



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



COVID-19 testing, case, and death rates and spatial socio-demographics in New York City: An ecological analysis as of June 2020

Byoungjun Kim^{a,*}, Andrew G. Rundle^a, Alicia T. Singham Goodwin^a,
Christopher N. Morrison^{a,c}, Charles C. Branas^a, Wafaa El-Sadr^{a,b}, Dustin T. Duncan^a

^a Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, United States

^b ICAP, Columbia University Mailman School of Public Health, New York, NY, United States

^c Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

ARTICLE INFO

Keywords:

COVID-19
Spatial epidemiology
Spatial demography
Spatial analysis
Neighborhood

ABSTRACT

We assessed the geographic variation in socio-demographics, mobility, and built environmental factors in relation to COVID-19 testing, case, and death rates in New York City (NYC). COVID-19 rates (as of June 10, 2020), relevant socio-demographic information, and built environment characteristics were aggregated by ZIP Code Tabulation Area (ZCTA). Spatially adjusted multivariable regression models were fitted to account for spatial autocorrelation. The results show that different sets of neighborhood characteristics were independently associated with COVID-19 testing, case, and death rates. For example, the proportions of Blacks and Hispanics in a ZCTA were positively associated with COVID-19 case rate. Contrary to the conventional hypothesis, neighborhoods with low-density housing experienced higher COVID-19 case rates. In addition, demographic changes (e.g. out-migration) during the pandemic may bias the estimates of COVID-19 rates. Future research should further investigate these neighborhood-level factors and their interactions over time to better understand the mechanisms by which they affect COVID-19.

1. Introduction

After the first detected COVID-19 case in New York City (NYC) on March 1, 2020, the city rapidly became the first epicenter of disease in the United States. As of April 2020, at its peak the NYC Department of Health and Mental Hygiene (NYCDOHMH) reported 15 days in that period with over 6000 new confirmed cases of COVID-19 and over 500 deaths (New York City Department of Health and Mental Hygiene, 2020a). There is emerging evidence that marginalized (i.e. low-income) and vulnerable (i.e. racial and ethnic minority) populations are disproportionately affected by COVID-19 (Webb Hooper et al., 2020; Alsan et al., 2020). Black and Hispanic/Latinx people in the U.S. experienced higher case rates and death rates compared to Whites, and there was a clear socio-demographic gradient in COVID-19 infection by income and poverty (New York University Furman Center, 2020; Price-Haywood et al., 2020; Raifman and Raifman, 2020). Occupational characteristics were associated with risk for the disease as well as secondary transmissions; e.g., workers in the healthcare sector and other essential

service occupations were at higher risk for infection, due to frequent interactions with possibly infected individuals, and being in close quarters for extended periods of time with other workers (Baker et al., 2020). In addition to the individual-level characteristics, environmental factors in urban contexts such as urban design, housing density, and transportation systems can impact the transmission of infectious diseases (Harlem, 2020). Highly populated neighborhoods and multi-family housing structures tend to increase person-to-person contacts, which in turn can exacerbate community transmission (Rocklov and Sjodin, 2020; Ghinai et al., 2020). The frequent use of public transportation systems has also been noted as a potential risk factor of COVID-19 in urban areas (Zheng et al., 2020).

Recent analytic studies looking at associations between neighborhood characteristics and the geographic distribution of COVID-19 have focused on socio-demographic factors, but have not explored the contribution of environmental factors. In addition, as SARS-CoV-2 testing and COVID-19 case/death rates show geographical clustering with high spatial autocorrelation (Kang et al., 2020), it is critical to

* Corresponding author. Columbia University Mailman School of Public Health, Department of Epidemiology, Columbia Spatial Epidemiology Lab, 722 West 168th Street, New York, NY, 10032, United States.

E-mail address: bk2767@cumc.columbia.edu (B. Kim).

<https://doi.org/10.1016/j.healthplace.2021.102539>

Received 4 September 2020; Received in revised form 11 February 2021; Accepted 12 February 2021

Available online 19 February 2021

1353-8292/© 2021 Elsevier Ltd. All rights reserved.

properly adjust for such spatial similarity in the modeling stage. Lastly, with the advent of COVID-19, the socio-demographic landscapes in NYC changed due to residents moving out of NYC in response to the pandemic. On average, 5% of NYC residents left the city between March 1 and May 1, but the proportion of residents who left the city was socially patterned and varied substantially across neighborhoods (Quealy, 2020). The COVID-19 rates provided by NYCDOHMH are based on estimated populations from census data before the pandemic; thus, the SARS-CoV-2 testing and COVID-19 case/death rates are likely to be biased estimates. To the best of our knowledge, none of the recent studies on socio-demographic predictors of COVID-19 cases, testing, or hospitalizations have incorporated data on NYC's population changes across neighborhoods, and few have accounted for spatial dependencies that could bias estimates. In this study, we examine the geographic variation in socio-demographic characteristics, migration patterns, mobility, and built environmental factors in NYC in relation to COVID-19 rates, using spatial analytic methods to address potential issues of spatial autocorrelation.

2. Methods

2.1. Data

COVID-19 statistics for testing, positive cases, and death counts of New York City residents by residential ZCTA (ZIP Code Tabulation Area) were obtained from the New York City Department of Health and Mental Hygiene on June 10, 2020 (New York City Department of Health and Mental Hygiene, 2020c). The NYCDOHMH reports data using *modified* ZCTAs, with certain boundaries modified to combine areas with small or no populations to allow for stable estimates of COVID-19 rates. After combining 34 such ZCTAs, there were a total of 177 *modified* ZCTAs (referred to simply as ZCTAs from now on) with valid COVID-19 data included in this analysis. The three outcomes – number of SARS-CoV-2 tests, COVID-19 cases, and deaths – were normalized by the population of their ZCTAs and used as outcomes. The total population by ZCTA was obtained from the 2018 American Community Survey 5-year estimates. As our method for spatial regression analysis can only account for areas with physically-touching neighbors, one ZCTA with no neighboring areas (10044: Roosevelt Island) was excluded from the analysis. Another island in NYC, Rikers Island, is included in the analysis as it is incorporated into ZCTA 11370, which contains land in Queens (Astoria Heights) that shares boundaries with neighboring ZCTAs.

Pre-pandemic socio-demographic characteristics were calculated by ZCTA from the 2018 American Community Survey 5-year estimates, including age, sex, race/ethnicity, median income, household size, occupation, and commuting characteristics (U.S. Census Bureau, 2020). Specifically, variables included in the present analysis as predictors of COVID-19 outcomes were: male-to-female ratio (number of male/number of female); percentages of the population under 18 (used as reference), 18–44, 45–64, 65–74, and over 75 years; percentages of non-Hispanic White (used as reference), Black, Asian, others, and Hispanic populations; median household income; average household size; percentages involved in essential service occupations (firefighting, law enforcement, building and ground cleaning and maintenances, food preparation and serving related, and personal care) and health-related occupations (healthcare practitioners and technical occupations, healthcare support services); and percentage commuting via public transit. The socio-demographic variables were estimates from 2014 to 2018 surveys, and therefore the data may not fully reflect characteristics during the pandemic. Socio-demographic variables expressed as proportions were re-scaled such that a 1-unit change reflected the inter-quartile range (IQR), that is, the difference between the 75th and 25th percentile ZCTA.

A zoning map for NYC was obtained from the NYC Department of City Planning (New York City Department of City Planning, 2020), and each ZCTA's percentages of land assigned to low and high-density

residential zonings were calculated using Quantum GIS v3.10. (QGIS Development Team, Open Source Geospatial Foundation Project). Residential zones R1-R5 are classified as low-density zones, predominantly consisting of detached or semi-detached single- and two-family housing. Zones R6-R10 are classified as higher-density residential zones, and allow for high-rise multifamily housing (New York City Department of City Planning, 2018).

Lastly, in order to account for residents moving out of the city during the COVID-19 pandemic, a dataset from cellular phone usage was utilized as a proxy measure of population changes. The data from cellular phone towers captures the mobility and migration patterns of a wide range of residents, as the towers interact with all types of cellular devices even when those devices are in stand-by and calls are not in progress. More than 1 million cellular devices that interacted with cellular towers in NYC were analyzed, and the percentage change in registered cellular phone signals between March 1, 2020 and May 1, 2020 were aggregated into census tracts by Teralytics Inc. (New York, NY). These data were provided to our research team by *The New York Times* (Quealy, 2020). The census-tract-level data were converted to ZCTAs with the help of Crosswalk Files provided by the U.S. Department of Housing and Urban Development. For ZCTAs containing area from multiple census tracts, we calculated a weighted average of the component census tracts' population change rates. Weights were determined by calculating the proportion of residential addresses in a ZCTA contained within each given census tract (U.S. Department of Housing and Urban Development). The resulting ZCTA-level percent decreases in residential population was conceptualized as an "out-migration" index, and this metric was included in our analyses as a covariate.

2.2. Statistical analysis

Spatial autocorrelation indicates geographical interdependencies among observations in data. When spatial autocorrelation is detected, the major assumptions of uncorrelated error terms and independence of observations are violated. This can lead to biased parameter estimates, necessitating adjustment for spatial clustering (LeSage and Pace, 2009; Ward and Gleditsch, 2019). Therefore, we first tested for spatial autocorrelations of all variables using the Global Moran's *I*. In this study, a row-standardized binary contiguity spatial weight matrix with first-order queen's criteria was employed, which is a conventional spatial matrix for areal data (Haining, 1991). A pseudo p-value of the Global Moran's *I* for each variable was estimated from a Monte Carlo simulation of 999 random iterations. Second, spatially adjusted Spearman correlations were tested to evaluate bivariate correlations between study variables and COVID-19 rates based on a spatial adjustment method proposed by Clifford and Richardson (Clifford and Richardson, 1985; Duncan et al., 2011).

After evaluating spatial autocorrelation for each variable of interest, we tested regression models for each exposure variable after adjusting for both the spatial autocorrelation and the out-migration index. This set of regression models provides unbiased crude associations between neighborhood-level factors and COVID-19 rates, adjusting for the potential confounding due to out-migration. Lastly, a set of spatially adjusted multivariable models with all neighborhood characteristic variables were specified. Because the neighborhood-level socio-demographic and built-environmental factors are interconnected (Leal et al., 2012), we tested variance inflation factors (VIF) for each variable to check the degree of multicollinearity (Song et al., 2017). We employed a cut-off point of 10, considering the sample size of the analysis and underlying correlations between socio-demographic characteristics (Craney and Surles, 2002).

In the model with COVID-19 case rate as the outcome, we included testing rate as a covariate because the case rate is associated with the number of tests conducted in a ZCTA. Likewise, the case rate was included as a covariate in the model with COVID-19 death rate as the outcome. The multivariable models were adjusted for spatial autocor-

relation contingent on Lagrange Multiplier (LM) test results. The LM tests evaluated spatial error and lag dependences from non-spatial ordinary least squares (OLS) models (LeSage and Pace, 2009). Based on the test results, spatial error models or spatial lag models were developed to account for spatial autocorrelation (LeSage and Pace, 2009; Ward and Gleditsch, 2019). The spatial error model equation can be represented as follows:

$$y = \beta X + \lambda W\epsilon + \mu$$

Where y is the dependent variable; β is the vector of the regression parameter associated with the matrix of observations on the covariates X ; λ is the spatial autoregressive coefficient that indicates the extent to which the spatial component of the errors is correlated with each other; W is the given spatial weight matrix; and μ is an independent error term.

The spatial lag model considered is given by:

$$y = \rho W y + \beta X + \epsilon$$

where y is the dependent variable; ρ is the spatial autoregressive coefficient for the lagged dependent variable matrix Wy (W is the given spatial weight matrix); β is the vector of coefficients of regression parameters associated with the independent variable matrix X ; and ϵ is an error term that is assumed to be independent and identically distributed. In the spatial lag model, therefore, spatial autocorrelation is introduced in the form of the spatially dependent variable, as the outcomes in one place predict an increased likelihood of similar outcomes in neighboring places.

The Akaike Information Criterion was examined for goodness-of-fit (Bozdogan, 1987). All statistical analyses were conducted in R statistical software version.

3. Results

Maps showing COVID-19 testing, case and death rates and selected

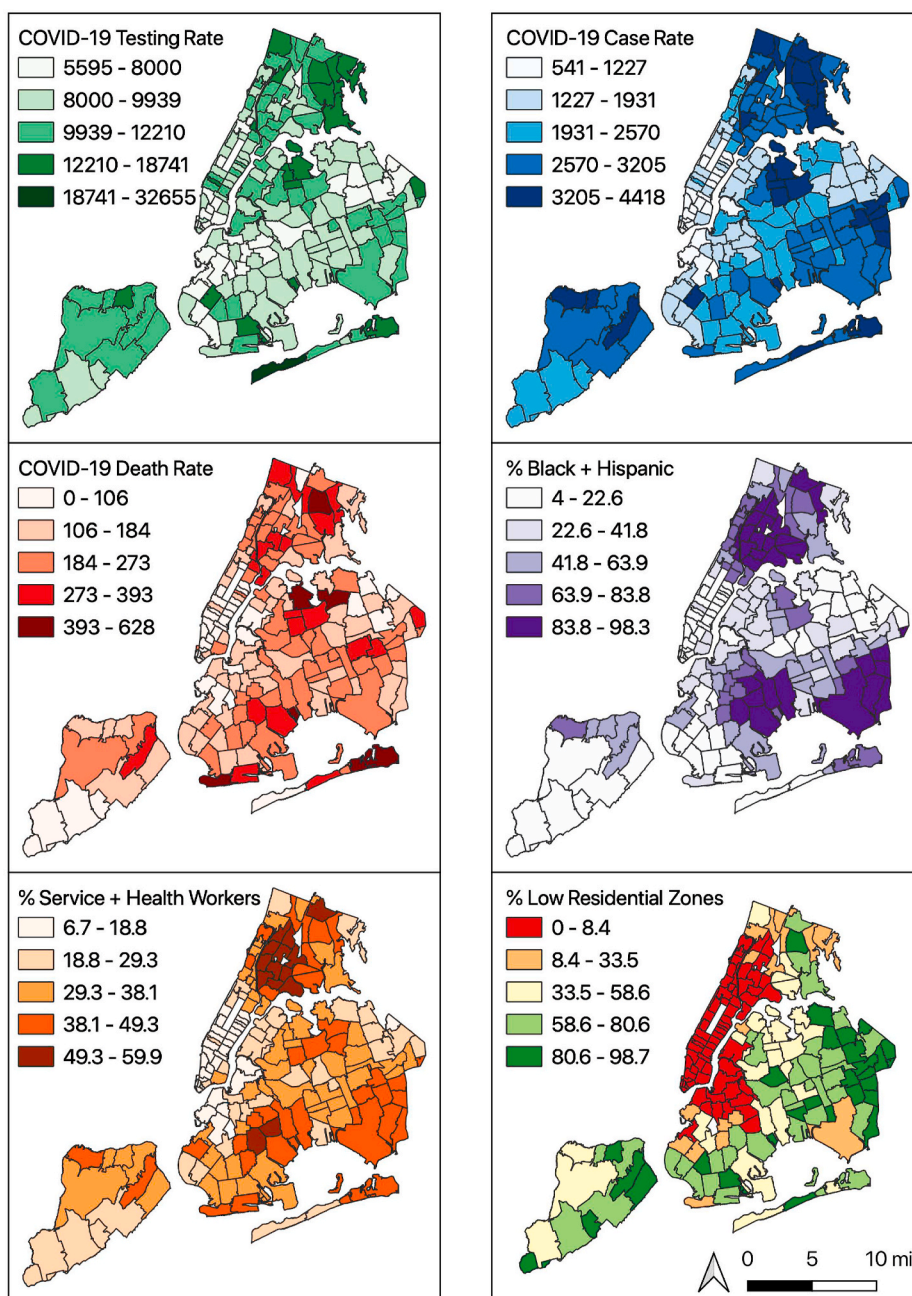


Fig. 1. COVID-19 testing, case, death rates (/100,000) and socio-demographics in New York city by ZCTA

socio-demographic characteristics by ZCTA are provided in Fig. 1. Table 1 provides descriptive statistics and the spatial autocorrelation statistic (Moran's *I*) for each study variable. The city-wide SARS-CoV-2 testing rate was 9569.9 per 100,000 population (standard deviation across ZCTAs (SD): 2641.5), and the COVID-19 case and death rates were 2286.8 (SD: 878.0) and 188.7 (SD: 102.0) per 100,000 respectively. The Moran's *I* statistics show strong spatial autocorrelation of all variables of interest, indicating the violation of independence of observations. For example, the COVID-19 case rate has high spatial homogeneity (Moran's *I* = 0.75, p-value < 0.01).

The spatially adjusted Spearman test illustrates the rank-order correlations between COVID-19 rates and variables of interest (Table 2). For example, a high percentage of the population in health-related occupations is highly correlated with SARS-CoV-2 testing ($\rho = 0.41$), case rate ($\rho = 0.62$), and death rate ($\rho = 0.50$). The out-migration index was associated with lower testing ($\rho = -0.29$), case ($\rho = -0.57$) and death ($\rho = -0.40$) rates, meaning that ZCTAs that lost more of their population during the pandemic had lower rates for all three outcomes assessed.

Bivariate and multivariable model estimates for each outcome are provided in Table 3. The bivariate analyses indicated predictors of each COVID-19 rate after adjusting for spatial autocorrelation and out-migration. The Moran's *I* values for the non-spatial multivariable OLS regression residuals showed evidence of spatial autocorrelation in each model, confirming the need for a spatial error model or a spatial lag model. The LM tests of OLS models indicate that spatial lag models would be better-fitting models for SARS-CoV-2 testing and case rates, whereas a spatial error model could be applied for COVID-19 death rate (Florax et al., 2003). Lastly, the VIF tests for each multivariable OLS model detected high multicollinearity of one independent variable: the

Table 1
Descriptive statistics and global spatial autocorrelation (N = 176, as of October 6, 2020).

| | Mean (SD) | Range | Moran's <i>I</i> | P-value ^b |
|----------------------------------|-----------------|--------------|------------------|----------------------|
| SARS-CoV-2 testing (/100,000) | 9569.6 (2641.5) | 5136–32,347 | 0.33 | 0.001 |
| COVID-19 cases (/100,000) | 2286.8 (878.0) | 540.6–4418.3 | 0.75 | 0.001 |
| COVID-19 percent positive (%) | 22.9 (6.4) | 9.0–33.7 | 0.77 | 0.001 |
| COVID-19 deaths (/100,000) | 188.7 (102.0) | 0.0–628.4 | 0.41 | 0.001 |
| Male to female ratio (%) | 92.2 (10.1) | 69.0–145.8 | 0.30 | 0.001 |
| % under 18 years | 20.9 (5.5) | 6.3–36.2 | 0.67 | 0.001 |
| % 18–44 years | 40.9 (8.3) | 23.4–74.4 | 0.67 | 0.001 |
| % 45–64 years | 24.9 (3.6) | 11.8–34.0 | 0.55 | 0.001 |
| % 65–74 years | 8.0 (2.5) | 0.5–14.4 | 0.46 | 0.001 |
| % over 75 years | 6.3 (2.8) | 0.0–14.7 | 0.42 | 0.001 |
| % non-Hispanic Black | 19.8 (23.9) | 0.4–90.5 | 0.73 | 0.001 |
| % non-Hispanic Asian | 14.6 (14.0) | 0.1–72.6 | 0.67 | 0.001 |
| % non-Hispanic others | 3.2 (3.1) | 0.4–28.9 | 0.45 | 0.001 |
| % Hispanic | 26.1 (19.5) | 1.1–75.8 | 0.76 | 0.001 |
| Median household income | 73.5K (36.6K) | 21.1K–250K | 0.71 | 0.001 |
| Average household size | 2.6 (0.5) | 1.6–4.0 | 0.71 | 0.001 |
| % service jobs | 16.0 (7.2) | 0.8–33.3 | 0.72 | 0.001 |
| % health-related jobs | 10.8 (4.2) | 2.6–21.8 | 0.73 | 0.001 |
| % pre-pandemic transit commuters | 53.1 (15.0) | 12.3–85.3 | 0.77 | 0.001 |
| % low density housing zones | 38.8 (35.9) | 0.0–98.7 | 0.77 | 0.001 |
| % high density housing zones | 29.4 (30.0) | 0.0–96.2 | 0.71 | 0.001 |
| Out-migration index ^a | 7.5 (8.5) | 0.0–37.2 | 0.80 | 0.001 |

^a % changes in unique cellular phone signals between March 1, 2020 and May 1, 2020.

^b p-value for global Moran's *I*.

Boldface indicates statistical significance.

percentage of people in the ZCTA working service jobs (VIF > 10). We therefore fitted spatial lag and error models without that variable, and the results remained stable in terms of the size and confidence interval of each coefficient (data not shown).

3.1. SARS-CoV-2 test rate

From the multivariable spatial lag model (Table 3), percentages of people aged over 75 years ($\beta = 1232.6$, CI: [515.9, 1949.2]), non-Hispanic people of "other" races (American Indian, Alaska Native, Pacific Islander, some other race, and two or more races: $\beta = 230.9$, CI: [22.2, 439.6]), median household income ($\beta = 2205.4$, CI: [375.6, 4035.2]), workers in essential service occupations ($\beta = 2237.5$, CI: [993.9, 3481.2]), and health-related occupations ($\beta = 896.9$, CI: [92.8, 1701.0]) were positively associated with SARS-CoV-2 testing rate. The out-migration index score was not associated with ZCTA-level test rates. Percentages of people aged 45–64 and 65–74 ($\beta = -1054.2$, CI: [1799.3, -309.2]; $\beta = -891.5$, CI: [1742.5, -40.5]), non-Hispanic Asians ($\beta = -948.0$, CI: [-1589.9, -306.1]), average household size ($\beta = -1802.1$, CI: [-3402.1, -202.1]), commuters using transit ($\beta = -1239.8$, CI: [-2109.5, -370.2]), and average household size ($\beta = -1802.1$, CI: [-3402.1, -202.1]) in ZCTAs show negative relationships with the SARS-CoV-2 testing rate. For example, for percent of the population employed in the service industry, comparing the 75th to the 25th percentile ZCTA (the interquartile range) shows a +2238 per 100,000-person difference in the testing rate, after adjusting for covariates.

3.2. COVID-19 case rate

Share of male population compared to female ($\beta = 75.8$, CI: [5.1, 146.6]), percentages of non-Hispanic Blacks ($\beta = 248.7$, CI: [136.0, 361.4]), Hispanics ($\beta = 368.9$, CI: [211.2, 526.6]), and low-density housing zone areas ($\beta = 5.6$, CI: [2.6, 8.7]) were positive predictors of ZCTA-level COVID-19 positive case rate (Table 3). The out-migration index score was positively associated with ZCTA-level case rates ($\beta = 11.5$, CI: [0.2, 22.8]).

3.3. COVID-19 death rate

Percentages of people aged 18–44 and 45–64 ($\beta = -37.3$, CI: [-67.8, -6.7]; $\beta = -22.9$, CI: [-44.9, -0.8]) and average household size ($\beta = -54.0$, CI: [-102.2, -5.8]) were negatively associated with COVID-19 death rate. Percentages of non-Hispanic Asians ($\beta = 22.7$, CI: [3.1, 42.4]) and commuters using transit ($\beta = 48.3$, CI: [20.9, 75.7]) were positive predictors of COVID-19 death rate. The out-migration index score was not associated with ZCTA-level COVID-19 death rates.

4. Discussion

The aim of this study was to examine social and environmental determinants of COVID-19 in NYC. To the best of our knowledge, this is the first study to address the potential bias due to residents moving out of the city at differential rates across neighborhoods. News-media reported that out-migration varied strongly by neighborhood socio-demographic characteristics, and our analysis found that neighborhood level out-migration, as measured by cell phone derived data, was also associated with COVID-19 case rates. As reported in the Supplemental table (Table S1), there appeared to be modest confounding by out-migration; the sizes of coefficients for parameter estimates differed between models that did and did not adjust for the out-migration. After adjustment for out-migration, we identified potential drivers of SARS-CoV-2 test, case, and death rates, including a ZCTA's composition by sex, age, race/ethnicity, income, and occupational risks. Specifically, our results showed that higher percentages of residents who were Black and Hispanic were positively associated with COVID-19 case rates. We also confirmed that neighborhoods with higher percentages of essential

Table 2
Spearman correlation between ZCTA level socio-demographics and COVID-19 rates (N = 176).

| | SARS-CoV-2 Test Rate | | | COVID-19 Case Rate | | | COVID-19 Death Rate | | |
|----------------------------------|----------------------|-----------------|---|--------------------|-----------------|---|---------------------|-----------------|---|
| | r _s | P-value | Spatially Adjusted P-value ^b | r _s | P-value | Spatially adjusted P-value ^b | r _s | P-value | Spatially adjusted P-value ^b |
| Male to female ratio (%) | 0.01 | 0.92 | 0.95 | -0.15 | 0.05 | 0.13 | 0.04 | 0.63 | 0.71 |
| % under 18 years | 0.23 | <0.01 | 0.10 | 0.58 | <0.01 | <0.01 | 0.42 | <0.01 | <0.01 |
| % 18–44 years | -0.38 | <0.01 | <0.01 | -0.52 | <0.01 | <0.01 | -0.35 | <0.01 | <0.01 |
| % 45–64 years | 0.13 | 0.08 | 0.29 | 0.30 | <0.01 | 0.07 | 0.08 | 0.28 | 0.53 |
| % 65–74 years | 0.17 | 0.03 | 0.12 | 0.07 | 0.37 | 0.61 | 0.02 | 0.72 | 0.80 |
| % over 75 years | 0.22 | <0.01 | 0.03 | 0.09 | 0.26 | 0.48 | 0.12 | 0.10 | 0.24 |
| % non-Hispanic White | -0.18 | 0.02 | 0.17 | -0.55 | <0.01 | <0.01 | -0.57 | <0.01 | <0.01 |
| % non-Hispanic Black | 0.24 | <0.01 | 0.04 | 0.45 | <0.01 | <0.01 | 0.47 | <0.01 | <0.01 |
| % non-Hispanic Asian | -0.28 | <0.01 | 0.03 | -0.31 | <0.01 | 0.06 | -0.26 | <0.01 | 0.05 |
| % non-Hispanic others | -0.26 | <0.01 | 0.01 | -0.25 | <0.01 | 0.05 | -0.13 | 0.09 | 0.23 |
| % Hispanic | 0.28 | <0.01 | 0.03 | 0.45 | <0.01 | <0.01 | 0.47 | <0.01 | <0.01 |
| Median household income | -0.24 | <0.01 | 0.08 | -0.49 | <0.01 | <0.01 | -0.65 | <0.01 | <0.01 |
| Average household size | 0.22 | <0.01 | 0.12 | 0.62 | <0.01 | <0.01 | 0.39 | <0.01 | <0.01 |
| % service jobs | 0.28 | <0.01 | 0.04 | 0.59 | <0.01 | <0.01 | 0.61 | <0.01 | <0.01 |
| % health-related jobs | 0.41 | <0.01 | <0.01 | 0.62 | <0.01 | <0.01 | 0.50 | <0.01 | <0.01 |
| % pre-pandemic transit commuters | -0.21 | <0.01 | 0.13 | -0.21 | <0.01 | 0.28 | 0.14 | 0.06 | 0.32 |
| % low density zones | 0.24 | <0.01 | 0.10 | 0.50 | <0.01 | <0.01 | 0.24 | <0.01 | 0.11 |
| % high density zones | -0.12 | 0.10 | 0.37 | -0.21 | <0.01 | 0.25 | 0.16 | 0.04 | 0.26 |
| Out-migration index ^a | -0.29 | <0.01 | <0.07 | -0.57 | <0.01 | <0.01 | -0.40 | <0.01 | <0.01 |
| Testing rate | - | - | - | 0.78 | <0.01 | <0.01 | 0.51 | <0.01 | <0.01 |
| Case rate | - | - | - | - | - | - | 0.71 | <0.01 | <0.01 |

^a % changes in unique cellular phone signals between March 1, 2020 and May 1, 2020.

^b P-value of Spearman correlation test accounting for spatial autocorrelation.

Boldface indicates statistical significance.

service and healthcare-related workers had increased SARS-CoV-2 testing and death rates. These findings are consistent with recent studies on the social determinants of COVID-19 (New York City Department of Health and Mental Hygiene, 2020a; Webb Hooper et al., 2020; New York University Furman Center, 2020; Baker et al., 2020).

Contrary to the existing findings (Harris, 2020), higher pre-pandemic transit ridership in a neighborhood was not associated with higher COVID-19 case or death rates. Subway use has decreased dramatically after the declaration of a local state of emergency (Sy et al., 2020), and this null association may reflect the reduced overall ridership during the pandemic. Another possible explanation could be increased vigilance in transit riders – and subsequent higher levels of health-promoting behaviors such as physical distancing and mask-wearing – due to public awareness of the risk for infection in enclosed spaces. Additionally, the NYC Metropolitan Transportation Authority (MTA) worked to mitigate COVID-19 transmission risk by regularly disinfecting all subways and buses, installing hand sanitizers in all stations, and marking six feet of distance on subway platforms (Goldbaum 2020) (New York City Office of the Mayor, 2020).

Our results indicated that neighborhoods zoned for predominantly low-density housing had higher COVID-19 case rates than those zoned for predominantly high-density housing. While this is not consistent with the conventional hypothesis (Rocklov and Sjodin, 2020), there are inconsistent findings regarding the association between population density and COVID-19 rates. Recent studies on neighborhood-level factors in Chinese and European cities reported that population density was negatively or not associated with COVID-19 (Liu, 2020; Gerli et al., 2020). One explanation could be the differential application of mitigation strategies. Early in the pandemic the Center for Disease Control and Prevention and the NYCDOHMH announced guidelines for maintaining safe operations of multifamily housing, including closing public areas in the building, disinfecting common areas, providing hand sanitizer in common areas, and mandatory mask-wearing (Centers for Disease Control and Prevention, 2020; New York City Department of Health and Mental Hygiene, 2020b). A myriad of high-rise residential buildings in NYC voluntarily implemented such recommendations (Amy Plitt, 2020), and such vigilance may be associated with the relatively low COVID-19 case rates in neighborhoods zoned for high-density

residential buildings. Visual inspection of the map (Fig. 1) shows that areas with the highest percentage of low-density housing are located on the periphery of the city. It is possible that such a geographic location could lead to lower access to healthcare and social support, and subsequent higher COVID-19 risk despite any protections conferred by living in low-density housing (Ji et al., 2020).

There are limitations in this analysis. First, the city’s COVID-19 testing and case data are not necessarily representative of the underlying populations at risk or experiencing COVID-19 infection in their ZCTA. Of note, until early May the NYC DOHMH discouraged people with mild and moderate symptoms from being tested due to limited testing resources. Therefore, these reported rates are subject to potential selection/sampling bias as well as misclassification. To illustrate, NYC overall is 52% female and 48% male (U.S. Census Bureau, 2020), but the testing breakdown by sex was 56% female and 44% male. Because COVID-19 testing and case data did not come from a randomly sampled or representative population, our analyses may be confounded by the skewedness of the underlying data. However, the COVID-19 testing, case, and death rates from the DOHMH were the best and only available data to estimate population-level COVID-19 in NYC. Second, this analysis is susceptible to many common problems in neighborhood-level analyses. It is susceptible to the ecological fallacy: that the associations found in aggregated data may not translate to corresponding associations at the individual level. Additionally, in ecological studies measurement errors in the predictor variables can bias results away from the null (Brenner et al., 1992). Our unit of analysis, ZCTA, and similar geographic boundaries can be subject to the modifiable area unit problem (MAUP) – potential bias due to the artificial aggregation of point-based data (Wong, 2009). Third, there is a temporal mismatch between the COVID-19 statistics and the socio-demographic data used in this analysis. COVID-19 data were retrieved on June 10, 2020, whereas the American Community Survey data are estimates from a 5-year survey conducted between 2014 and 2018. Fourth, the analysis is also susceptible to potential residual confounding, such as by the uneven distribution of underlying health conditions by neighborhood. For example, chronic respiratory and cardiovascular diseases can increase the risk for COVID-19 death (Jordan et al., 2020), and may also be associated with neighborhood conditions, but such measures were not

Table 3
Crude and multivariable estimations of the relationships (N = 176).

| | Testing rate (Spatial Lag Model) | Case rate (Spatial Lag Model) | Death rate (Spatial Error Model) |
|--------------------------------------|----------------------------------|--------------------------------|----------------------------------|
| Crude Models^a | | | |
| Male to female ratio (%) | 38.1 (-353.5, 429.7) | 45.9 (-30.3, 122.0) | -4.9 (-33.4, -3.1) |
| % 18-44 years | -780.9 (-1295.9, -265.8) | -123.9 (-225.6, -22.1) | -35.3 (-56.9, -13.7) |
| % 45-64 years | 5.9 (-461.8, 473.5) | -31.2 (-122.3, 59.9) | -9.8 (-29.6, 10) |
| % 65-74 years | 569.6 (110.5, 1028.7) | 7.4 (-82.5, 97.3) | 19.5 (0.1, 38.8) |
| % over 75 years | 806 (409.2, 1202.9) | 62.5 (-16.1, 141.0) | 36 (19.7, 52.2) |
| % non-Hispanic Black | -48.4 (-441.6, 344.8) | 93 (15.2, 170.7) | 27.6 (9.9, 45.4) |
| % non-Hispanic Asian | -561.9 (-1019.5, -104.3) | -92.3 (-182.1, -2.4) | -7.7 (-28.3, 12.8) |
| % non-Hispanic others | -147.7 (-342.4, 47) | -6.1 (-44.2, 31.9) | -1.7 (-9.9, 6.6) |
| % Hispanic | 86.2 (-408.6, 581) | 160.5 (59.8, 261.3) | 29.2 (6.5, 51.9) |
| Median household income ^a | 56.7 (-899.5, 1012.9) | -311.3 (-500.5, -122.1) | -117.6 (-157.4, -77.9) |
| Average household size | -293.2 (-1183.7, 597.4) | 299.7 (121.7, 477.7) | -4.5 (-41.6, 32.7) |
| % service jobs | 167.4 (-417.2, 752) | 241.4 (124.4, 358.4) | 54.9 (31.9, 77.9) |
| % health-related jobs | 447.1 (-126.2, 1020.4) | 178.2 (62.9, 293.4) | 42.8 (18.5, 67.1) |
| % pre-pandemic transit commuters | -831.5 (-1387.1, -275.9) | -56.6 (-162.8, 49.7) | 24.5 (-1.3, 50.3) |
| % low density housing zones | 2.8 (-8.8, 14.3) | 1.4 (-0.9, 3.7) | 0 (-0.5, 0.5) |
| % high density housing zones | -6.6 (-18.3, 5) | -0.9 (-3.1, 1.4) | 0.4 (-0.1, 1.0) |
| Testing rate | - | 0.1 (0.08, 0.13) | - |
| Case rate | - | - | 0.09 (0.07, 0.11) |
| Multivariable Models | | | |
| Male to female ratio (%) | 441.2 (-12.6, 895.0) | 75.8 (5.1, 146.6) | -5.2 (-18.4, 8.0) |
| % 18-44 years | -842.1 (-1851, 166.8) | -104.1 (-260.4, 52.2) | -37.3 (-67.8, -6.7) |
| % 45-64 years | -1054.2 (-1799.3, -309.2) | -79.1 (-196.2, 38.0) | -22.9 (-44.9, -0.8) |
| % 65-74 years | -891.5 (-1742.5, -40.5) | 9.8 (-122.6, 142.2) | -6.5 (-30.9, 17.9) |
| % over 75 years | 1232.6 (515.9, 1949.2) | 61.8 (-51.6, 175.3) | 19.6 (-1.6, 40.8) |
| % non-Hispanic Black | -353.3 (-1081.6, 375.1) | 248.7 (136.0, 361.4) | 13.7 (-9.4, 36.8) |
| % non-Hispanic Asian | -948.0 (-1589.9, -306.1) | 29.9 (-70.9, 130.6) | 22.7 (3.1, 42.4) |
| % non-Hispanic others | 230.9 (22.2, 439.6) | 14.4 (-18.3, 47.2) | 3.9 (-2.3, 10.2) |
| % Hispanic | -189.1 (-1167.6, 789.4) | 368.9 (211.2, 526.6) | -2.7 (-34.3, 28.8) |
| Median household income ^b | 2205.4 (375.6, 4035.2) | -263.3 (-549.4, 22.7) | -17.0 (-71.1, 37.0) |
| Average household size | -1802.1 (-3402.1, -202.1) | 132.7 (-116.9, 382.3) | -54.0 (-102.2, -5.8) |
| % service jobs | 2237.5 (993.9, 3481.2) | -47 (-245.5, 151.6) | 18.3 (-17.8, 54.5) |
| % health-related jobs | 896.9 (92.8, 1701.0) | -16.1 (-142.8, 110.5) | -4.1 (-28.2, 19.9) |
| % pre-pandemic transit commuters | -1239.8 (-2109.5, -370.2) | -29.8 (-164.7, 105.2) | 48.3 (20.9, 75.7) |
| % low density housing zones | -5.1 (-24.6, 14.3) | 5.6 (2.6, 8.7) | 0.2 (-0.4, 0.7) |
| % high density housing zones | -5.8 (-24.7, 13) | -1.7 (-4.6, 1.2) | -0.1 (-0.7, 0.4) |
| Out-migration index ^c | -29.9 (-102.2, 42.4) | 11.5 (0.2, 22.8) | 1.1 (-1.0, 3.3) |
| Testing rate | - | 0.11 (0.09, 0.13) | - |
| Case rate | - | - | 0.09 (0.07, 0.11) |
| Non-Spatial OLS Moran's I (p-value) | 0.13 (<0.01) | 0.15 (<0.01) | 0.13 (<0.01) |
| Spatial Model Moran's I (p-value) | -0.02 (0.62) | -0.03 (0.69) | 0.00 (0.49) |
| LM Spatial Error Model (p-value) | 5.88 (0.02) | 7.48 (<0.01) | 5.41 (0.02) |
| LM Spatial Lag Model (p-value) | 7.74 (<0.01) | 27.34 (<0.01) | 1.24 (0.27) |
| Non-Spatial OLS AIC | 3218.5 | 2584.3 | 1991.1 |
| Spatial Model AIC | 3212.5 | 2558.2 | 1986.9 |

Boldface indicates statistical significance.

^a Adjusted for Out-migration index.

^b Annual income in \$1000.

^c % changes in unique cellular phone signals between March 1, 2020 and May 1, 2020.

incorporated in this analysis due to lack of available data. Lastly, the cellular phone data-based estimates of out-migration provided by *The New York Times* may not fully capture people's movement in and out of the city, since phone usage may systematically differ from actual mobility of the residents in corresponding ZCTA.

5. Conclusions

This study provides important information on neighborhood-level factors and their association with COVID-19, in the context of a large metropolitan city with a high burden of COVID-19 in the United States. In addition to socio-demographic characteristics like neighborhood-level distributions of sex, age, and race/ethnicity, we must also focus on the impacts of the built environment on COVID-19 transmission and mortality. Future research should emphasize interactions between health behaviors (i.e. social distancing and commuting behaviors) and built environments in order to shed light on the environmental determinants of COVID-19.

Funding

This research did not receive any specific grant or funding from agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors gratefully acknowledge Kevin Quealy at the New York Times for providing the cellular phone usage data that supported this analysis. The authors also thank the New York City Department of Health and Mental Hygiene for the rapid and open provision of COVID-19 data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.healthplace.2021.102539>.

References

- Alsan, M., Stantcheva, S., Yang, D., Cutler, D., 2020. Disparities in coronavirus 2019 reported incidence, knowledge, and behavior among US adults. *JAMA Network Open* 3 e2012403-e2012403.
- Amy Plitt, 2020. How New York City Residential Buildings Are Tackling Coronavirus. *Curbed*, New York, 16 March 2020.
- Baker, M.G., Peckham, T.K., Seixas, N.S., 2020. Estimating the burden of United States workers exposed to infection or disease: a key factor in containing risk of COVID-19 infection. *PLoS One* 15, e0232452.
- Bozdogan, H., 1987. Model selection and Akaike information criterion (aic) - the general theory and its analytical extensions. *Psychometrika* 52, 345–370.
- Brenner, H., Savitz, D.A., Jöckel, K.-H., Greenland, S., 1992. Effects of nondifferential exposure misclassification in ecologic studies. *Am. J. Epidemiol.* 135, 85–95.
- Centers For Disease Control And Prevention, 2020. COVID-19 Guidance for Shared or Congregate Housing [Online]. <https://www.cdc.gov/coronavirus/2019-ncov/community/shared-congregate-house/guidance-shared-congregate-housing.html>.
- Clifford, P., Richardson, S., 1985. Testing the association between two spatial processes. *Stat. Decis.* 2, 155–160.
- Crane, T.A., Surlles, J.G., 2002. Model-dependent variance inflation factor cutoff values. *Qual. Eng.* 14, 391–403.
- Duncan, D.T., Aldstadt, J., Whalen, J., Melly, S.J., Gortmaker, S.L., 2011. Validation of Walk Score® for estimating neighborhood walkability: an analysis of four US metropolitan areas. *Int. J. Environ. Res. Publ. Health* 8, 4160–4179.
- Florax, R.J.G.M., Folmer, H., Rey, S.J., 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Reg. Sci. Urban Econ.* 33, 557–579.
- Gerli, A.G., Centanni, S., Miozzo, M.R., Virchow, J.C., Sotgiu, G., Canonica, G.W., Soriano, J.B., 2020. COVID-19 mortality rates in the European Union, Switzerland, and the UK: effect of timeliness, lockdown rigidity, and population density. *Minerva Med.* 111(4), 308–314.
- Ghinai, I., Woods, S., Ritger, K.A., Mcpherson, T.D., Black, S.R., Sparrow, L., Fricchione, M.J., Kerins, J.L., Pacilli, M., Ruestow, P.S., Arwady, M.A., Beavers, S.F., Payne, D.C., Kirking, H.L., Layden, J.E., 2020. Community transmission of SARS-CoV-2 at two family gatherings - Chicago, Illinois, February-March 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69, 446–450.
- Goldbaum, C., 2020. N.Y.C.'s Subway, a 24/7 Mainstay, Will Close for Overnight Disinfection. *The New York Times*. April 30, 2020.
- Haining, R., 1991. Bivariate correlation with spatial data. *Geogr. Anal.* 23, 210–227.
- Harlem, G., 2020. Descriptive analysis of social determinant factors in urban communities affected by COVID-19. *Journal of Public Health* 42 (3), 466–469.
- Harris, J.E., 2020. The Subways Seeded the Massive Coronavirus Epidemic in New York City. *NBER Working Paper*.
- Ji, Y., Ma, Z., Peppelenbosch, M.P., Pan, Q., 2020. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Global Health* 8, e480.
- Jordan, R.E., Adab, P., Cheng, K.K., 2020. Covid-19: risk factors for severe disease and death. *BMJ* 368, m1198.
- Kang, D., Choi, H., Kim, J.H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int. J. Infect. Dis.* 94, 96–102.
- Leal, C., Bean, K., Thomas, F., Chaix, B., 2012. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am. J. Epidemiol.* 175, 1152–1162.
- Lesage, J.P., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, Boca Raton.
- Liu, L., 2020. Emerging study on the transmission of the Novel Coronavirus (COVID-19) from urban perspective: evidence from China. *Cities* 103, 102759.
- New York City Department Of City Planning, 2018. *Zoning Handbook 2018 Edition*. <https://www1.nyc.gov/site/planning/zoning/zh-2016.page>.
- New York City Department Of City Planning, 2020. *NYC GIS Zoning Features*. New York City Department of City Planning. <https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-gis-zoning.page>.
- New York City Department Of Health And Mental Hygiene, 2020a. COVID-19: Data Summary. <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>.
- New York City Department Of Health And Mental Hygiene, 2020b. COVID-19: FAQ for Residential and Commercial Buildings. <https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-residential-buildings-faq.pdf>.
- New York City Department Of Health And Mental Hygiene, 2020c. Incident Command System For COVID-19 Response - NYC Coronavirus (COVID-19) Data. New York City Department of Health and Mental Hygiene. <https://github.com/nychealth/coronavirus-data>.
- New York City Office Of The Mayor, 2020. As New York Nears First Phase of Reopening. Mayor de Blasio Calls on MTA to Take Concrete Steps to Keep New Yorkers Safe on Mass Transit. *The Official Website of the City of New York*, June 2, 2020.
- New York University Furman Center, 2020. COVID-19 Cases in New York City, a Neighborhood-Level Analysis.
- Price-Haywood, E.G., Burton, J., Fort, D., Seoane, L., 2020. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *N. Engl. J. Med.* 382 (26), 2534–2543.
- Quealy, K., 2020. The Richest Neighborhoods Emptied Out Most as Coronavirus Hit New York City. *The New York Times*.
- Raifman, M.A., Raifman, J.R., 2020. Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. *Am. J. Prev. Med.* 59, 137–139.
- Rocklöv, J., Sjödin, H., 2020. High population densities catalyse the spread of COVID-19. *J. Trav. Med.* 27.
- Song, C., Mei-Po, K., Song, W., Zhu, J., 2017. A Comparison between Spatial Econometric Models and Random Forest for Modeling Fire Occurrence. *Sustainability* 9 (5), 819.
- Sy, K.T.L., Martinez, M.E., Rader, B., White, L.F., 2020. Socioeconomic Disparities in Subway Use and COVID-19 Outcomes in New York City. *medRxiv*.
- U.S. Census Bureau, 2020. 2018 American Community Survey 5-year Estimates. U.S. Census Bureau.
- U.S. Department Of Housing And Urban Development. *Hud Usps Zip Code Crosswalk Files, 6/20/2020*. https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook.
- Ward, M.D., Gleditsch, K.S., 2019. *Spatial Regression Models*. SAGE, Los Angeles.
- Webb Hooper, M., Napoles, A.M., Perez-Stable, E.J., 2020. COVID-19 and Racial/Ethnic Disparities. *JAMA* 323 (24), 2466–2467.
- Wong, D., 2009. The modifiable areal unit problem (MAUP). In: FOTHERINGHAM, A.S., ROGERSON, P.A. (Eds.), *The SAGE Handbook of Spatial Analysis*. SAGE, London.
- Zheng, R., Xu, Y., Wang, W., Ning, G., Bi, Y., 2020. Spatial transmission of COVID-19 via public and private transportation in China. *Trav. Med. Infect. Dis.* 34, 101626.