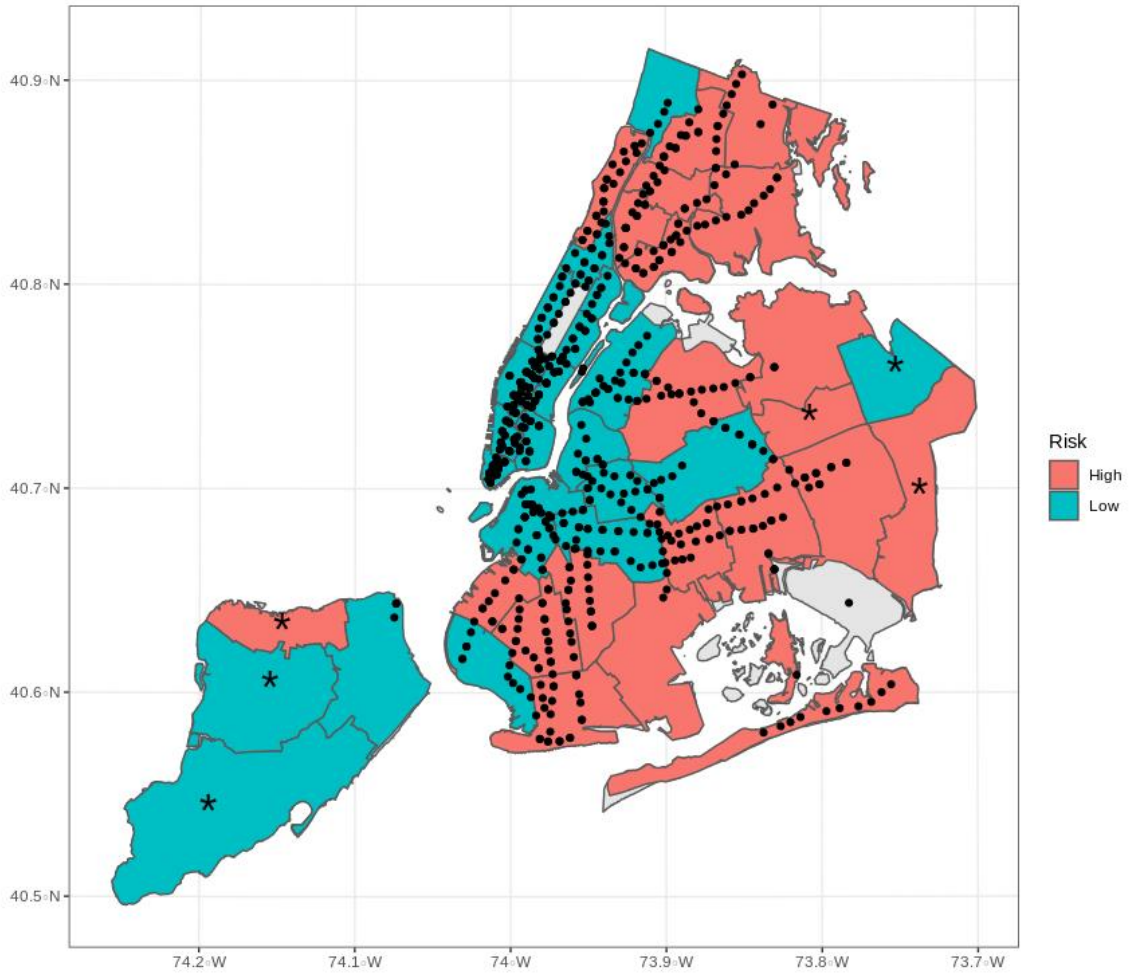
Supplemental Figure 1: Cumulative infections and COVID mortality per 100,000 by ZCTA.



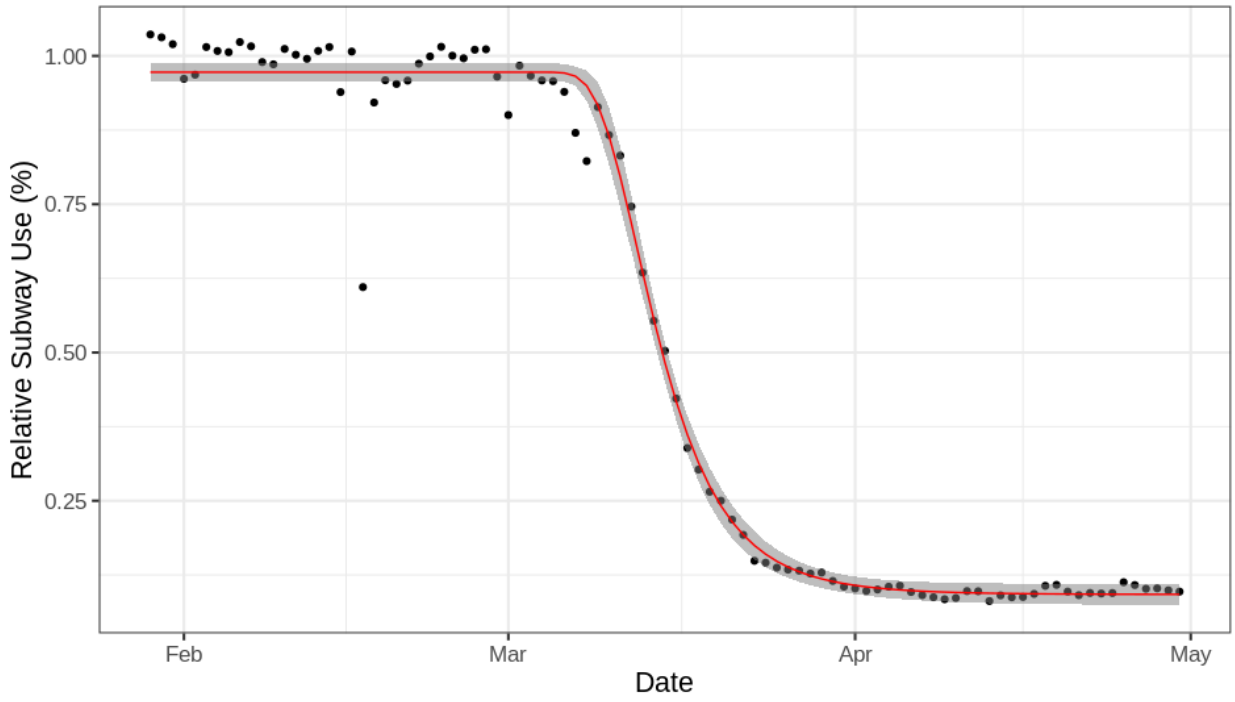
Supplemental Table 2: Bivariate kendall's tau correlations for social determinants with per capita cumulative SARS-CoV-2 infection rate as of May 7, 2020 by NYC ZCTA.

Variable	Correlation	P value
1/median income	0.317	<0.0001
Not insured (%)	0.231	<0.0001
Unemployed (%)	0.296	<0.0001
1/grocers per 1000 people	0.261	<0.0001
Essential workers (%)	0.544	<0.0001
Essential workers commuting via public transit (%)	0.174	0.0001
Essential workers commuting via car (%)	0.493	<0.0001
People who do not work from home (%)	0.454	<0.0001
Population density according to housing volume (people per cubic foot)	0.096	0.058
Average household size (#)	0.467	<0.0001

Supplemental Figure 2: UHF neighborhoods by population-weighted neighborhood infection risk index. Those excluded from the subway ridership analysis indicated by asterisks. Dots represent subway stations with available data. New Jersey PATH stations excluded from this analysis.



Supplemental Figure 3: Fit of the Weibull distribution curve on the citywide mean of UHF-level ridership. Outlier on February 15th, a national holiday. Values above 1 indicate above-average ridership per our day of week comparison of years prior.



References:

Acevedo-Garcia, D. (2000). Residential segregation and the epidemiology of infectious diseases. Social Science & Medicine, 51(8), 1143–1161. https://doi.org/10.1016/S0277-9536(00)00016-2

Acs, G., & Loprest, P. J. (2009). *Job Differences by Race and Ethnicity in the Low-Skill Job Market* (No. 4; Employers in the Low-Skill Labor Market, p. 6). The Urban Institute. https://www.urban.org/sites/default/files/publication/30146/411841-Job-Differences-by-Race-and-Ethnicity-in-the-Low-Skill-Job-Market.PDF

Berchick, E. R., Barnett, J. C., & Upton, R. D. (2019). *Health Insurance Coverage in the United States: 2018* (No. P60-267(RV); p. 44). U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf

Chowkwanyun, M., & Reed, A. L. (2020). Racial Health Disparities and Covid-19—Caution and Context. *New England Journal of Medicine*, *0*(0), null. https://doi.org/10.1056/NEJMp2012910

Colicino, E., Pedretti, N. F., Busgang, S. A., & Gennings, C. (2020). Per- and poly-fluoroalkyl substances and bone mineral density: Results from the Bayesian weighted quantile sum regression. *Environmental Epidemiology*, *4*(3), e092. https://doi.org/10.1097/EE9.0000000000000092

Dorn, A. van, Cooney, R. E., & Sabin, M. L. (2020). COVID-19 exacerbating inequalities in the US. *The Lancet*, *395*(10232), 1243–1244. https://doi.org/10.1016/S0140-6736(20)30893-X

Garg, S. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019—COVID-NET, 14 States, March 1–30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6915e3

Gold, J. A. W. (2020). Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19—Georgia, March 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6918e1

Google. (2020). *COVID-19 Community Mobility Report*. COVID-19 Community Mobility Report. https://www.google.com/covid19/mobility?hl=en

Greco, F., Ventrucci, M., & Castelli, E. (2018). P-spline smoothing for spatial data collected worldwide. *Spatial Statistics*, *27*, 1–17.

Griffith, D. A., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology*, *87*(10), 2603–2613.

Guzman, S. (2019). *Multigenerational Housing on the Rise, Fueled by Economic and Social Changes*. AARP Public Policy Institute. https://doi.org/10.26419/ppi.00071.001

Messer, L. C., Laraia, B. A., Kaufman, J. S., Eyster, J., Holzman, C., Culhane, J., Elo, I., Burke, J. G., & O'Campo, P. (2006). The Development of a Standardized Neighborhood Deprivation Index. *Journal of Urban Health*, *83*(6), 1041–1062. https://doi.org/10.1007/s11524-006-9094-x

Murray, C. J. L., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., Iandiorio, T. J., & Ezzati, M. (2006). Eight Americas: Investigating Mortality Disparities across Races, Counties, and Race-Counties in the United States. *PLoS Medicine*, *3*(9), e260. https://doi.org/10.1371/journal.pmed.0030260

NYC DOHMH. (2020, May 13). *Age-adjusted rates of lab-confirmed COVID-19* [Government]. Age-Adjusted Rates of Lab-Confirmed COVID-19. https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-05142020-1.pdf

Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*, *323*(18), 1775–1776. https://doi.org/10.1001/jama.2020.4683

R Core Team. (2019). *R: A Language and Environment for Statistical Computing* (Version 3.6.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-Response Analysis Using R. *PLOS ONE*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Walker, R. E., Keane, C. R., & Burke, J. G. (2010). Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & Place*, *16*(5), 876–884. https://doi.org/10.1016/j.healthplace.2010.04.013

Webb Hooper, M., Nápoles, A. M., & Pérez-Stable, E. J. (2020). COVID-19 and Racial/Ethnic Disparities. *JAMA*. https://doi.org/10.1001/jama.2020.8598

Yancy, C. W. (2020). COVID-19 and African Americans. *JAMA*. https://doi.org/10.1001/jama.2020.6548