Pandemic Injustice: Spatial and Social Distributions of the first wave of COVID-19 in the US Epicenter

Timon McPhearson, Zbigniew Grabowski, Pablo Herreros-Cantis, Ahmed Mustafa, Luis Ortiz, Christopher Kennedy, Claudia Tomateo, Bianca Lopez, Veronica Olivotto, Avigail Vantu

Urban Systems Lab, The New School, 79 Fifth Avenue, 16th Fl., New York, NY, 10003, USA

Email:

urbansystemslab@newschool.edu

Classification

Social Science, Environmental Science

Keywords

COVID-19, social justice, New York City, social vulnerability

Author Contributions

All authors contributed equally to the conceptual development of the study, TM, PH, LO, AM, ZG, and CK led writing and data collection, ZG, AM, PH, and LO led spatial analysis, CT led graphics, and all authors contributed to final writing, editing and submission.

This manuscript is under review by the Journal of Extreme Events (JoEE).

Abstract

We examine the uneven social and spatial distributions of COVID-19 and their relationships with indicators of social vulnerability in the U.S. epicenter, New York City (NYC). As of July 17th, 2020, NYC, despite having only 2.5% of the US population, has ~6% of all confirmed cases, and ~16% of all deaths, making it a key learning ground for the social dynamics of the disease. Our analysis focuses on the multiple potential social, economic, and demographic drivers of disproportionate impacts in COVID-19 cases and deaths, as well as population rates of testing. Findings show that immediate impacts of COVID-19 largely fall along lines of race and class. Indicators of poverty, race, disability, language isolation, rent burden, unemployment, lack of health insurance, and housing crowding all significantly drive spatial patterns in prevalence of COVID-19 testing, confirmed cases, death rates, and severity. Income in particular has a consistent negative relationship with rates of death and disease severity. The largest differences in social vulnerability indicators are also driven by populations of people of color, poverty, housing crowding, and rates of disability. Results highlight the need for targeted responses to address injustice of COVID-19 cases and deaths, importance of recovery strategies that account for differential vulnerability, and provide an analytical approach for advancing research to examine potential similar injustice of COVID-19 in other U.S. cities.

Significance Statement

Communities around the world have variable success in mitigating the social impacts of COVID-19, with many urban areas being hit particularly hard. Analysis of social vulnerability to COVID-19 in New York City, the US national epicenter, shows strongly disproportionate impacts of the pandemic on low income populations and communities of color. Results highlight the class and racial inequities of the coronavirus pandemic in NYC, and the need to unpack the drivers of social vulnerability. To that aim we provide a replicable framework for examining patterns of uneven social vulnerability to COVID-19 using publicly available data which can be readily applied in other study regions, especially within the USA. This study is important to inform public and policy debate over strategies for short- and long-term responses that address the injustice of disproportionate impacts of COVID-19.

1. Introduction

1.1 Coronavirus impacts in the U.S. and New York City

The first confirmed case of 2019 novel coronavirus (COVID-19) in the United States occurred on January 19, 2020 in Snohomish County, Washington. On February 25, 2020, the Centers for Disease Control and Prevention (CDC) issued their first warning to the American public about a local outbreak. A national emergency was declared on March 13, 2020, and by mid-April, the U.S. death toll had reached 20,000; the highest in the world at the time. As the coronavirus pandemic continues to spread, impacts in the United States have been highest in dense urban areas such as San Francisco, Miami, Chicago, Houston, and New York City (Rosenthal 2020; Desai 2020; Rocklöv and Sjödin 2020). New York City (NYC) in particular emerged as a pandemic “vanguard,” with the largest number of confirmed cases and deaths in the United States (Angel, Tamayo, Lamson-Hall, and Blei 2020). As other regions and cities across the U.S. face a new wave of COVID-19 cases, the NYC case is particularly instructive for understanding how a major city that implemented significant measures to contain the disease nevertheless experienced extreme, and highly socially uneven, incidence of COVID-19 cases, deaths, and disease severity.

In this study, we provide context for NYC’s outbreak and immediate social responses, and provide a replicable analytical framework for using widely available 2018 American Community Survey Data in combination with NYC Department of Health and Mental Hygiene (DOHMH) data on COVID-19 impacts (New York State Department of Health 2020a) (available through their Github repository on zip code level disease incidence) in order to examine the spatial distribution of population normalized prevalence and severity of the disease. Building on the CDC’s Social Vulnerability (SV) indicators (Flanagan et al. 2011), we examined their relationships with COVID-19 indicators of testing, positive cases, deaths, and severity at the zip code level in NYC. We asked: what is the social distribution of COVID-19 testing and associated indicators of confirmed positive cases, deaths, and disease severity? To answer this question we examined the city wide relationships between SV indicators, tests, prevalence, deaths, and severity. Further, we provide a robust examination of how areas with high and low disease-related impacts compare and contrast in relation to their SV indicators.

1.1 The NYC Experience with COVID-19

In New York State, the first confirmed case emerged on March 1, 2020, a 39-year old woman who was a healthcare worker living in Manhattan. Within a week the confirmed cases had risen to 44, with a major outbreak identified in the town of New Rochelle. These early confirmed incidents prompted Governor Andrew Cuomo to declare a state of emergency on March 7, 2020, with New York City to remain on mandatory “P.A.U.S.E” until June 8, 2020 (New York State Department of Health 2020b). Current research indicates that COVID-19 was circulating in New York City through community transmission weeks before the first reported case (NYU Langone 2020). At the time of this writing, there have been 225,045 confirmed cases and 22,845 deaths in New York City, and the city has begun to experience significant declines in daily disease incidence. NYC presently makes up ~6% of national confirmed cases of, and ~16% of deaths from COVID-19 (John Hopkins University 2020), despite only having ~2.5% of

the national population, indicating that other cities and states loosening restrictions may have yet to experience their peak infection levels and mortality (Dave et al. 2020).

Figure 1. Spatial Variation of COVID-19 Impacts. Spatial variation in primary indicators of COVID-19 testing and impacts including % population tested (A), % population with confirmed case (B), % population deceased (C), and estimated severity (D).

In NYC, both overall incidence and population prevalence of the disease show highly uneven spatial distributions with some populations, neighborhoods, and boroughs being much more severely impacted than others (Figure 1; SI Figures S1 to S12). These uneven spatial patterns can be partially explained by existing work on the geography of COVID-19 identifying factors affecting disease prevalence and severity, including population density and age distributions, disease specific factors (e.g. timing of introduction), and social dimensions of the response and reporting mechanisms (CDC COVID-19 Response Team 2020a). Others have found relationships with mortality and susceptibility due to factors such as air pollution (Wu et al. 2020). Both media (Kendi 2020; Mays and Newman 2020; Cruz, 2020; Godoy and Wood 2020; Schwartz and Cook 2020) and scholarly (New York City Department of Health and Mental Hygiene 2020; Perry, Harshbarger, and Romer 2020; Dorn, Cooney and Sabin 2020) work report highly uneven patterns of disease incidence, prevalence, and severity among different races, age groups, and income brackets. Here we expand on these efforts by integrating information on the margins of error (MOEs) inherent to the American Community Survey (ACS), allowing for robust testing of statistical significance of differences between the social characteristics of groups experiencing different rates of disease. In addition, other social vulnerability criteria besides race, income and age, such as rent burden, health insurance coverage rate, housing vacancy, employment and others remain underexplored. Here, we rely on the indicators proposed by the CDC's Social Vulnerability Index to provide a more comprehensive examination of SV indicators, which highlight consistent patterns of disproportionate COVID-19 impacts, and provide analysis important to contribute to ongoing policy discussions over appropriate responses to the pandemic and analytical approaches for examining other aspects of differential vulnerability and exposure both in NYC and other U.S. cities.

In using the indicators chosen by the CDC to assess social vulnerability in the context of emergency management, we frame COVID-19 as an *extreme event* which severely compromised city-wide resilience and communities' capacity to adapt, respond and recover from such a shock. The identification of key vulnerability indicators can help to reveal legacies of social injustice that have led some communities to be especially vulnerable to this extreme event. We hope this will improve ability to the target planning and policy decisions to alleviate specific vulnerabilities and thus allow for a more just distribution of resources, both for the current crisis and future pandemics.

2. Results

2.1 City Wide Patterns and Relationships indicated by Correlations

Relationships between COVID-19 and SV indicators at the zip code level (the highest spatial resolution available) across the city were examined using Spearman ranked order correlations (Figure 2). These correlations provide a broad estimate of the relationships between social vulnerability indicators and COVID-19 testing, confirmed positive and mortality rates. Median income appears to dominate the distribution of all COVID-19 indicators, especially for positive cases and death rates (-0.53 and -0.67, respectively). The inverse correlation between income

and COVID-19 cases and deaths may be related to accessibility to testing as well as the relatively low incomes of essential workers that kept working during shelter-in-place orders. We also note that income impacts may actually be wider, since the ACS caps reported values at 250,000 USD.

Testing positive case rates are highly correlated with the percent of population with a disability, with the second strongest positive link with percent of rent burdened population. Percent of population without health insurance is weakly and positively correlated with testing rates (0.19), while exhibiting a stronger positive association with mortality (0.49). Zip codes with higher proportions of people of color also have strong positive correlations with increased cases (0.57) and death rates (0.53). Death rates also show relatively strong correlations across nearly all social vulnerability indicators, with notable exceptions in % multi-unit lots and % population above 65.

The lack of statistically significant links between % population above 65 and COVID-19 indicators in spite of higher biological susceptibility to the disease among these populations is notable, and suggests that other socioeconomic factors may be more important determinants of severity and death. The lack of statistically significant links may be partially explained by the positive relationship between % population above 65 and median income as well as negative correlations with % crowded units and % people of color (POC), all of which have strong links to COVID-19 indicators.

Meanwhile, indicators that are usually used nationwide as a measure of urban decay and vulnerability such as percent vacancy are associated here with higher incomes and lower COVID-19 risk. This is consistent with reports of vacant luxury apartments having much lower vacancy rates in low- and medium-income areas (Chen 2019). Overall, these measures of statistical association indicate that the distribution of COVID-19 and SV indicators are uneven, and potentially non-linear, prompting the need for further evaluation of zip code level associations of COVID-19 indicators and aggregated SV characteristics.

Fig. 2. Spearman correlation. Spearman correlation between all social vulnerability and COVID-19 indicators for NYC. Hatched squares mark correlations that are not statistically significant ($p > 0.05$).

2.2 Classification of High, Medium, and Low COVID-19 indicators

We used two different methods to examine the high and low prevalence of COVID-19 across neighborhoods in NYC. Zip codes were grouped in quantiles as well as geometric classes to study how COVID-19 impacts are distributed across SV indicators. Quantile based clustering (Figure 3) shows the broad differences between areas by grouping them into equally sized groups, whereas geometric interval clustering isolates areas with particularly high values (for results of geometric interval analyses see SI Figures 14, and 15). Due to the right skewness of the COVID-19 indicator distributions, particularly in severity and mortality rates, geometric clusters in the high end have both much higher values and fewer associated zip codes than in the quantile classification (Table 1). Notably, zip codes in the high COVID-19 response clusters in both methods have (respectively) 30 to 45% higher testing rates, 60 to 45% higher confirmed case rates, 58 to 119% higher mortality, and 37 to 76% higher estimated severity (SI Figure S13) than areas in the low category. We focus here on quantile results in order to emphasize

broad differences between zip codes experiencing divergent incidence of disease, and below on more extreme cases identified through geometric clustering.

Table 1. Values of Quantile and Geometric interval breaks and differences, numerically defined ranges for each low, medium, and high category are provided, as is the % difference between the top of the low range and the bottom of the high grouping (% diff)

COVID-19 Indicator	Quantile				Geometric Interval			
	Low	Med	High	% diff	Low	Med	High	% diff
% Pop. Tested	0.00 - 5.18	5.18 - 6.72	6.72 - 15.08	29.7	0.00 - 6.17	6.17 - 8.92	8.92 - 15.08	44.6
% Pop. Confirmed Cases	0.00 - 1.57	1.57 - 2.52	2.52 - 4.41	60.5	0.00 - 1.80	1.80 - 2.61	2.61 - 4.41	45
% Pop. Deceased	0.00 - 0.12	0.12 - 0.19	0.19 - 0.57	58.3	0.00 - 0.16	0.16 - 0.35	0.35 - 0.57	118.8
Estimated Severity	0.00 - 6.70	6.70 - 9.17	9.17 - 21.71	36.9	0.00 - 7.86	7.86 - 13.86	13.86 - 21.71	76.3

Spatial distribution of high, medium, and low groupings varies across the different COVID-19 response variables studied. Clusters of tests, cases and deaths have a similar distribution, with zip codes in the Bronx, North Queens, North Staten Island, all consistently classified as high. Fewer zip codes in Staten Island are classified as high for deaths than for testing and positive cases, whereas Brooklyn shows the opposite trend. A similar result is observed for the distribution of severity clusters, in which a group of neighboring zip codes in Brooklyn is classified as high, while some parts of the Bronx and almost all the zip codes in Staten Island are classified in the low quantile. Eleven zip codes were consistently classified as a high value cluster according to the quantile classification method for tests, cases, deaths, and severity. Of these eleven zip codes, five are in the Bronx (10469 and 10475 in Northeast Bronx, 10463 in Kingsbridge - Riverdale, 10459 in Hunts Point - Mott Haven, and 10451 in High Bridge - Morrisania), three in Queens (11691 and 11694 in Rockaway, and 11369 in West Queens), one in Brooklyn (11239 in Canarsie - Flatlands), one Manhattan (10035 in East Harlem), and one in Staten Island (10304 in Stapleton - St. George). Excluding tests to only examine areas of negative impacts including high confirmed prevalence, deaths, and severity, four additional zip codes stand out (11203 in East Flatbush - Flatbush, 11236 in Canarsie - Flatlands, 11432 in Jamaica, and 11693 in Rockaway) (Figure 4). See SI Table S1 for a zip code and neighborhood correspondence.

Fig. 3. Spatial distribution of COVID-19 indicators. Spatial distribution of high, medium, and low clusters of COVID-19 indicators (% population tested (A), % population with confirmed case (B), % population deceased (C), and estimated severity (D) for quantile-based classification of high, medium, and low impact clusters.

Fig. 4. COVID-19 impact clusters. Zip codes with consistent groupings as high and low impact clusters for COVID-19 cases, deaths, and estimated severity.

2.3 Differences in underlying SV characteristics by SV dimension

Our findings definitively show that COVID-19 disproportionately impacts communities with lower incomes and a higher proportion of people of color (POC) in NYC. Zip codes in the top quantiles of testing, confirmed cases, mortality, and severity all show consistently higher social vulnerability as indicated by poverty, unemployment, disability, population under 17, language isolation, rent burden, and housing crowding (Figure 5). The largest SV differences occur in percentage of the population identifying as POC (over 80% in the top third of cases and deaths vs ~55% in the bottom third), rent burden (over 70% in the top third of COVID-19 cases and deaths vs 55% in the bottom third), disability rates (~27% in the top quantiles vs ~17% in the bottom, for all indicators) and median household income (~80,000 to 100,000 in the top quantiles and ~50 to 60,000 in the bottom). Income shows a consistent negative significant relationship with rates of death and disease severity, although zip codes with high and medium testing rates and confirmed cases do not have statistically significant differences in median income. Other indicators also show mixed patterns. The percentage of the elderly population (over 65 years old), for example, is consistently lower in the top quantiles of tests, cases and deaths, while it is higher in the top quantiles of severity. This is likely related to the lower

survival rates that have been observed among the elderly. Housing vacancy rates also showed unexpected results by being considerably higher in zip codes with low COVID-19 impacts.

Figure 5. Social vulnerability and COVID-19 indicator values. Comparisons of aggregated SV indicator values for quantile based grouping based on four COVID-19 indicators; percent population tested, with confirmed case, and deceased, as well as estimated severity. Bars represent standard error of aggregated ACS 5-year estimates, letters (a,b,c) indicate statistically significant differences of $p < 0.05$, with each letter corresponding to statistically distinguishable groups.

A closer examination of specific zip codes illustrates the distributional injustices of COVID-19 in NYC. For example, Canarsie-Flatlands (Brooklyn) (zip code 11239) shows the highest percentage of deaths relative to its total population (0.57%). Additionally, this zip code is fourth in severity (15.67%), seventh in percentage of its total population being infected (3.66%) and eighth in percentage of the population tested (9.67%). In this zip code, median annual household income is only \$27,104, 81.2% identifies as POC, and 71.3% of the population is rent burdened. Additionally, 39.4% of the households have at least one person with a disability, and 28.4% of the population lives below the federal poverty line. Hunts Point - Mott Haven (zip code 10459), provides another case in point, with 7.26% of the population (at the time of analysis) tested for COVID-19, 2.95% confirmed, 0.31% deceased, with resultant estimated severity of 10.53%. Besides a high percentage of POC (98.4%), this zip code has a median income of \$27,687. A majority (68.8%) of the population is experiencing rent burden, and 27.9% of its households have one or more people with a disability. Finally, 34% of the population lives below the poverty line, and 18.2% may be experiencing language isolation (For a visualization of the variation per SV indicator across gradients of COVID-19 variables, see SI Figures S12 to S19. COVID-19 and SV data may be directly consulted in the supplementary dataset S1 provided).

Grouping zip codes by COVID-19 response indicators using a geometric approach, as opposed to quantiles, reveals important differences for certain indicators, especially death rates and severity. In severity clusters, rates of language isolation range between ~13% and ~25%, while in the quantiles the values ranged between ~9% and ~14%. The percentage of people above 65 years old also shows a large increase between the high severity geometric and quantile groups, jumping from ~15% to ~20%. Finally, we also observe this change in the percentage of the population living in multi-unit structures, which jumps from ~60% to ~78%. These differences are a reflection of the long tailed distributions of most of the COVID-19 response indicators, a hallmark of social unevenness. Since in the geometric approach low and medium clusters contain a larger sample, median income is closer to the city-wide median and hence returns a lower value than in the quantile approach.

3. Discussion

3.1 Direct impacts of COVID-19 are highly unequal

Overall, the uneven patterns of COVID-19 highlight the role of social vulnerability to the disease in driving health impacts. Vulnerability in the context of a respiratory disease like COVID-19 affects susceptibility, recovery, and exposure, making it a key, if often hidden factor in many epidemiological models of transmission (Michaud, Kates, and Levitt 2020). SV impacts may also help explain coronavirus disease occurrence at the county level across the U.S. (Chin et al. 2020). Many scholars have employed the concept of social vulnerability more generally to explain how socioeconomic status, poor access to healthcare, labor inequalities, household overcrowding, racism, and other factors increase the likelihood of adverse outcomes from disease, natural hazards, and inadequate medical care (Cutter, Boruff and Shirley 2003; Cutter 2020; Link 2008; Drago and Miller 2020; Leclerc 1994). Social vulnerability is also employed by the CDC as a key indicator of why certain subsets of the population are more likely to be

impacted by disasters, including pandemics, as well as face higher risks of more adverse outcomes (Flanagan 2011; CDC COVID-19 Response Team 2020b; DeCaprio et al. 2020).

Strikingly, the indicator for elderly populations (percentage of the population older than 65 years old) shows no significant links to any of the COVID-19 variables at the city-scale in spite of documented physiological susceptibility to the virus (Liu, Chin, Lin and Han 2020). However, zip codes with high severity such as those in the Bronx neighborhoods (Kingsbridge - Riverdale, Hunts Point - Mott Haven, HighBridge and HighBridge – Morrisania) as well as in the Jackson Heights and the Rockaways area of Queens all have significantly higher proportions of elderly than low severity areas. The age disparity is widened in areas with the highest severity, as shown in the geometric interval clustering, which may reflect the higher risks reported in elder care facilities (Applegate and Ouslander 2020), as well as the need to examine age-based vulnerability in relation to other risk factors.

The largest majority of the zip codes identified as “high” across several COVID-19 indicators in the Bronx are home to some of the poorest congressional districts in the country, and with some of the highest rates of asthma, cardiovascular disease, mental illness, and other chronic health conditions, all risk factors for COVID-19 (NYC Health and Hospitals Corporation, 2019). Similar to the Bronx, areas in Queens including the Rockaways and Jackson Heights are home to some of the most diverse communities in the country. However, 18% of residents in the Rockaways and 25% in Jackson Heights are living in poverty according to the City’s 2018 Community Health Assessment. Epidemiological studies highlight the effects of environmental health factors that range from poor air quality, access to quality food and green spaces as playing a key role in determining health impacts (Warman, Silver and Wood 2009; New York City Department of Health and Mental Hygiene 2018). The Citizens’ Committee for Children of New York for instance, found higher rates of asthma among youth in the Bronx, as well significant rates of heart disease, low birth rate, cancer and metabolic diseases (Citizens’ Committee for Children of New York 2013). Organizations such as WE ACT (2017) point to the location of waste transfer stations, bus depots, hazardous waste facilities and large volumes of truck traffic as an environmental justice concern, an example of historical and current structural racism, and may be driving increased social vulnerability to COVID-19 in low income and neighborhoods predominantly home to POC (Calcagno 2013).

Our analysis highlights the class and racial inequities of the coronavirus pandemic in NYC, which have significant disproportionate impacts on communities of color and economically precarious communities. The results of our clustering technique and hotspot analysis point to key dimensions of disproportionate impacts, and in particular to racial and ethnic marginalization generally and especially for extreme cases such as language isolation. The inequality of impacts is reflected not just at the zip code level but also the demographics of COVID-19 deaths. For example, Black and Hispanic/Latinos account for 64% of all deaths, while accounting for about half of the total population of NYC (Kendi 2020).

In NYC, issues of poverty, race-based inequity and access to healthcare are pronounced for POC including Latino/a, African American, Black, and low-income communities (Calcagno 2013; Parrot 2019) and have been well described for vulnerability to heat waves, flooding or other extremes (Anguelovski et al. 2016). In relation to a pandemic, O’Sullivan and Bourgoin (2010) identify a social gradient of risk related to social vulnerability increasing exposure probability linked to living conditions rather than lifestyle choices. This phenomenon can be

clearly seen in NYC subway ridership in the early days of the outbreak (SI Figure S20). Consistently, subway ridership in zip codes with lower median income and higher proportions of people below poverty and without health insurance was higher, with ridership reductions ranging from 65% reduction in low income areas to nearly 100% in the wealthiest neighborhoods.

3.2 Additional considerations

Although the majority of the CDC's SV indicators were considered throughout our analysis, other aspects of vulnerability may also need to be considered to further improve understanding of prevalence, distribution, and severity of COVID-19 in U.S. cities. For example, factors such as differences in life expectancy (Lamantia 2019), family structure, wealth, and ability to vacate may help explain COVID-19 impacts. Additionally, relationships between labor and housing precarity (Urban Systems Lab 2020), for instance, may require further study, as well as environmental conditions that may impact health outcomes such as poor indoor and outdoor air quality, inadequate access to food, population density and intergenerational cohabitation rates, or uneven access to greenspace (important for social distancing and mental and physical health).

The varied responses to the COVID-19 crisis are also a critical component to consider. In New York State and around the region, there was little consistency in shelter-in-place and social distancing policies at the federal, state and local level. This variability in responses may have played a critical role in health and economic impacts, especially for those most vulnerable. Furthermore, labor force characteristics may also have contributed to COVID-19 vulnerability. For example, according to the U.S. Bureau of Labor Statistics (2018), people identifying as Black and Hispanic are more likely to be employed in services occupations (healthcare support, food preparation and serving, building cleaning, and personal care) and in production, transportation, and material moving occupations, than self-identified Whites or Asians. Thus, POC are more likely to be employed in essential service industries with increased risk of exposure to COVID-19.

Daily updates of available COVID-19 data imply potential lags in reporting of the tests, positive cases and deaths that happened the previous days due to delays in the testing and reporting processes. Moreover, the CDC's framework for assessing social vulnerability may not be comprehensive in some respects, especially for urban areas. Additionally, we found high correlation between some SV indicators, including high variance inflation factor values for some of CDCs indicators used, which were subsequently removed from analysis. It is not surprising that many SV indicators are strongly correlated with one another, given that vulnerability is often compounding and intersectional. Sadly, it has been understood for quite some time that persistent racism has extreme impacts on health and well-being of POC in America, through intersecting factors affecting economic opportunities, uneven patterns of policing, access to quality housing, health insurance, and likelihood of experiencing poverty, along with other stress factors (Rothstein 2017; Harrell 2000). Such intersections of different kinds of vulnerability highlight that social vulnerability does not occur in a vacuum, rather it is produced in particular contexts by entrenched modes of decision-making and broader patterns of urban governance (O'Brien et al. 2007). Paying attention to the social forces producing vulnerability, rather than simply taking indicators at face value, is important in order to avoid framings of resilience that

obscure, rather than illuminate the underlying social drivers requiring resilience to hazards with uneven social distributions (Kaika 2017).

4. Materials and Methods

4.1 Approach

Data from the American Community Survey (ACS) 5-year estimates was gathered to develop the 15 social vulnerability indicators defined in CDC's Social Vulnerability Index (Centers for Disease Control and Prevention 2020), as well as three additional indicators (availability of health insurance, rent burden and vacant housing). We used a Variance Inflation Factor (VIF) analysis to remove indicators that showed collinearity, selecting 15 out of the initial 18. Social vulnerability indicators at the zip code level were joined with data on COVID-19 testing, positive cases, and deaths available at New York City Department of Health and Mental Hygiene's (2020) github data repository. A city-wide analysis was first carried out to explore the magnitude and direction of associations between the social vulnerability and the COVID-19 indicators. To incorporate the influence of the margins of error attached to ACS data, we proceeded to group zip codes into clusters of Low, Medium and High incidence of each COVID-19 indicator. Clusters were then compared by testing the statistical significance of their differences.

4.2 Data and Sample

4.2.1 ACS Data

The most updated 5-year estimates of the ACS (2014-2018) were retrieved at the zip code (zip code) level. We developed the 15 social vulnerability indicators originally selected by the CDC to develop its Social Vulnerability Index (Flanagan et al. 2018). These 15 indicators are classified in four groups: socioeconomic status, housing composition and disability, minority status and language, and housing type and transportation. Three additional indicators were added to the ones defined by CDC in order to represent access to healthcare (*percent population without health insurance*), a measure of the economic stress induced by housing costs (*percent population experiencing rent burden*), and a coarse indicator of housing availability (*percent vacant housing units*).

Building the indicators required aggregating records from the ACS and calculating their percentages based on total population or household counts. For instance, the indicator *percentage of the total population above 65 years old* was developed by first combining the estimates of female and male individuals across several age brackets higher than 65. During the aggregation of records and the calculation of the percentage over the total population, we relied on the guidelines published by the US Census Bureau (2018) on handling margins of error during calculations.

We examined the multicollinearity of the indicators selected using the variance inflation factors (VIF) test to avoid two or more indicators from being linearly related. Some guidelines have set $VIF \geq 5$ or even higher as a cutoff point to indicate serious

multicollinearity (Snee 1973; New York State Department of Health 2020c). We, therefore, considered only 15 indicators that showed VIF values ≤ 5 (Table 2) in our further analysis. Examples of indicators with VIF above this threshold include % households with single parents and % of households without a car.

Table 2: Summary statistics for the selected social vulnerability indicators in New York City's zip codes, arranged in the different thematic groups defined by CDC with the addition of population without health insurance, rent burden, and vacant housing indicators. All indicators represent percent of the zip code's population (a), occupied households (b), or available housing units (c), except for median income in US\$.

Group	Indicator	Minimum	Maximum	Average	Median
Socioeconomic status	Below Poverty ^(a)	2.05	45.36	16.33	13.70
	Unemployment ^(a)	0.60	15.76	6.61	6.13
	Median Income	21149.00	250001.00	73901.11	66483.00
	Health Insurance ^(a)	0.59	23.82	7.60	6.96
	Rent burden ^(b)	23.26	76.84	58.84	60.18
Household composition and disability	Above 65 ^(a)	0.46	28.98	14.30	13.56
	Below 17 ^(a)	6.35	36.23	19.93	20.18
	Disability ^(b)	0.00	40.01	21.69	21.49
Minority status and language	People of Color ^(a)	8.39	99.24	63.52	63.78
	Language Isolation ^(a)	0.08	41.91	10.75	8.25
Housing type	Multi-Unit ^(a)	0.39	99.68	50.58	51.19
	Mobile Homes ^(a)	0.00	2.08	0.14	0.10

	Crowded households ^(a)	0.94	29.65	8.29	7.21
	Group Quarters ^(a)	0.00	22.02	2.21	1.12
	Vacant Housing ^(c)	1.68	46.53	9.67	7.86

4.2.2 COVID-19 Data.

Absolute numbers of COVID-19 laboratory tests, laboratory confirmed cases and deaths in NYC at the zip code level were obtained from the DOHMH on May 19, 2020. This data has been available for NYC since April 1, 2020, with ongoing daily updates as of the time of this article's preparation, with a single missing day (April 2, 2020). The DOHMH case data is collected in real-time from a combination of state-level records and calls to individual hospitals. Testing totals, positive counts and deaths are assigned to zip codes based on each test subject's residence, with positive counts sourced directly from testing laboratories.

Examining testing is important because the official guidance from the DOHMH, while recommending tests for symptomatic individuals or those with known exposure to other confirmed cases, provides significant physician discretion (New York State Department of Health 2020c). Both testing and confirmed case data therefore are estimates of overall population prevalence using best available knowledge, which may be updated at a later date as new information comes to light. In addition to the indicators gathered from the DOHMH repository (tests, positive cases and deaths), we estimated zip code level severity, by calculating percentage of confirmed cases resulting in deaths, an important metric of disease impact in the absence of more detailed clinical information (Ruan 2020).

Data distributions were examined visually and using Shapiro-Wilks tests for normality, yielding the insight that while tests are roughly normally distributed among zip codes, confirmed cases, deaths, and estimated severity are not, displaying long tails of high-end values.

4.3 Analysis

4.3.1 City wide

The nonparametric Spearman correlation was employed to assess the monotonic relationship between our indicators (Spearman 1906). Spearman correlation is an extension of Pearson's product-moment correlation that, unlike Pearson's correlation, does not require normally distributed data in order to examine associations between ranked observations of each indicator, computing statistical significance based on observation order (Borkowf 2002).

4.3.2 Clustering Zip Codes based on COVID-19 Indicators

Clustering techniques widely used to detect spatial clustering of high or low values include Getis-Ord G_i^* (Hotspot analysis), K-means, and agglomerative clustering. In the case of NYC, complex geometries of zip codes and sharp breaks in COVID-19 indicators between adjacent neighborhoods (Choi and Velasquez 2020) preclude the use of common spatial clustering algorithms. We therefore explored two clustering methods based on the non-spatial distributions of COVID-19 indicators, using a standard method of quantile-based grouping in even thirds, and a geometrical interval classification algorithm, to set three clusters of COVID-19 identifying factors (low, medium, and high) (Figure 4). While quantile-based grouping creates evenly sized groups, it may mask differences in highly uneven distributions, whereas geometric interval classification algorithms attempt to maximize differences between even size classes while minimizing the sum of squares of the number of elements in each class and is useful for non-normally distributed data (Arlinghaus and Kerski 2013).

4.3.3 Analysis of SV of COVID-19 clusters

For each of the high, medium, and low groups of zip codes generated for each COVID-19 indicator, we reaggregated their population and calculated their social vulnerability indicators. This aggregation allowed for testing the significance in the difference between clusters using the MOEs provided in the ACS estimates. Significance was tested using the guidelines developed by the US Census Bureau (2018). Including the MOEs in the analysis incorporates the various forms of error within the census estimates themselves, which is critical for testing significance of differences.

5. Conclusion

New York City serves as a warning and a lesson for other cities weighing their own responses to the ongoing crisis. As the coronavirus pandemic continues to evolve, other U.S. cities urgently require a critical examination of what characteristics of their urban populations make them more or less vulnerable to COVID-19. It should be obvious that underlying disparities in health outcomes make clear that persistent environmental and social injustice calls for increased consideration in both responses seeking to contain the spread of the virus, as well as economic and social support for communities disproportionately impacted. Results of this study help make visible how decades of uneven investment in cities, disparities in access to education, affordable food, housing, and healthcare, and disproportionate economic impacts among other factors (Parrott and Moe 2020), which lead to social vulnerability, are also strongly related to disproportionate impacts of COVID-19, especially on low-income and POC populations. These social, economic and health issues are intimately linked to community resilience and reflect a long history of racism and equity issues that have laid the groundwork for a lack of preparedness across city, state and Federal scales (Wallace and Wallace 1997). As cities come to terms with the full extent of the coronavirus crisis, it will be important to undertake a deeper examination of how structural injustices have impacted, and may continue to impact, key decision making, emergency planning and response, and further drive increased (or decreased) social vulnerability to impacts of COVID-19.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 2029918 and 1934933. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Angel, S, Tamayo, MM, Lamson-Hall, P, and Blei, A (2020). The Coronavirus and the Cities: Explaining Variations in U.S. Metropolitan Areas as of 27 March 2020. Maron Institute of Urban Management. <https://marroninstitute.nyu.edu/papers/coronavirus-and-the-cities>
- Angelovski, I, Shi, L, Chu, E, Gallagher, D, Goh, K, Lamb, Z, Reeve, K, and Teicher, H (2016). Equity Impacts of Urban Land Use Planning for Climate Adaptation: Critical Perspectives from the Global North and South. *Journal of Planning Education and Research*, 36(3), 333–348.
- Applegate, WB, and Ouslander, JG (2020). COVID-19 Presents High Risk to Older Persons. *Journal of the American Geriatrics Society*, 68(4), 681.
- Arlinghaus, SL, and Kerski, JJ (2013). *Spatial Mathematics: Theory and Practice through Mapping*. CRC Press.
- Borkowf, CB (2002). Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation. *Computational Statistics and Data Analysis*, 39(3), 271–286.
- Calcagno, J (2013). Trends in Poverty Rates Among Latinos in New York City and the United States, 1990 - 2011. Center for Latin American, Caribbean, and Latino Studies. https://academicworks.cuny.edu/clacsl_pubs/37
- CDC COVID-19 Response Team.(2020a). Geographic Differences in COVID-19 Cases, Deaths, and Incidence - United States, February 12-April 7, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(15), 465–471.
- CDC COVID-19 Response Team (2020b). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69.
- Centers for Disease Control and Prevention (2020). CDC Social Vulnerability Index (SVI) | Data. Data.CDC.Gov. Retrieved July 16, 2020, from <https://data.cdc.gov/Health-Statistics/CDC-Social-Vulnerability-Index-SVI-u6k2-rtt3/data>
- Chen, S (13 September 2019). One in Four of New York's New Luxury Apartments Is Unsold. *The New York Times*. <https://www.nytimes.com/2019/09/13/realestate/new-development-new-york.html>
- Chin, T, Kahn, R, Li, R, Chen, JT, Krieger, N, Buckee, CO, Balsari, S, and Kiang, MV (2020). U.S. county-level characteristics to inform equitable COVID-19 response. *MedRxiv*, 2020.04.08.20058248.
- Choi, A, and Velasquez, J (3 May 2020). Early Precautions Draw a Life-and-Death Divide Between Flushing and Corona. *THE CITY*.

- <https://www.thecity.nyc/health/2020/5/3/21247136/early-precautions-draw-a-life-and-death-divide-between-flushing-and-corona>
- Citizens' Committee for Children of New York (2013). Concentrations of Risk: Asthma and Poor Housing Conditions. <https://www.cccnewyork.org/blog/concentrations-of-risk-asthma-and-poor-housing-conditions/>
- Cruz, D (4 April 2020). Racial Breakdown Of NYC's COVID-19 Cases Necessary To Track The Pandemic's Inequities. Gothamist. <https://gothamist.com/news/racial-breakdown-nycs-covid-19-cases-necessary-track-pandemics-inequities>
- Cutter, SL (2020). The Changing Nature of Hazard and Disaster Risk in the Anthropocene. *Annals of the American Association of Geographers*, 0(0), 1–9.
- Cutter, SL, Boruff, BJ, and Shirley, WL (2003). Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, 84(2), 242–261.
- Dave, DM, Friedson, AI, Matsuzawa, K, and Sabia, JJ (2020). When Do Shelter-in-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across States and Adoption Time (Working Paper No. 27091). National Bureau of Economic Research.
- DeCaprio, D, Gartner, J, Burgess, T, Kothari, S, Sayed, S, and McCall, CJ (2020). Building a COVID-19 Vulnerability Index. ArXiv:2003.07347 [Cs, Stat]. <http://arxiv.org/abs/2003.07347>
- Desai, D (2020). Urban densities and the Covid-19 pandemic: Upending the sustainability myth of global megacities. Observer Research Foundation. Retrieved June 22, 2020, from <https://www.orfonline.org/research/urban-densities-and-the-covid-19-pandemic-upending-the-sustainability-myth-of-global-megacities-65606/>
- Dorn, A, Cooney, RE, and Sabin, ML (2020). COVID-19 exacerbating inequalities in the US. *Lancet* (London, England), 395(10232), 1243–1244.
- Drago, R, and Miller, K (2020). Sick at Work: Infected Employees in the Workplace During the H1N1 Pandemic. Institute for Women's Policy Research. Retrieved June 22, 2020, from <https://iwpr.org/publications/sick-at-work-infected-employees-in-the-workplace-during-the-h1n1-pandemic/>
- Evelyn, K (8 April 2020). "It's a racial justice issue": Black Americans are dying in greater numbers from Covid-19. *The Guardian*. <https://www.theguardian.com/world/2020/apr/08/its-a-racial-justice-issue-black-americans-are-dying-in-greater-numbers-from-covid-19>
- Flanagan, BE, Gregory, EW, Hallisey, EJ, Heitgerd, JL, and Lewis, B (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, 8(1).
- Flanagan, BE, Hallisey, EJ, Adams, E, and Lavery, A (2018). Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers for Disease Control and Prevention's Social Vulnerability Index. *Journal of Environmental Health*, 80(10), 34–36.
- Godoy, M, and Wood, D (2020). What Do Coronavirus Racial Disparities Look Like State By State? NPR.org. Retrieved June 22, 2020, from <https://www.npr.org/sections/health-shots/2020/05/30/865413079/what-do-coronavirus-racial-disparities-look-like-state-by-state>

- Harrell, SP (2000). A multidimensional conceptualization of racism-related stress: Implications for the well-being of people of color. *American Journal of Orthopsychiatry*, 70(1), 42–57.
- John Hopkins University (2020). COVID-19 Map. Johns Hopkins Coronavirus Resource Center. Retrieved June 26, 2020, from <https://coronavirus.jhu.edu/map.html>
- Kaika, M (2017). ‘Don’t call me resilient again!’: the New Urban Agenda as immunology ... or ... what happens when communities refuse to be vaccinated with ‘smart cities’ and indicators. *Environment and Urbanization*, 29(1), 89–102.
- Kendi, I. X (6 April 2020). What the Racial Data Show. *The Atlantic*.
<https://www.theatlantic.com/ideas/archive/2020/04/coronavirus-exposing-our-racial-divides/609526/>
- Lamantia, J (9 July 2019). Life expectancy rises in NYC as disparities persist. *Crain’s New York Business*. <https://www.crainsnewyork.com/health-care/life-expectancy-rises-nyc-disparities-persist>
- Leclere, FB, Jensen, L, and Biddlecom, AE (1994). Health care utilization, family context, and adaptation among immigrants to the United States. *Journal of Health and Social Behavior*, 35(4), 370–384.
- Link, BG (2008). Epidemiological sociology and the social shaping of population health. *Journal of Health and Social Behavior*, 49(4), 367–384.
- Liu, K, Chen, Y, Lin, R, and Han, K (2020). Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *The Journal of Infection*, 80(6), e14–e18.
- Mays, JC, and Newman, A (8 April 2020). Virus Is Twice as Deadly for Black and Latino People Than Whites in N.Y.C. *The New York Times*.
<https://www.nytimes.com/2020/04/08/nyregion/coronavirus-race-deaths.html>
- Michaud, J, Kates, J, and Levitt, L (16 April 2020). COVID-19 Models: Can They Tell Us What We Want to Know? KFF. <https://www.kff.org/coronavirus-policy-watch/covid-19-models/>
- New York City Department of Health and Mental Hygiene (2018). Community Health Profiles - NYC Health. <https://www1.nyc.gov/site/doh/data/data-publications/profiles.page>
- New York City Department of Health and Mental Hygiene (2020). Disparity in NYC deaths between NYC DOHMH, NY State, and Johns Hopkins · Issue #27 · nychealth/coronavirus-data. GitHub. Retrieved June 22, 2020, from <https://github.com/nychealth/coronavirus-data/issues/27>
- New York State Department of Health (2020a). Novel Coronavirus (COVID-19). New York State Department of Health. Retrieved June 22, 2020, from <https://coronavirus.health.ny.gov/home>
- New York State Department of Health (2020b). New York State on PAUSE. Department of Health. Retrieved July 17, 2020, from <https://coronavirus.health.ny.gov/new-york-state-pause>
- New York State Department of Health (2020c). COVID-19 Testing. Department of Health. Retrieved June 22, 2020, from <https://coronavirus.health.ny.gov/covid-19-testing>
- NYC Health and Hospitals Corporation (2019). Community Health Needs Assessment (CHNA) and Implementation Strategy Plan.

- <https://www.nychealthandhospitals.org/publications-reports/2019-community-health-needs-assessment/>
- NYU Langone (2020). NYU Langone Scientists Track Evolution of Coronavirus in New York City. NYU Langone News. Retrieved June 22, 2020, from <https://nyulangone.org/news/nyu-langone-scientists-track-evolution-coronavirus-new-york-city>
- O'Brien, K, Eriksen, S, Nygaard, LP, and Schjolden, A (2007). Why different interpretations of vulnerability matter in climate change discourses. *Climate Policy*, 7(1), 73–88.
- O'Sullivan, T, and Bourgoin, M (2010). Vulnerability in an Influenza Pandemic: Looking Beyond Medical Risk | The Homeless Hub. <https://homelesshub.ca/resource/vulnerability-influenza-pandemic-looking-beyond-medical-risk>
- Parrot, J (2019). Economy: Inequality in New York City: The Intersection of Race and Class. In *Racial Inequality in New York City since 1965*. Retrieved June 22, 2020, from <https://www.sunypress.edu/p-6776-racial-inequality-in-new-york-c.aspx>
- Parrott, J, and Moe, L (2020). The New Strain of Inequality: The Economic Impact of Covid-19 In New York City. Center for New York City Affairs. Retrieved June 26, 2020, from <http://www.centernyc.org/reports-briefs/2020/4/15/the-new-strain-of-inequality-the-economic-impact-of-covid-19-in-new-york-city>
- Perry, A, Harshbarger, D, and Romer, C (2020, April 16). Mapping racial inequity amid COVID-19 underscores policy discriminations against Black Americans. Brookings. <https://www.brookings.edu/blog/the-avenue/2020/04/16/mapping-racial-inequity-amid-the-spread-of-covid-19/>
- Rocklöv, J, and Sjödin, H (2020). High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine*, 27(3).
- Rosenthal, BM (23 March 2020). Density Is New York City's Big 'Enemy' in the Coronavirus Fight. The New York Times. <https://www.nytimes.com/2020/03/23/nyregion/coronavirus-nyc-crowds-density.html>
- Rothstein, R (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing Corporation.
- Ruan, S. (2020). Likelihood of survival of coronavirus disease 2019. *The Lancet. Infectious Diseases*, 20(6), 630–631.
- Salmerón, R, García, CB, and García, J (2018). Variance Inflation Factor and Condition Number in multiple linear regression. *Journal of Statistical Computation and Simulation*, 88(12), 2365–2384.
- Schwartz, M, and Cook, LR (18 May 2020). These N.Y.C. Neighborhoods Have the Highest Rates of Virus Deaths. The New York Times. <https://www.nytimes.com/2020/05/18/nyregion/coronavirus-deaths-nyc.html>
- Snee, RD (1973). Some Aspects of Nonorthogonal Data Analysis. *Journal of Quality Technology*, 5(2), 67–79.
- Spearman, C (1906). 'Footrule' for Measuring Correlation. *British Journal of Psychology*, 1904-1920, 2(1), 89–108.
- U.S. Bureau of Labor Statistics (2018). Labor force characteristics by race and ethnicity, 2018 : BLS Reports: U.S. Bureau of Labor Statistics. <https://www.bls.gov/opub/reports/race-and-ethnicity/2018/home.htm>

- United States Census Bureau (2018). Using ACS Estimates and Margins of Error. Retrieved July 16, 2020, from <https://www.census.gov/programs-surveys/acs/guidance/training-presentations/acs-moe.html>
- Urban Systems Lab (15 May 2020). Covid-19 and housing precarity: from systemic failure towards a just recovery. Medium. <https://medium.com/urban-resilience/covid-19-and-housing-precarity-from-systemic-failure-towards-a-just-recovery-4083b48535a5>
- Wallace, R, and Wallace, D (1997). Resilience and Persistence of the Synergism of Plagues: Stochastic Resonance and the Ecology of Disease, Disorder and Disinvestment in US Urban Neighborhoods. *Environment and Planning: Economy and Space*, 29(5), 789–804.
- Warman, K, Silver, EJ, and Wood, P. R. (2009). Modifiable Risk Factors for Asthma Morbidity in Bronx Versus Other Inner-City Children. *The Journal of Asthma: Official Journal of the Association for the Care of Asthma*, 46(10), 995–1000.
- WE ACT for Environmental Justice (2017). Unequal Air and Care: Asthma Disparities. <https://www.weact.org/campaigns/asthmadisparities/>
- Wu, X, Nethery, RC, Sabath, BM, Braun, D, and Dominici, F (2020). Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *MedRxiv*, 2020.04.05.20054502.