Subway Ridership, Crowding, or Population Density: Determinants of COVID-19 Infection Rates in New York City

Shima Hamidi, PhD,[1] Iman Hamidi, MSc[2]

Introduction: This study aims to determine whether subway ridership and built environmental factors, such as population density and points of interests, are linked to the per capita COVID-19 infection rate in New York City ZIP codes, after controlling for racial and socioeconomic characteristics.

Methods: Spatial lag models were employed to model the cumulative COVID-19 per capita infection rate in New York City ZIP codes (N=177) as of April 1 and May 25, 2020, accounting for the spatial relationships among observations. Both direct and total effects (through spatial relationships) were reported.

Results: This study distinguished between density and crowding. Crowding (and not density) was associated with the higher infection rate on April 1. Average household size was another significant crowding-related variable in both models. There was no evidence that subway ridership was related to the COVID-19 infection rate. Racial and socioeconomic compositions were among the most significant predictors of spatial variation in COVID-19 per capita infection rates in New York City, even more so than variables such as point-of-interest rates, density, and nursing home bed rates.

Conclusions: Point-of-interest destinations not only could facilitate the spread of virus to other parts of the city (through indirect effects) but also were significantly associated with the higher infection rate in their immediate neighborhoods during the early stages of the pandemic. Policymakers should pay particularly close attention to neighborhoods with a high proportion of crowded households and these destinations during the early stages of pandemics.
Am J Prev Med 2021;60(5):614−620. © 2021 American Journal of Preventive Medicine. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

INTRODUCTION

New York City (NYC) has been particularly hit hard by coronavirus disease 2019 (COVID-19). As of May 25, about a quarter of total COVID-19 deaths in the U.S. occurred in NYC. Research efforts to investigate the determinant factors of the COVID-19 outbreak in NYC mostly focused on socioeconomic factors and reported significant associations between socioeconomic and racial variations and the COVID-19 per capita infection rate, COVID-19 testing rates, and proportion of positive tests.[1−4]

However, there is very little empirical evidence on the effects of subway ridership on COVID-19. The only existing evidence is a non−peer reviewed working paper released by the National Bureau of Economic Research

entitled "The Subways Seeded the Massive Coronavirus Epidemic in New York City." Without any statistical analysis and largely based on observational data, the study argued that the New York subway system was a major disseminator and likely served as the transmission vehicle for the spread of the COVID-19 pandemic, particularly in early days during the first 2 weeks of March. The

From the [1]Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland; and [2]School of Engineering and Computing Sciences, New York Institute of Technology, Vancouver, British Columbia, Canada

Address correspondence to: Shima Hamidi, PhD, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore MD 21205. E-mail: shamidi2@jhu.edu.

study concluded that ZIP codes that are located along the subway lines had a higher number of confirmed cases than ZIP codes that were not served by subway.[5] In the absence of data and statistical analysis, claims in this paper have fueled political debates on conservative media outlets and among policymakers. In NYC, 4 council members cited this paper in their letter to New York Governor Cuomo demanding the complete shutdown of the New York subway system. The petition was largely pushed back by the Metropolitan Transit Authority, emphasizing the critical role of public transit in providing mobility for the frontline essential workers during the pandemic.[6,7]

In addition, there is very little evidence on the relationship between population density and crowding and spatial variations in COVID-19 infection rates at the ZIP code level in NYC. The effects of population density on COVID-19 have been at the center of attention; however, population density is distinct from crowding, which is defined as a large number of people gathered closely together. Crowding could happen in bars, restaurants, sport events, and any other destination that could attract visitors; in other words, points of interest (POIs).[8,9] The Pearson correlation coefficient between population density and POIs per 1,000 population in NYC ZIP codes is <0.052, which also confirms the distinction between the 2 measures. Very little is known about the relationship between different types of crowding venues at the neighborhood level and the COVID-19 infection rate.

Another factor that has been largely missed by existing studies is the extent to which NYC neighborhoods have been emptying out to escape the pandemic. According to the New York Times, as of May 1, in many neighborhoods in Manhattan, between 30% and 50% of residents were gone.[10] It is impossible to contract the virus in NYC if a person is not physically living there. Similarly, nursing homes facilities have been major COVID-19 hotspots in NYC and other parts of the country. In the State of New York, nursing home facilities accounted for >20% of all COVID-19 death cases.[11]

This study is the first to conceptualize and integrate 3 dimensions of crowding, including households, businesses, and subways, in a comprehensive framework. The major aim of this study is to investigate the relationship among these 3 crowding variables, population density, and other confounding factors and the COVID-19 (per capita) infection rate during the early stages (as of April 1) and after the epidemic curve was flattened (as of May 25) at the ZIP code level in NYC. Spatial autoregressive modeling techniques were employed to control for the spatial dependency of observations (ZIP codes) in the sample. The authors hypothesize that, during the early stages, crowding-related factors such as POIs and crowded housing explain the spatial distributions of infection rates, whereas

on May 25, racial and socioeconomic characteristics had the strongest relationship with the per capita infection rate.

METHODS

Study Sample
The sample in this study consisted of 177 ZIP Code Tabulation Areas (ZCTAs) in 5 boroughs of NYC. Data on the cumulative number of COVID-19 tests and the cumulative number of confirmed cases were downloaded from the NYC Department of Health from March 2, 2020 through April 1 and May 25, 2020.[12,13] The outcome variables were the cumulative COVID-19 per capita infection rates at 2 points in time to account for the different nature of the pandemic spread at early stages (April 1) and after the epidemic curve was flattened (from March 2 to May 25). The 2 outcome variables were mapped using quantile categorization in ArcMap, version 10.7.1. In addition, hotspot analyses were performed using the Getis−Ord method to identify clusters of ZCTAs with a high concentration of infection rates (hotspots) and a low concentration of infection rates (coldspots) in NYC (Figure 1).

Measures
The independent variable of greatest interest is subway ridership. Raw data on transit ridership were obtained from the Metropolitan Transit Authority.[14] The Metropolitan Transit Authority releases daily subway ridership data based on entrees and exits for each turnstile by station, which were downloaded and cleaned to compute 3 ridership variables. The first ridership variable represented the prepandemic baseline ridership and was computed as the average weekday ridership in the last week of February, before the first COVID-19 case was confirmed in NYC on February 29. The second and third ridership variables represented the percentage changes in subway ridership relative to the baseline during 2-week time periods before the confirmed positive cases in each model (April 1 and May 25). These 2 variables were estimated backward from observed confirmed cases to estimate transmission that occurred several weeks previously, allowing for the time lag between infection and positive COVID-19 test.[15]

This analysis also accounted for the number of POIs within each ZCTA in NYC, utilizing data from SafeGraph.[16] The SafeGraph database measures foot traffic patterns to POIs based on GPS data from >45 million smartphones in the U.S. POIs include restaurants, cafes, retail shops, movie theaters, parks, and other public places that could attract visitors. Initially, 2 sets of POI variables representing the level of crowding at the baseline and in March were computed for each ZCTA. However, checking the face validity of these variables[17] via ArcMap and Google Maps showed that the most reliable and accurate variable was the number of POIs in each ZCTA per 1,000 population, which was computed and used as a proxy for business crowding in this study.

In addition, analyses controlled for the percentage of residents in each ZCTA who left NYC to escape the pandemic in March and April. The data were borrowed from the New York Times based on aggregated smartphone location data from Descartes Lab and measured the proportion of population who lived in NYC during the last 2 weeks of February but were not living there on May 1.[10] The population-weighted average of Census tracts was calculated to obtain the ZCTA-level variable.
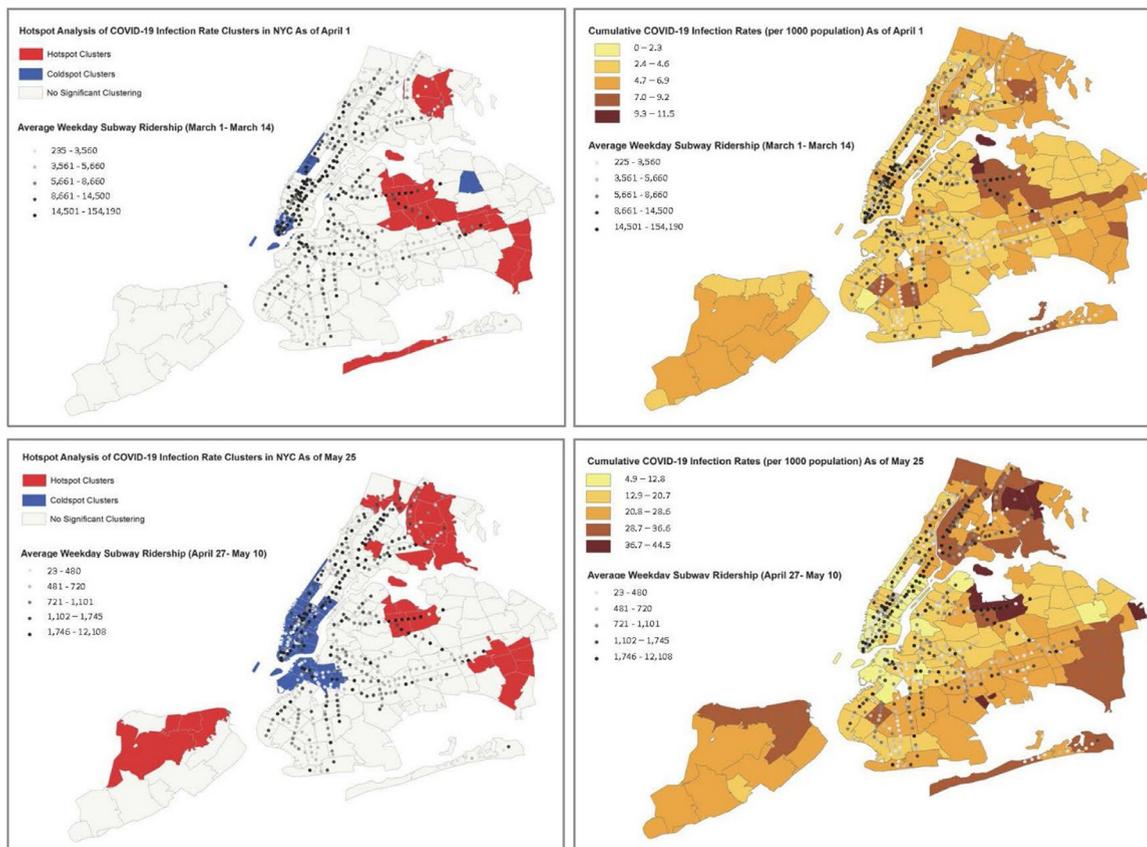
Figure 1. Spatial distribution and hotspot analysis of COVID-19−positive cases per 1,000 population as of April 1 (top right and top left) and May 25 (bottom right and bottom left) by ZIP code in NYC.
COVID-19, coronavirus disease 2019; NYC, New York City.

Employing the same methodology as Yost et al.,[18] an SES index was developed for each ZCTA based on the following variables from the 2018 American Community Survey (5-year estimates)[19]: median household income in the past 12 months; median gross rent; median home value; percentage unemployed (aged ≥16 years); percentage working class (aged ≥16 years); percentage living <150% of poverty line; and education index, which is a weighted combination of the percentage below high school education, high school graduates, and more than high school degrees (adults aged ≥25 years). Higher value of education index represents higher educational attainments. Using principal component analysis, these variables were combined into 1 score for each ZCTA with an eigenvalue of 5.2, which explains 74.8% of the variance following this equation:

$$(\text{median home value} \times 0.141) + (\text{median gross rent} \times 0.17)$$

$$+ (\text{median household income} \times 0.181)$$

$$+ (\text{percentage below poverty} \times -0.161)$$

$$+ (\text{percentage unemployed} \times -0.147)$$

$$+ (\text{percentage working class} \times -0.174)$$

$$+ (\text{education score} \times 0.177).$$

The score was standardized to have a mean of 100 and an SD of 25.

Measures of racial composition characteristics, including percentage Black, percentage Hispanic, and average household size, were computed based on data from the 2018 American Community Survey (5-year estimates).[19] In addition, population density was computed by dividing the ZCTA's total population by the land area in square miles. Finally, using ArcMap, the number of beds in nursing homes and assisted living facilities for each ZCTA was calculated based on data from the Homeland Infrastructure Foundation-level Data[20] and was converted to a per capita rate variable by dividing the number of beds in each ZCTA by ZCTA population. Pearson correlation coefficients between explanatory variables are presented in Appendix Table 1, available online.

Statistical Analysis

The nature of virus spread is a spatialized phenomenon, meaning that the per capita rate of infection rate in a ZCTA is not independent of the infection rate in surrounding ZCTAs. People move beyond the boundary of ZIP codes and so does the virus. The spatial relationship between ZCTAs violates the assumption of ordinary least squares, which requires the unexplained error term to be randomly distributed across observations.[21] This was also confirmed with Moran's *I* analysis of ordinary least squares regression

residuals with a coefficient value of 0.38, which was statistically significant at <0.001 level.

Two forms of spatial autoregressive modeling methods, spatial lag and spatial error, are used to account for spatial dependency among observations.[21] Based on the results of Lagrange multiplier tests, the spatial lag model was selected and performed using R, version 4.0.2 software. The spatial lag model estimates both direct and indirect effects of explanatory variables on COVID-19 infection rates. The indirect effects are through the spatial relationship between observations (ZCTAs). The total effect is the sum of direct and indirect effects, which is also presented in the Results tables. Except for subway ridership variables and nursing home bed rate, all other variables were log-transformed to achieve a better fit with the data, reduce the influence of outliers, and adjust for nonlinearity of the data. Therefore, the coefficients in the Results tables are interpreted as elasticities. The collinearity diagnostic test was also performed, and the tolerance values of explanatory variables, in both models, were higher than the 0.2 threshold,[22] which suggested no issue of multicollinearity.

RESULTS

The results of spatial lag models for the COVID-19 infection rate per 1,000 population as of April 1 and May 25 are shown in Tables 2 and 3, respectively. The comparison between the 2 tables shows noticeable differences between factors that significantly explained the infection rate at these 2 times during the COVID-19 pandemic.

The comparison between the 2 models revealed that, at early stages of the pandemic and before NYC reached the apex, ZCTAs with the higher number of POIs (per capita) as potential venues for crowding reported significantly higher per capita infection rates. The concentration of POIs in a ZIP code facilitates social interactions and closer contacts and could lead to the transmission of disease in the immediate neighborhood. This was no longer the case on May 25, possibly because of business closures and the implementation of stay-at-home orders.

However, the average household size, representing the level of crowding in households, was the only crowding variable that was significant in both the April 1 and May 25 models. On April 1, doubling household size was associated with a 36% increase in COVID-19 infection rate per 1,000 population. The spread of COVID-19 may begin in schools, workplaces, or POIs, but eventually neighborhoods with relatively larger households are the most vulnerable to the possibilities of transmission. These findings suggest that neighborhoods with relatively larger households, such as immigrant communities, are more vulnerable to the spread of virus during the pandemic.

After controlling for the variables that represented the level of crowding, population density had no significant relationship with the COVID-19 infection rate on April 1 and May 25. These findings indicate that variables representing different dimensions of crowding might be better predictors of the per capita infection rate than population density. Recent national polls show that residents in dense places are more likely to voluntarily

Table 1. Variable Descriptions, Data Sources, and Descriptive Statistics

Variable/description	Data sources	Mean (SD)
Dependent variables		
ln of confirmed cases per 1,000 (as of April 1)	NYC Department of Health 2020	4.59 (1.7)
ln of confirmed cases per 1,000 (as of May 25)	NYC Department of Health 2020	21.9 (8.5)
Independent variables		
ln of percent Black population	ACS 2018	21.7 (24.9)
ln of percent Hispanic population	ACS 2018	26.1 (19.5)
ln of average household size	ACS 2018	2.6 (0.51)
ln of standardized SES index	Developed by authors based on data from ACS 2018	100 (25)
ln of the number of POIs per 1,000	SafeGraph 2020	11.28 (13.1)
ln of population density	ACS 2018 (5-year estimates)	39,886 (25,067)
ln of percent emptying out	The New York Times 2020	11.0 (10.4)
Number of nursing home beds per 1,000	HIFLD 2019	5.7 (10.4)
Subway ridership in 1,000s (baseline)	MTA 2020	172.1 (254.6)
% change in subway ridership (March 1−March 14, relative to the baseline)	MTA 2020	−1.82 (5.94)
% change in subway ridership (April 27−May 10, relative to the baseline)	MTA 2020	−60.1 (40.41)
ln of tests per 1,000 (as of April 1)	NYC Department of Health 2020	9.1 (2.6)
ln of tests per 1,000 (as of May 25)	NYC Department of Health 2020	72.2 (21.9)

Note: Descriptive statistics were calculated before log-transformation.
ACS, American Community Survey; HIFLD, Homeland Infrastructure Foundation-Level Data; ln, natural logarithm; MTA, Metropolitan Transit Authority; NYC, New York City; POI, point of interest.

Table 2. Results of the Spatial Lag Model as of April 1

Variables	b	SE	t-ratio	p-value	Total effects
Intercept	−0.1847	0.6710	−0.2752	0.783	—
ln of percent Black	0.0239	0.0122	1.9574	**0.047**	0.0245
ln of percent Hispanic	0.0067	0.0267	0.2499	0.803	0.0068
ln of average household size	0.7158	0.1138	6.2890	**<0.001**	0.7350
ln of SES index	−0.3398	0.1083	−3.1378	**0.002**	−0.3488
ln of POI per 1,000 population	0.0722	0.0372	1.9632	**0.049**	0.0742
ln of population density	0.0054	0.0226	0.2387	0.811	0.0055
Subway ridership per 1,000 population (baseline)	0.000086	0.00007	1.1202	0.263	0.000088
% change in subway ridership (March 1−March 14, relative to the baseline)	−0.0039	0.0030	−1.3059	0.192	−0.0040
Number of nursing home beds per 1,000 population	0.0018	0.0014	1.2647	0.206	0.0018
ln of tests per 1,000 population (April 1)	1.0778	0.0538	20.0433	**<0.001**	1.1066

Note: Boldface indicates statistical significance ($p<0.05$). Outcome variable is the natural log of the number of confirmed cases per 1,000 population as of April 1.
ln, natural logarithm; POI, point of interest.

Table 3. Results of the Spatial Lag Model as of May 25

Variables	b	SE	t-ratio	p-value	Total effects
Intercept	10.785	0.470	22.94	**<0.001**	—
ln of percent Black	0.0272	0.0080	3.39	**<0.001**	0.0269
ln of percent Hispanic	0.0432	0.0173	2.49	**0.013**	0.0428
ln of average household size	0.3621	0.0764	4.74	**<0.001**	0.3592
ln of SES index	−0.2436	0.0713	−3.42	**<0.001**	−0.2417
ln of POI per 1,000 population	−0.0418	0.0244	−1.71	0.087	−0.0414
ln of population density	−0.0259	0.0168	−1.54	0.123	−0.0257
Subway ridership (baseline)	0.000082	0.00007	1.67	0.094	0.000081
% change in subway ridership (April 27−May 10, relative to the baseline)	0.0002	0.0003	0.75	0.453	0.00024
Number of nursing home beds per 1,000 population	0.0027	0.0010	2.81	**0.005**	0.0027
ln of % emptying out	−0.1114	0.0187	−5.97	**<0.001**	−0.110
ln of tests per 1,000 population (as of May 25)	0.9581	0.0433	22.15	**<0.001**	0.951

Note: Boldface indicates statistical significance ($p<0.05$). Outcome variable is natural log of the number of confirmed cases per 1,000 population as of May 25.
ln, natural logarithm; POI, point of interest.

engage in social distancing, being more cognizant of the threat.[23]

After controlling for the level of crowding and population density, the baseline subway ridership per 1,000 population had no significant relationship with the cumulative ZCTA per capita infection rates on April 1 and May 25. Similarly, the changes in subway ridership relative to baseline were not significantly related to the COVID-19 infection rates on April 1 and May 25. These findings were confirmed with follow-up *t*-tests, which showed no significant differences between ZCTAs with no subway station and ZCTAs that were served by subway in terms of the per capita infection rate on April 1 and May 25 (*p*-values of 0.685 and 0.735, respectively).

In contrast, from the list of control variables, racial and socioeconomic compositions were among the most significant predictors of the spatial variation in COVID-

19 per capita infection rates in NYC, even more so than variables such as POI rates, density, and nursing home bed rates. These findings align with recent findings about the increased prevalence of COVID-19 in low-income, Hispanic-, and Black-majority neighborhoods in NYC, possibly because of their greater risk of occupational exposure and other key social determinants of health.[1,2,24−26]

DISCUSSION

This study found no evidence that subway ridership was related to the COVID-19 infection rate in NYC. The recent experience of a few developed countries in tracing infection clusters confirms this finding. In Japan, since the state of emergency was lifted in late May, the majority of infection clusters were traced to gyms, bars, music clubs, and karaoke rooms, whereas not even a single

infection cluster, defined as ≥3 COVID-19 infections linked by contact, were associated with its highly popular and often crowded commuter trains.[27] Similarly, according to the National Public Health Institute in France, between May 9 and June 15, from 150 clusters of new COVID-19 infections, none were traced to the nation's public transit system, consisting of 6 subway systems, trams, light rail, and bus networks. In fact, most of these clusters had emerged in hospitals, workplaces, and homeless shelters.[28]

In addition, findings about the insignificant link between population density and the per capita COVID-19 infection rate run counterintuitive to recent dialogues in news media outlets and among policymakers that highlight the role of density on the COVID-19 spread, particularly in NYC.[29] CNN, for instance, quoted Governor Cuomo of New York in an article on May 2, 2020 and wrote "It's very simple. It's about density. It's about the number of people in a small geographic location allowing that virus to spread.... Dense environments are its feeding grounds."[30] Before the COVID-19 pandemic, extensive research has confirmed the environmental and public health benefits of dense, compact, and transit-accessible developments.[31−34] This study found no evidence that population density was associated with a higher per capita COVID-19 infection rate. Indeed, crowding (and not density) was associated with the higher infection rate on April 1.

Limitations

One limitation of this study is that the analyses were based on ZCTA-level aggregated data and did not control for the individual-level variations and interactions among variables. Therefore, findings could not draw individual-level conclusions, particularly related to socioeconomic factors. In addition, the aggregated nature of this study limits the ability to control for individual-level factors, such as underlying health conditions that might be associated with the severity of disease and the likelihood of testing. Also, the transit ridership variables only represent the subway ridership, and findings are not generalizable to other modes of public transit, such as bus or ride-hailing services. It is possible that other modes of public transportation, such as bus transit, which are more widely accessible across all ZCTAs in the study area have a significant relationship with the COVID-19 per capita infection rate. In addition, the POI variables were computed based on GPS data from smartphones and may underrepresent those who do not have a smartphone or opt to turn off the location feature of their smartphone. Furthermore, NYC is the densest U.S. city, has the highest transit ridership, and may not represent a typical American city. Finally, the socioeconomic and demographic variables in this study are mainly based on Census data and may underrepresent noncitizens and undocumented immigrants. Therefore, the SES and racial composition of ZCTAs may not have been fully captured with measures in this study.

CONCLUSIONS

This study offers empirical evidence that distinguishes between population density and different forms of crowding and shows that crowded households, measured in terms of household size, are associated with the significantly higher per capita infection rate across NYC ZIP codes. In addition, destinations (POIs) that could attract visitors not only could facilitate the spread of virus to other parts of the city (through indirect effects) but also are significantly associated with the higher per capita infection rate in their immediate neighborhoods, particularly during the early stages of the pandemic. Policymakers should pay particularly close attention to neighborhoods with a high proportion of crowded households and these destinations (or POIs) during the early stages of pandemics.

Another major takeaway of this study is that investigators found no evidence that a higher per capita subway ridership and percentage changes in subway ridership are related to the COVID-19 infection rate across the NYC ZIP codes. These findings challenge Harris,[5] who argued that the ZCTAs along the subway lines had significantly higher infection rates than ZIP codes that were not served by subway. Still, it may be too early to draw a definitive conclusion, and more studies are needed to further investigate the role of the transit system (including other transit modes) on COVID-19 pandemic spread through contact tracing.

SUPPLEMENTAL MATERIAL

Supplemental materials associated with this article can be found in the online version at https://doi.org/10.1016/j.amepre.2020.11.016.

REFERENCES

1. Borjas GJ. Demographic determinants of testing incidence and COVID-19 infections in New York City neighborhoods. HKS Working Paper No. RWP20-008. *SSRN*. Online April 10, 2020. https://doi.org/10.2139/ssrn.3572329.

2. Credit K. Neighborhood inequity: exploring the factors underlying racial and ethnic disparities in COVID-19 testing and infection rates using ZIP code data in Chicago and New York. *Reg Sci Policy Pract.* 2020;12(6):1249–1271. https://doi.org/10.1111/rsp3.12321.

3. Lieberman-Cribbin W, Tuminello S, Flores RM, Taioli E. Disparities in COVID-19 testing and positivity in New York City. *Am J Prev Med.* 2020;59(3):326–332. https://doi.org/10.1016/j.amepre.2020.06.005.

4. Almagro M, Orane-Hutchinson A. JUE insight: the determinants of the differential exposure to COVID-19 in New York City and their evolution over time. *J Urban Econ.* 2020 In press. Online October 28, 2020. https://doi.org/10.1016/j.jue.2020.103293.

5. Harris JE. The subways seeded the massive coronavirus epidemic in New York City. NBER working paper 27021. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27021. Revised August 2020. Accessed March 31, 2021.

6. Bliss L. The New York subway got caught in the coronavirus culture war. *Bloomberg CityLab.* April 21, 2020. https://www.bloomberg.com/news/articles/2020-04-21/the-tenuous-link-between-the-subway-and-covid-19. Accessed September 10, 2020.

7. Sadik-Khan J, Solomonow S. Fear of public transit got ahead of the evidence. *The Atlantic.* June 14, 2020. https://www.theatlantic.com/ideas/archive/2020/06/fear-transit-bad-cities/612979/. Accessed September 10, 2020.

8. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *J Am Plann Assoc.* 2020;86(4):495–509. https://doi.org/10.1080/01944363.2020.1777891.

9. Hamidi S, Ewing R, Sabouri S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: early evidence from 1,165 metropolitan counties in the United States. *Health Place.* 2020;64:102378. https://doi.org/10.1016/j.healthplace.2020.102378.

10. Quealy K. The richest neighborhoods emptied out most as coronavirus hit New York City. *The New York Times.* May 15, 2020. https://www.nytimes.com/interactive/2020/05/15/upshot/who-left-new-york-coronavirus.html. Accessed September 10, 2020.

11. The New York Times. More than one-third of U.S. coronavirus deaths are linked to nursing homes. *The New York Times.* Updated November 20, 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html. Accessed September 10, 2020.

12. New York City Department of Health. Confirmed and probable COVID-19 deaths. New York, NY: New York City Department of Health. https://www1.nyc.gov/site/doh/covid/covid-19-data-archive.page. Accessed January 8, 2021.

13. Buchanan L, Patel JK, Rosenthal BM, Singhvi A. A month of coronavirus in New York City: see the hardest-hit areas. *The New York Times.* April 1, 2020. https://www.nytimes.com/interactive/2020/04/01/nyregion/nyc-coronavirus-cases-map.html. Accessed on September 10, 2020.

14. Metropolitan Transit Authority. Turnstile data. http://web.mta.info/developers/turnstile.html. Accessed July 6, 2020.

15. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* 2020;584(7820):257–261. https://doi.org/10.1038/s41586-020-2405-7.

16. SafeGraph. Places schema. https://docs.safegraph.com/docs. Accessed July 6, 2020.

17. Duke C, Hamidi S, Ewing R. Validity and reliability. In: Ewing R, Park K, eds. *Basic Quantitative Research Methods for Urban Planners.*

18. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control.* 2001;12(8):703–711. https://doi.org/10.1023/a:1011240019516.

19. American Community Survey. American Community Survey 5-Year Estimates. U.S. Census Bureau. https://www.census.gov/programs-surveys/acs/data.html. Updated March 30, 2020. Accessed January 8, 2021.

20. U.S. Department of Homeland Security. Homeland Infrastructure Foundation-Level Data (HIFLD). https://hifld-geoplatform.opendata.arcgis.com/. Accessed January 8, 2021.

21. Anselin L. *Spatial Econometrics: Methods and Models.* Dordrecht, Netherlands: Springer Science + Business Media, 1988.

22. O'brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 2007;41:673–690. https://doi.org/10.1007/s11135-006-9018-6.

23. Saad L. Americans rapidly answering the call to isolate, prepare. Gallup. March 20, 2020. https://news.gallup.com/poll/297035/americans-rapidly-answering-call-isolate-prepare.aspx. Accessed September 10, 2020.

24. Yancy CW. COVID-19 and African Americans. *JAMA.* 2020;323(19):1891–1892. https://doi.org/10.1001/jama.2020.6548.

25. Quinn SC, Kumar S. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur Bioterror.* 2014;12(5):263–273. https://doi.org/10.1089/bsp.2014.0032.

26. Kumar S, Quinn SC, Kim KH, Daniel LH, Freimuth VS. The impact of workplace policies and other social factors on self-reported influenza-like illness incidence during the 2009 H1N1 pandemic. *Am J Public Health.* 2012;102(1):134–140. https://doi.org/10.2105/AJPH.2011.300307.

27. Normile D. Japan ends its COVID-19 state of emergency. *Science.* May 26, 2020. https://doi.org/10.1126/science.abd0092. Accessed March 31, 2021.

28. O'Sullivan F. In Japan and France, riding transit looks surprisingly safe. Bloomberg CityLab. June 9, 2020. https://www.bloomberg.com/news/articles/2020-06-09/japan-and-france-find-public-transit-seems-safe. Accessed September 10, 2020.

29. Rosenthal BM. Density is New York City's big "enemy" in the coronavirus fight. *The New York Times.* March 23, 2020. https://www.nytimes.com/2020/03/23/nyregion/coronavirus-nyc-crowds-density.html. Accessed September 10, 2020.

30. Shoichet CE, Jones A. Coronavirus is making some people rethink where they want to live. CNN. May 2, 2020. https://www.cnn.com/2020/05/02/us/cities-population-coronavirus/index.html. Accessed September 10, 2020.

31. Hamidi S. Urban sprawl and the emergence of food deserts in the USA. *Urban Stud.* 2020;57(8):1660–1675. https://doi.org/10.1177/0042098019841540.

32. Hamidi S, Ewing R, Tatalovich Z, Grace JB, Berrigan D. Associations between urban sprawl and life expectancy in the United States. *Int J Environ Res Public Health.* 2018;15(5):861. https://doi.org/10.3390/ijerph15050861.

33. Ewing R, Hamidi S, Grace JB. Urban sprawl as a risk factor in motor vehicle crashes. *Urban Stud.* 2016;53(2):247–266. https://doi.org/10.1177/0042098014562331.

34. Ewing R, Hamidi S. *Costs of Sprawl.* New York, NY: Routledge, 2017.

New York, NY: Routledge, 2020:88–106. https://doi.org/10.4324/9780429325021-6.