Received: 8 May 2020 | Revised: 22 July 2020 | Accepted: 27 July 2020

DOI: 10.1111/rsp3.12321

ORIGINAL ARTICLE

Neighbourhood 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

Kevin Credit

Center for Spatial Data Science, University of Chicago, IL, USA

Correspondence
Kevin Credit, Center for Spatial Data Science, University of Chicago, 1155 E 60th St, Room 211A, Chicago, IL 60637, USA.
Email: kcredit@uchicago.edu

Abstract

This paper compares ZIP code-level data on observed COVID-19 testing and case rates for the City of Chicago and New York City to better understand both: (i) the extent to which racial and ethnic disparities in COVID-19 testing and case rates exist at the neighbourhood level; and (ii) the most important neighbourhood-level drivers of these observed disparities. Through exploratory spatial mapping and econometric approaches, the paper finds that, across both cities, Hispanic-majority neighbourhoods have significantly lower testing rates than other racial/ethnic neighbourhood types, even when controlling for observed infection rates—which are also significantly higher for Hispanic-majority neighbourhoods. At the same time, White-majority neighbourhoods have significantly higher testing rates and lower observed infection rates. Given this observed disparity, the paper also examines a range of underlying factors that are potentially driving observed neighbourhood-level COVID-19 case rates. The findings suggest that higher socio-economic status and the provision of healthy, active built environments are significantly negatively associated with COVID-19 infection rates, while several aspects of social vulnerability are significant positive predictors of COVID-19 infection rates. These findings suggest that the health benefits from higher density, walkable built environments may play a larger “protective” role from observed COVID-19 case rates at the neighbourhood-level than previously assumed, while at the same time indicating that the increased prevalence of COVID-19 in Hispanic- and Black-majority neighbourhoods may be in part due to their greater risk of occupational exposure and multi-generational household structure (particularly for Hispanic-majority neighbourhoods).

KEYWORDS

cross-sectional models, economics of minorities and races, health, education, and welfare, land use patterns, non-labor discrimination, quantile regressions, spatial models, treatment effect models

JEL CLASSIFICATION

I1; I14; I31; C21; R14

1 | INTRODUCTION AND BACKGROUND

The novel coronavirus SARS-CoV-2 (COVID-19) was first discovered in 2019 in the Hubei province of China and has since spread across the world, creating a global pandemic that has resulted in over 7,500,000 confirmed cases and over 548,000 deaths (as of 8 July 2020) (Worldometer, 2020). While significant research effort has been dedicated to tracking the dynamics of the spread of the virus at the national and county levels (CSDS, 2020; Dong, Du, & Gardner, 2020; JHU, 2020), data at smaller spatial scales—such as ZIP code level—has only recently been made available in many jurisdictions, allowing for neighbourhood-level analyses. These data provide the first opportunity to begin to explore the spatial associations between socio-demographic, economic, and built environment characteristics and neighbourhood-level virus activity.

Neighbourhood-level analyses are particularly important in the context of social determinants of health (SDOH), a body of research that recognizes that the social and economic conditions in which people live shape their ultimate health outcomes in a number of critically-important ways (Kolak, Bhatt, Park, Padrón, & Molefe, 2020; Solar & Irwin, 2010). For example, in *The spirit level*, Wilkinson and Pickett (2009) clearly illustrate the “pernicious effects” that economic inequality has on numerous health outcomes, including chronic physical disease (such as obesity), mental health, drug addiction, and violence. In the US, historic and continuing patterns of racial and ethnic segregation often overlap with environmental degradation and socio-economic deprivation, creating a context in which non-White communities are often systematically faced with higher rates of chronic disease, lower levels of infrastructure investment, lower access to healthcare and healthy food options, higher crime, and lower life expectancies (Braveman, Cubbin, Egerter, Williams, & Pamuk, 2010; Smith, 1977, 1994; Williams, Mohammed, Leavell, & Collins, 2010; Williams, Priest, & Anderson, 2016). At the same time, recent work suggests that “deaths of despair” related to structural inequality have significantly impacted lower income and working class Whites (in addition to those in racial and ethnic minorities) (Case & Deaton, 2020), constituting a public health crisis in the US that may be exacerbated by the COVID-19 pandemic. Still, a comprehensive analysis of federal data from the Centers for Diseases Control and Prevention (CDC) by the *New York Times* on 5 July 2020 indicates that rates of COVID-19 infection in Latino and Black people are much higher than rates for White people – 73 cases per 100,000 people for Latino people and 62 per 100,000 for Black people, compared to only 23 cases per 100,000 for White people, demonstrating a systematic racial/ethnic disparity that persists across the country (Oppel, Gebeloff, Lai, Wright, & Smith, 2020).

As one of the most segregated cities in the US, Chicago, IL provides an important case in which to explore the possible socio-demographic disparities in COVID-19 case rates and testing: according to a recent analysis of life expectancy at the Census tract scale, life expectancies in Cook County (containing the City of Chicago) range from 59.9 (lower than countries such as Angola, Mali, and Zimbabwe) to 90 (which, as a country, would be the highest in the world—higher than Monaco at 89.4) (Arias, Escobedo, Kennedy, Fu, & Cisewski, 2018; CIA, 2017; NCHS, 2015). Following this pattern, hospitalizations and deaths from COVID-19 at the individual case level have—to this point—been strongly associated with minority groups and those with underlying conditions. According to the City of Chicago Department of Public Health (CDPH), as of 7 July, 75.5% of deaths have been from Latinx or Black residents, compared to only 19.2% of deaths from White non-Hispanic residents, while 92.1% of Chicago residents who have died from COVID-19 have had evidence of at least one chronic underlying health condition[1] (CDPH, 2020).

New York City, site of the nation's largest COVID-19 outbreak, has also seen similar disparities in case outcomes: as of 8 July, according to the New York City Health Department, 58.6% of confirmed deaths from COVID-19 have come from Black and Hispanic residents (NYCHD, 2020a), while 87.6% of confirmed deaths have been from patients with underlying health conditions (NYCHD, 2020b).

Given these initial indications of starkly disparate outcomes by racial/ethnic group and underlying healthiness at the individual level, the goal of this paper is to explore two primary questions: first, foundationally, to what extent is there racial or ethnic disparity in testing and observed infection rates at the neighbourhood level in Chicago and New York? Second, if disparities based on racial/ethnic neighbourhood type are observed, the paper then asks what underlying (associated) features of neighbourhoods—including socio-economic status, health-related characteristics of the physical environment, and other indicators of COVID-19 vulnerability such as age, occupation, and household structure—most strongly predict observed COVID-19 case rates (CDC, 2020b). This second question is particularly useful because it allows us to better understand how *specific* underlying features of neighbourhoods appear to be driving systematic differences in COVID-19 outcomes. This knowledge can then help to suggest targeted policy interventions to reduce these disparities in this and future health crises.

As the results of this paper show—through a comparison of exploratory mapping and econometric analyses at the ZIP code-level in Chicago and New York—Hispanic-majority neighbourhoods have significantly lower testing rates than other racial/ethnic neighbourhood types, even when controlling for observed infection rates, which are also significantly higher for Hispanic-majority neighbourhoods, as well as Black (non-Hispanic)-majority neighbourhoods in New York. At the same time, White (non-Hispanic)-majority neighbourhoods have significantly higher testing rates and lower observed infection rates than other neighbourhood types. As for the underlying factors driving observed neighbourhood-level COVID-19 case rates, higher socio-economic status and the provision of healthy, active built environments in the form of higher population density and higher prevalence of pedestrian and bicycle commuting are significantly negatively associated with COVID-19 infection rates (even when controlling for neighbourhood accessibility to healthy food and healthcare services), while several aspects of social vulnerability—in particular, the prevalence of crowded housing and the percentage of workers in healthcare service occupations—are significant positive predictors of COVID-19 infection rates. These findings suggest that the health benefits from higher density, walkable built environments may play a larger "protective" role from observed COVID-19 case rates at the neighbourhood-level than previously assumed, which is in line with a large body of literature that shows direct links between a neighbourhood's built environment and positive health outcomes (Booth, Pinkston, & Poston, 2005; Cerin, Nathan, van Cauwenberg, Barnett, & Barnett, 2017; Feng, Glass, Curriero, Stewart, & Schwartz, 2010; Frank et al., 2006; Frank, Andresen, & Schmid, 2004; Lindstrom, 2008; Saelens, Sallis, Black, & Chen, 2003; Wang, Chau, Ng, & Leung, 2016).

[1]Through 5 June, the last day for which this data was provided publicly by CDPH.

At the same time, these results indicate that the increased prevalence of COVID-19 in Hispanic- and Black (non-Hispanic)-majority neighbourhoods may be in part due to their greater risk of occupational exposure (given higher rates of healthcare service employment) and multi-generational household structure, particularly for Hispanic-majority neighbourhoods.

2 | DATA AND METHODS

The data for this paper, shown with their descriptive statistics in Table 1, come from a variety of sources: the number of COVID-19 tests and confirmed cases by ZIP code as of 16 April and 1 May 2020, come from the websites of the Illinois Department of Public Health (IDPH, 2020) and the New York City Health Department (NYCHD, 2020c). With no comprehensive federal policy on testing, decisions on testing requirements and the availability of tests varied significantly from state to state (CDC, 2020a). For instance, in Illinois, residents needed to display CDC-approved symptoms or obtain a referral from a doctor in order to be tested up to May in order to obtain a COVID-19 test (Petrella, Munks, & Lourgos, 2020). This means that the officially-recorded recorded tests and cases used in this paper (through the early stage of the pandemic) are almost certainly skewed towards more serious symptomatic infections, since this was a direct criteria for being tested in the first place (with tests often in short supply) (Goodnough, Thomas, & Kaplan, 2020).

Confirmed cases and tests were divided by ZIP code population in order to obtain a normalized rate per population. This information was joined to census data at the ZIP code Tabulation Area (ZCTA) from the 2014–2018 American Community Survey (ACS) for ZIP codes with centroids within the City of Chicago and New York City boundaries. A range of variables assessing various aspects of social vulnerability were considered, building on the CDC's Social Vulnerability Index (SVI) framework (Flanagan, Gregory, Hallisey, Heitgerd, & Lewis, 2011), with a specific focus on aspects of vulnerability that intersect with known COVID-19 risk factors (including older age, obesity, and racial and ethnic minority group status) and infection pathways (such as crowded housing and healthcare service occupations, among others) (CDC, 2020).

As Table 1 shows, the chosen explanatory variables were grouped into four categories based on four different conceptual aspects of COVID-19 vulnerability: (i) racial/ethnic neighbourhood type—percentage White (non-Hispanic), Black (non-Hispanic), and Hispanic population by ZIP code were used to classify each ZIP code as either majority-White (non-Hispanic), majority-Black (non-Hispanic), or majority-Hispanic, based on a level of >50% population of the given racial/ethnic group in that ZIP code; given the strong historical pattern of segregation in US urban areas, these are key neighbourhood indicators of interest. (ii) Socio-economic status is proxied for by median household income. (iii) Variables related to "healthy, active built environments" capture various aspects of a neighbourhood's physical environment that influence health outcomes, including population density (per m^2) and percent pedestrian and bike commuting, which are both indicators of the kinds of dense, walkable built environments that encourage physical activity and generally lead to lower rates of chronic disease and obesity (Besser & Dannenberg, 2005; Booth et al., 2005; Cerin et al., 2017; Feng et al., 2010; Frank et al., 2004, 2006; Lindstrom, 2008; Saelens et al., 2003; Wang et al., 2016), as well as a hospital accessibility score[2] that proxies for underlying spatial accessibility to healthcare resources, and the percentage of tracts identified as "food deserts[3]" by the 2017 USDA report "Low-Income and Low-Supermarket-Access Tracts, 2010-2015" within each ZCTA was included as a general indicator of spatial access to healthy foods (Rhone et al., 2017). Finally, (iv) several COVID-19

[2]The Euclidean distance from each ZCTA centroid in the Chicago metropolitan region to the point location of hospitals provided by the Department of Homeland Security (HIFLD, 2020) was calculated using the "access" package in Python 3.7 and then weighted by a distance decay parameter (β) of −.5 within a cutoff threshold of 16,000 m (roughly 10 miles). These weighted distances were then summed by ZCTA centroid, and a max-min transformation was performed on the dataset to place the resulting scores on a range of 0–100 (100 being the highest hospital accessibility).

[3]Tracts identified as "low-income" and "low-access" at ½ mile in urban areas and 10 miles in rural areas (LILATracts_halfAnd10) were classified as food deserts (Rhone, Ver Ploeg, Dicken, Williams, & Breneman, 2017); at the ZIP code level, the variable was constructed by taking the percentage of these food desert tracts out of all tracts whose centroids were located within each ZCTA.

TABLE 1 Descriptive statistics for variables of interest

City	Concept	Variable description	Variable name	Min	Mean	Median	Max
New York City	COVID-19	Case Rate (4/16) (per pop.)	CASER4_16	0.003	0.014	0.014	0.029
		Testing Rate (4/16) (per pop.)	CASER4_16	0.010	0.025	0.024	0.053
		Case Rate (5/1) (per pop.)	CASER5_1	0.004	0.019	0.019	0.038
		Testing Rate (5/1) (per pop.)	TESTR5_1	0.017	0.042	0.040	0.082
	Racial/Ethnic neighbourhood types	Black non-Hispanic-Majority Neighbourhood	BLKNH	0	0.153	0	1
		Hispanic-Majority Neighbourhood	HISPNH	0	0.153	0	1
		White non-Hispanic-Majority Neighbourhood	WNH	0	0.356	0	1
	Socio-economic status	Median household income	MEDINC	$21,149	$73,674	$66,483	$250,001
	Healthy, active built environments	Population density (per m^2)	POPDENS	0.0005	0.0174	0.0150	0.0589
		Percent pedestrian and bike commuters	PERPEDB	0.9%	11.2%	7.6%	45.4%
		Hospital accessibility score	WS__5	0	51.74	56.90	100
		Percent food desert tracts	FDTRTPER	0.0%	1.8%	0.0%	50.0%
	Vulnerability (age, occupation, household structure)	Percent population 65+	PER65	0.5%	14.3%	13.6%	29.0%
		Percent healthcare service workers	PERHSRV	0.0%	5.2%	4.7%	15.5%
		Percent housing units w/ > 1 person per room	PERCROWD	0.9%	8.3%	7.2%	29.7%
Chicago	COVID-19	Case Rate (4/16) (per pop.)	CASER4_16	0.002	0.005	0.004	0.009
		Testing Rate (4/16) (per pop.)	CASER4_16	0.011	0.018	0.016	0.043
		Case Rate (5/1) (per pop.)	CASER5_1	0.003	0.008	0.008	0.015
		Testing Rate (5/1) (per pop.)	TESTR5_1	0.020	0.033	0.031	0.076
	Racial/Ethnic neighbourhood types	Black non-Hispanic-Majority Neighbourhood	BLKNH	0	0.273	0	1
		Hispanic-Majority Neighbourhood	HISPNH	0	0.145	0	1
		White non-Hispanic-Majority Neighbourhood	WNH	0	0.418	0	1

(Continues)

TABLE 1 (Continued)

City	Concept	Variable description	Variable name	Min	Mean	Median	Max
	Socio-economic status	Median household income	MEDINC	$20,991	$63,586	$60,073	
	Healthy, active built environments	Population density (per m^2)	POPDENS	0.0005	0.0063	0.0053	0.0157
		Percent pedestrian and bike commuters	PERPEDB	0.2%	10.9%	5.9%	55.0%
		Hospital accessibility score	WS__5	0	42.06	47.90	100
		Percent food desert tracts	FDTRTPER	0.0%	20.8%	11.8%	95.5%
	Vulnerability (age, occupation, household structure)	Percent population 65+	PER65	2.2%	12.3%	12.9%	21.8%
		Percent healthcare service workers	PERHSRV	0.0%	3.6%	3.2%	11.0%
		Percent housing units w/ > 1 person per room	PERCROWD	0.4%	3.8%	3.2%	10.9%

TABLE 2 Comparison of regression results for COVID-19 testing rate by racial/ethnic neighbourhood type

DV: Log (Testing Rate 5/1)		
Variable	**Chicago**	**New York City**
Black non-Hispanic-Majority Neighbourhood		
Hispanic-Majority Neighbourhood	–	–
White non-Hispanic-Majority Neighbourhood	+	+
Case Rate 5/1	+	+
Lambda	NA	+
Pseudo–/Adjusted R^2	0.4118	0.8853
Observations	54	177
Robust SE	No	Yes

Note: Due to the difference in sample size and model type between regressions run for each frame of reference, this table displays only the sign of significant (at $p \le .10$) results for coefficients for comparison; full regression results and diagnostics available in Appendix A.1.

TABLE 3 Comparison of regression results for COVID-19 case rate by racial/ethnic neighbourhood type

DV: Log (Case Rate 5/1)		
Variable	**Chicago**	**New York City**
Black non-Hispanic-Majority Neighbourhood		+
Hispanic-Majority Neighbourhood	+	+
White non-Hispanic-Majority Neighbourhood	–	–
Testing Rate 4/16	+	+
W_Log(Case Rate)	+	NA
Lambda	NA	+
Pseudo–/Adjusted R^2	0.6595	0.8025
Observations	54	177
Robust SE	No	Yes

Note: Due to the difference in sample size and model type between regressions run for each frame of reference, this table displays only the sign of significant (at $p \le .10$) results for coefficients for comparison; full regression results and diagnostics available in the Appendix A.2.

vulnerability indicators were included, including percent population 65 and older, percent workers in healthcare service occupations, and percentage housing units with more than 1 person per room (considered "crowded" based on the SVI) (CDC, 2020b).

This paper takes an exploratory approach to mapping, analysing, and modeling these data. Maps (shown in Figures 1 and 2) and parallel co-ordinate plots display the general spatial patterns of the variables of interest are used to better understand how they co-vary (GeoDa v.1.14 was used to produce the plots with linked selection). To formalize the exploratory analysis, three sets of linear regression models were estimated in GeoDaSpace v.1.2. The first set of models assess whether or not particular types of neighbourhoods demonstrate lower-than-expected testing rates. This is important to understand given the widespread lack of testing availability in the early stages of the pandemic. To examine whether the testing rate is similar across all three racial/ethnic neighbourhood groups, two models were run (one each for Chicago and New York) with testing rate (as of 5/1) as the dependent variable,

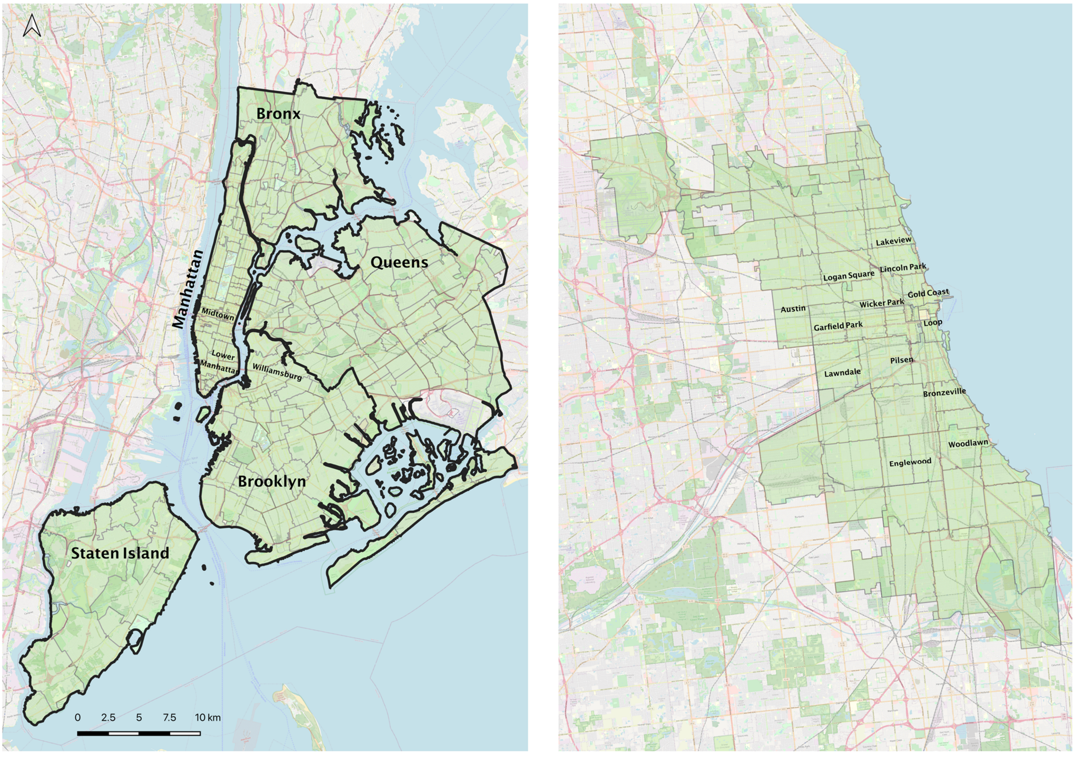

FIGURE 1 Map showing boroughs and neighbourhoods of note in New York City (left panel) and Chicago (right panel)

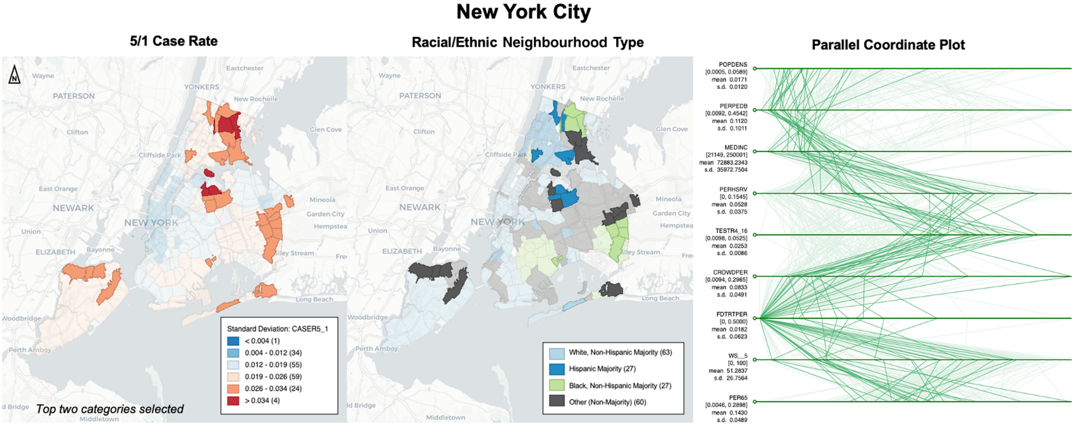

FIGURE 2 New York City maps comparing COVID-19 case rate and racial/ethnic neighbourhood type, with a parallel co-ordinate plot for other variables of interest

case rate (as of 5/1) as an independent variable—to importantly control for the fact that observations of large or small testing rates may in fact be due to the level of infection occurring within each neighbourhood—and dummies for each of the three neighbourhood types. Following the approach of Anselin and Rey (2014), estimation began with baseline ordinary least squares (OLS), as shown Equation 1, where y is the dependent variable of interest, **X** is a vector of explanatory variables, and *u* is the error term:

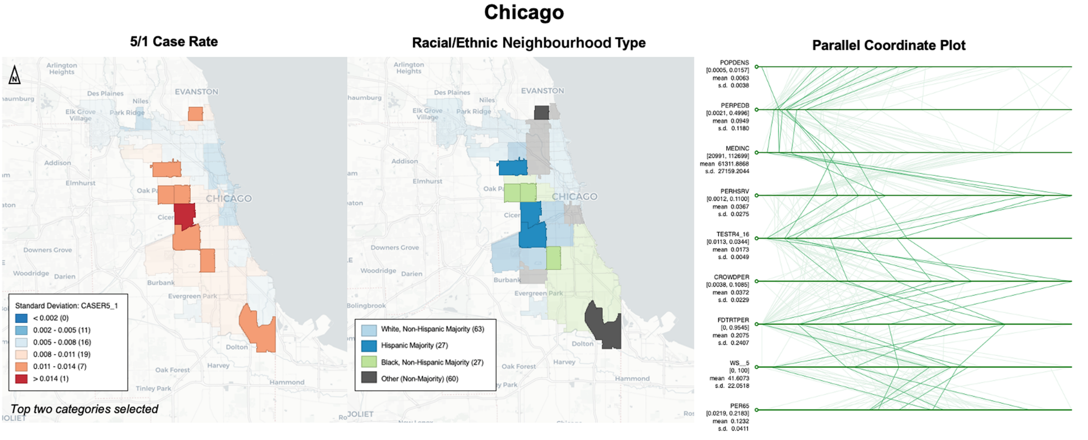

FIGURE 3 Chicago maps comparing COVID-19 case rate and racial/ethnic neighbourhood type, with a parallel co-ordinate plot for other variables of interest

TABLE 4 Comparison of regression results for COVID-19 case rate based on underlying neighbourhood health/vulnerability indicators

DV: Log (Case Rate 5/1)			
Concept	Variable	Chicago	New York City
Socio-economic status	Median household income		–
Healthy, active built environments	Population density (per m^2)	–	
	Percent pedestrian and bike commuters		–
	Hospital accessibility score		
	Percent food desert tracts		
Vulnerability (age, occupation, household structure)	Percent population 65+		
	Percent healthcare service workers		+
	Percent housing units w/ > 1 person per room	+	+
Controls	Testing Rate 4/16	+	+
	Lambda	NA	+
	Pseudo–/Adjusted R^2	0.7309	0.9033
	Observations	54	177
	Robust SE	No	Yes

Note: Due to the difference in sample size and model type between regressions run for each frame of reference, this table displays only the sign of significant (at p ≤ .10) results for coefficients for comparison; full regression results and diagnostics available in the Appendix A.3.

$$y = \mathbf{X}\beta + u. \tag{1}$$

Based on the results of Jarque-Bara tests on the normality of errors from these baseline OLS regressions, the natural log of the testing rate variable was used.[4] In addition, the results of the robust Lagrange multiplier

[4]While the use of the logged variables did not completely eliminate the significance of the Jarque-Bara test in the case of Chicago, it did reduce the value of the statistic.

(LM) tests suggested a different specification for each of the models: for the Chicago model, all of the robust LMs were highly-insignificant, so standard OLS was used. For the New York City model, the robust LM for the spatial error specification suggested the use of a spatial error model. This was estimated using the generalized method of moments (GMM) approach outlined in Anselin and Rey (2014) using a first order queen contiguity spatial weights matrix and the KP HET correction for heteroscedastic errors (based on the suggestions of the results of the Breusch-Pagan and Koenker-Bassett tests). The basic concept of the spatial error model is that the error term in Equation 1, *u*, follows a spatial autoregressive (SAR) process, shown in reduced form in Equation 2, such that:

$$\mathrm{u} = (\mathbf{I} - \lambda \mathbf{W})^{-1} \epsilon, \tag{2}$$

where **I** is an identity matrix and ϵ is a vector of idiosyncratic (i.e., non-spatial) errors and **W** is the appropriate spatial weights matrix. If we substitute this reduced form of the SAR process into the error term in the standard model in Equation 1 and simplify, we can see in Equation 3:

$$(\mathbf{I} - \lambda \mathbf{W})\mathrm{y} = (\mathbf{I} - \lambda \mathbf{W})\mathbf{X}\beta + \epsilon, \tag{3}$$

that we have removed the spatial autocorrelation from the error term by multiplying both sides of the equation by the spatial filter ($\mathbf{I} - \lambda \mathbf{W}$). Thus, with an estimate of the spatial autoregressive parameter λ, Equation 3 can be reduced to a simple linear regression with a resulting “spatially filtered” dependent variable (Anselin & Rey, 2014). This solves problems with the efficiency of the regression estimates due to the observed spatial pattern in the error term based on omitted variables, etc.

Next, a set of models (one for each city) was run using case rate[5] (as of 5/1) as the dependent variable, dummies for each of the three racial/ethnic neighbourhood types, and testing rate (as of 4/16[6]) (to control for the fact that higher observed case rates in a given ZIP code may be due to differential accessibility to testing). In this case, a spatial lag specification was suggested by the robust LMs for the Chicago model, which was estimated using a first order queen contiguity spatial weights matric according to maximum likelihood (ML) approach. The standard spatial lag specification is shown in Equation 4:

$$\mathrm{y} = \rho \mathbf{Wy} + \mathbf{X}\beta + u, \tag{4}$$

where **Wy** is the spatial lag parameter included on the right-hand side of the model to control for spatial dependence in the dependent variable, which, in the case of observed COVID-19 infection rates, also makes sense from a theoretical standpoint, as the disease appears to be very literally spatially contagious.[7]

The final set of models addresses the second research question—exploring the underlying factors driven by observed disparity in infection rates by racial/ethnic neighbourhood type—by again using case rate (5/1)[8] as the dependent variable, and the full suite of independent variables, including population density, percentage

[5]Based on the results of Jarque-Bara tests on the normality of errors from baseline OLS regressions for both models, the natural log of the case rate variable was used.

[6]A lag of 15 days between testing rate and case rate was chosen in order to account for the fact that the relationship between testing rate and observed case rate may be the result of a temporal adjustment process, where, a larger testing effort in a given ZIP code may lead to additional cases being recorded later on, which may lead to increased testing afterwards, etc. Also, introducing a temporal lag in this direction controls for the effects of differential testing accessibility by neighbourhood while mitigating the issues that arise from the fact that testing rate is endogenous to the observed case rate when measured simultaneously (a case can only be observed from a positive test, making case rate a direct function of testing rate, whereas tests can be conducted completely independently of observed cases).

[7]While in general viral infection activity suggests a spatial process that involves direct spillovers to/from neighbouring units, given the fact that we are still in the early stages of understanding the exact behaviour of COVID-19, the approach taken here was to let the data drive the choice of model specification in each case.

[8]The natural log of the case rate was used due to the results of Jarque-Bara tests on the normality of errors from baseline OLS regressions.

pedestrian and bike commuters, hospital accessibility score, percent food desert tracts, percent population 65+, percentage workers in healthcare service occupations, percentage housing units with greater than 1 person per room, and testing rate (as of 4/16). In this case, a spatial error model was suggested by the robust LM for New York City, and thus this model was estimated according to Equations 2 and 3. For the City of Chicago, the robust LMs were again not significant, so standard OLS regression was used for this frame of reference.

3 | RESULTS

3.1 | Study area geography

Before explaining the exploratory spatial analysis results, it is useful to provide a general context for the urban geography of the two selected cities, so that the maps and other results can be interpreted by readers more fully. New York City is the largest city in the US and its single most important cultural and economic hub, with an estimated 2019 population of 8,336,817 (US Census, 2020b). As shown in Figure 1, the city consists of five boroughs and hundreds of individual neighbourhoods, each of which have unique development trajectories, histories, and socio-demographic characteristics. Broadly, the major employment centres in the city are located in Manhattan; Lower Manhattan (or downtown), the first part of the city to be developed, houses the city's financial district (Wall Street), while Midtown Manhattan contains the city's central business district, containing the Empire State Building, Times Square, the Broadway Theater District, Madison Square Garden, the headquarters for many corporations and TV news networks (on Sixth Avenue) and one of the world's most expensive retail districts (on Fifth Avenue) (Wikipedia, 2020). Manhattan is known for its high income, high density residential neighbourhoods (particularly along Central Park), while northern Manhattan comprises the more racially-diverse neighbourhoods of Washington Heights and Harlem, which border the Bronx borough, which also contains some of the lowest income neighbourhoods in the city. However, gentrification has begun in many of these neighbourhoods (e.g., Harlem and

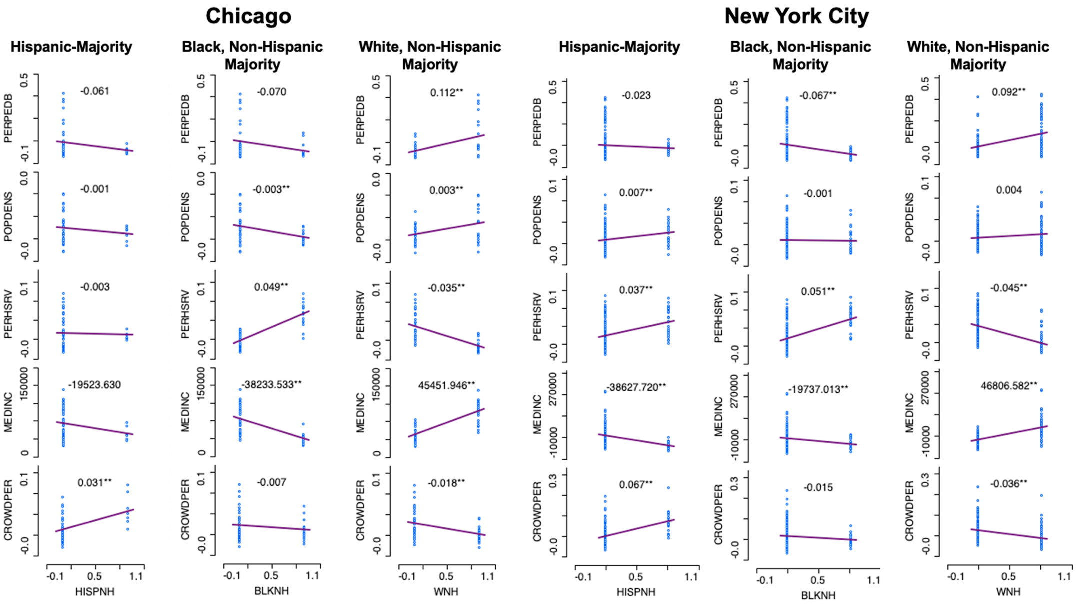

FIGURE 4 Scatterplots comparing racial/ethnic neighbourhood type and various significant indicators of COVID-19 case rate for both Chicago and New York City

Washington Heights), with rising rents and higher proportions of higher income White residents moving in; this has also been true of portions of Brooklyn closest to Manhattan (e.g., Williamsburg) for some time (Kilgannon, 2018). Queens is the most racially-diverse borough, home to many unique immigrant communities and a slightly more suburban development pattern, which is characteristic of Staten Island, the lowest density, least connected, and most suburban of the five boroughs.

Chicago, on the other hand, is the country's third most-populous city at 2,693,976 (US Census, 2020a). It is commonly referred to as one of the most segregated cities in the US, and while the socio-demographic makeup of the city is changing, many of its neighbourhoods are starkly divided along income and racial/ethnic lines (Silver, 2015). Shown in Figure 1, the Loop is the city's central business district and the largest regional employment centre, with a high concentration of financial activity (including the Chicago Mercantile Exchange and Chicago Board of Trade), corporate headquarters (including Boeing and McDonald's, which are both technically outside of the "Loop" neighbourhood but located nearby, just west across the Chicago River), and cultural attractions (Millennium Park and the Chicago Theater, among others). North from the Loop, neighbourhoods along Lake Michigan such as the Gold Coast, Lincoln Park, and Lakeview are traditionally high income, high density, and majority-White; taken very broadly, the city's north side sees the highest concentration of business activity, investment, and deployment of city services. On the other hand, black- and Hispanic-majority neighbourhoods on the city's south and west sides (e.g., Garfield Park, Lawndale, Austin, Woodlawn, and Englewood) have historically experienced high levels of disinvestment and crime. In recent years, the gentrification of centrally-located west and south side neighbourhoods such as Wicker Park, Logan Square, Pilsen, and Bronzeville has also accelerated the concentration of poverty and segregation of these lower income south and west side neighbourhoods, further exacerbating the overlapping economic, social and health disadvantages of these areas.

3.2 | Exploratory spatial analysis

To examine the relationships between COVID-19 case rate, racial/ethnic neighbourhood type, and the other variables of interest in each city, the investigation starts with an exploratory spatial analysis using linked mapping and parallel co-ordinate plots. Figure 2 shows the results of this analysis for New York City—the left panel shows the distribution of observed case rates (as of 5/1) according to a standard deviation classification scheme. ZIP codes in the top two standard categories for case rate (observations two standard deviations from the mean and above) are selected—a selection which is linked in the middle panel, showing racial/ethnic majority neighbourhood type, and the right panel, which shows a parallel co-ordinate plot for other variables of interest. In this way, we can start to understand the "raw" non-parametric relationships between observed case rate, racial/ethnic neighbourhood type, and underlying variables driving any observed racial/ethnic disparity.

As Figure 1 shows, the neighbourhoods with the highest observed case rates are generally located in the Bronx and Queens, including neighbourhoods like the East Bronx, Jackson Heights, and Laurelton. As the middle panel shows, these areas are primarily Hispanic- or Black (non-Hispanic)-majority; in fact, no White non-Hispanic-majority ZIP code can be found within this selection of high case rate outliers. What is driving this uneven distribution? The parallel co-ordinate plot in the right panel shows a strikingly clear pattern: these highest case rate ZIP codes (selected in bright green) overwhelmingly tend to have lower population densities, lower pedestrian/bike commuting, and lower median incomes; they also have somewhat higher proportions of healthcare service workers and testing rate, as well as somewhat middling percentages of crowded housing units (with some high outliers), and a pretty even spread across low, medium, and high values of food desert tracts, hospital accessibility, and percentage population 65 and older. From this analysis, it appears that lower population density, walkability, and socio-economic status, coupled with higher proportions of healthcare service employment, are the variables that primarily link the ZIP codes in the top end of observed COVID-19 case rates, perhaps driving the racial/ethnic disparity displayed in the middle panel.

Interestingly, the pattern for Chicago, shown in Figure , is strikingly similar—ZIP codes 2 standard deviations or higher from the mean in terms of case rate tend to be located on the City's west and south sides (other than the Rogers Park neighbourhood on the north side) in majority-Hispanic and Black (non-Hispanic) neighbourhoods; like New York, no White non-Hispanic-majority ZIP code can be found within this selection of high case rate outliers. Similarly, these neighbourhoods tend to have lower population densities, lower pedestrian/bike commuting, and lower median incomes, as well as higher percentages of healthcare service workers and middling values of food desert tracts, hospital accessibility, and percentage population 65 and older. The main deviations are found in the relatively lower testing rates and higher proportions of crowded housing units for the highest case rate neighbourhoods in Chicago (compared to New York).

3.3 | Regression analysis

While the exploratory analysis seems to point to significant disparity in COVID-19 case rates by racial/ethnic neighbourhood type, it could be the case that the observed case rates are related to the level of testing accessibility in these neighbourhoods. To test this, a regression between neighbourhood type and testing rate—while controlling for observed case rate—is run. As the results in Table show,[9] White non-Hispanic neighbourhoods have a significant positive relationship with testing rate in both cities. At the same time, the results show that Hispanic neighbourhoods have a significant negative relationship with testing rate even after controlling for the fact that these neighbourhoods may have lower case rates than others. This suggests that there may be some systematic disparity in access to testing for Hispanic neighbourhoods, or perhaps a lack of propensity to be tested for residents with uncertain legal immigration status.

Table shows the results of the models predicting COVID-19 case rate by neighbourhood type, while controlling for the testing rate 15 days earlier (to control for the general pattern of differential testing accessibility across neighbourhoods). Again, the results are similar across both cities—Hispanic-majority neighbourhoods have significantly higher case rates and White non-Hispanic-majority neighbourhoods have significantly lower case rates in both cities, while in New York (only) Black non-Hispanic neighbourhoods also have significantly higher case rates. These significant results—even after controlling for the differential testing accessibility suggested above—point to a clear disparity in COVID-19 burden for non-White (non-Hispanic) racial and ethnic groups.

But what factors are primarily driving this disparity? Table shows the results of the third set of models that predict COVID-19 case rate based on a number of underlying vulnerability characteristics (while also controlling for the temporally-lagged testing rate). Again, the results are relatively similar across both cities:[10] measures of healthy, active built environments are significantly *negatively* related to COVID-19 case rate (population density in Chicago and pedestrian/bike commuting percentage in New York), while specific measures of vulnerability (crowded housing percentage in both cities and percentage of healthcare service workers in New York) are significantly positively associated with COVID-19 case rate. In New York, median household income is also significantly negatively related to observed case rates.

These results point to the importance of; (i) neighbourhood physical environments that promote physical activity—even when controlling for other important neighbourhood health factors, such as access to sources of fresh food and healthcare;—and (ii) occupational and household structure, in addition to socio-economic status, as the key mechanisms through which Hispanic- and Black non-Hispanic-majority neighbourhoods have been disproportionately affected by COVID-19. Using scatter plots and simple regression coefficients, Figure examines the direct relationships between racial/ethnic neighbourhood type and the various significant variables demonstrated in the

[9]The results from all three New York models are reported using robust (KP-HET) standard errors.

[10]In addition to the significant results, the direction and sign of non-significant relationships are also quite similar across the two cities, as shown in Appendix A.3.

regression of underlying factors in order to flesh out the direct connections between the significant predictors of COVID-19 case rate and racial/ethnic disparity more specifically.

As Figure shows, in Chicago, Hispanic-majority neighbourhoods tend to demonstrate higher levels of crowding, while Black (non-Hispanic) neighbourhoods tend to have lower population densities, higher proportions of healthcare service workers, and lower median incomes. While White (non-Hispanic) neighbourhoods tend to have higher densities, percentages of pedestrian and bike commuting, and median incomes, as well as lower levels of crowding and healthcare service workers. Thus we can assume that the increased COVID-19 rates observed in Black (non-Hispanic)-majority neighbourhoods may be driven by occupational exposure and less active physical environments leading to increased prevalence of chronic disease (in addition to the more general vulnerability of lower socio-economic status), while in Hispanic-majority neighbourhoods, the primary factor is overcrowded housing (supported by cultural values that encourage multi-generational households), where one infected individual is more likely to infect a larger number of people within their household. The significant display of crowding in Hispanic-majority neighbourhoods is also born out in New York, as are most of the other relationships; the key differences are a relatively higher prevalence of healthcare service workers in Hispanic-majority neighbourhoods, and a bigger difference in pedestrian and bike commuting for White non-Hispanic neighbourhoods compared to the other two groups (rather than population density). Again, this analysis supports the idea that, beyond socio-economic status, active built environments that encourage lower rates of chronic disease, occupational exposure, and crowded housing are the primary drivers of racial/ethnic disparities in observed COVID-19 infection rates at the neighbourhood level.

4 | DISCUSSION AND CONCLUSION

The results of this analysis, while exploratory, provide some useful information about neighbourhood-level disparities in COVID-19 testing rates and how racial/ethnic, socio-economic, occupational, and built environment factors relate to infection rates across the City of Chicago and New York City.

First, there appears to be a disparity in COVID-19 testing accessibility in Hispanic-majority neighbourhoods in both cities, even when controlling for the observed case rates in these neighbourhoods. Whether this represents a systematic disparity in physical (transportation-related) access to testing locations, a propensity to avoid testing for those with uncertain legal immigration status, or other social, cultural, or economic reasons, remains to be seen. However, this observed disparity could have consequences in the future course of the pandemic, particularly because this paper's other results indicate that Hispanic neighbourhoods have also been more heavily affected by the virus (even after controlling for this differential access to testing). These findings could also suggest that actual rates of infection in Hispanic neighbourhoods are higher than reported, creating an important information gap in our understanding of the virus' spread and ultimate toll.

Second, this paper's results show that non-White (non-Hispanic) neighbourhoods have been disproportionately affected by COVID-19, confirming the available individual-level case information, and also pointing to the fundamental importance of social vulnerability factors and social determinants of health in channelling the spread of (even) this novel virus. This is an important finding in-and-of-itself, as these neighbourhoods also tend to be the most lacking in economic and political capital and access to quality healthcare, and thus require additional resources to combat the pandemic in the short term. In the long term, policies to effect significant structural change in these neighbourhoods should be pursued to help majority-minority neighbourhoods (and thus our society as a whole) become more resilient to future pandemics and other public health crises.

While the confirmation of racial/ethnic disparity in COVID-19 burden is important, this paper also attempts to disentangle some of the specific features of neighbourhoods driving this observed disparity in order to help support direct policy recommendations and expand our understanding of the mechanisms through which social determinants of health impact disparate health outcomes. To that end, the results suggest that neighbourhoods with higher levels of occupational exposure (from healthcare service workers), crowded housing (where one infected individual is more likely to infect a greater number of people), and lower density and levels of pedestrian commuting tend to have significantly higher rates of COVID-19 infection. Density and walkability in particular are widely known to be important factors in supporting the underlying healthiness of a neighbourhood and reducing rates of chronic disease (Besser & Dannenberg, 2005; Booth et al., 2005; Cerin et al., 2017; Feng et al., 2010; Frank et al., 2004, 2006; Lindstrom, 2008; Saelens et al., 2003; Wang et al., 2016),

which have been observed to be significant predictors of COVID-19 case severity and death. Interestingly, exercise specifically has been forwarded as one of the primary factors that may tend to decrease COVID-19 case severity (Barney, 2020; Yan, 2020).

Naturally, this analysis also comes with some limitations that could be improved in future work: given that much remains unknown about the COVID-19 virus—particularly the role of asymptomatic transmission, which, by some estimates may be as high as 45% of all actual cases, (Oran & Topol, 2020)—and that the pandemic remains ongoing, it is certainly possible that data collected at a later date will provide a more complete picture of the actual prevalence of COVID-19 infection that could change the interpretation of the results presented here. In addition, there are several factors that have been suggested to be of importance to COVID-19 transmission, including mobility, government interventions such as lockdown and mask orders, and temperature (both as a factor causing people to stay inside and as a physical/biological factor in virus virility) that (while important) cannot be adequately tested here, given the spatial cross-sectional nature of this study (with its focus on evaluating the effect of longer-term structural issues). However, even within this context, improved estimates of healthcare and food accessibility and the underlying prevalence of physical activity (and chronic disease) in future extensions of this work would help to further disentangle the specific effect of particular neighbourhood features on COVID-19 outcomes.

Given these limitations, the results support several direct policy interventions: (i) increased testing efforts in Hispanic-majority neighbourhoods; (ii) increased protective equipment provision, alternative housing provision, and financial assistance to larger or multi-generational households, as well as households with healthcare service workers, to enable those who are infected to safely socially distance from the other members of their household; and, from a long term perspective (perhaps most importantly); (iii) transportation and urban planning policy interventions to create more walkable urban environments that foster lower rates of chronic disease. While many of these interventions—changing zoning requirements, increased infrastructure investment and maintenance, etc.—are long-term, politically-complicated decisions, cities can begin the process immediately by providing more open space and allowing the conversion of some streets to pedestrian/bike-only (or limited-automobile) access in order to encourage safe, socially-distant physical activity, particularly in Hispanic- and Black-majority neighbourhoods. While "shared street" interventions have begun in Chicago and New York during the pandemic, they should be expanded and deployed in non-White neighbourhoods that have traditionally been underserved in terms of healthy, active neighbourhood physical infrastructure (Kamin, 2020).

ORCID

Kevin Credit https://orcid.org/0000-0002-3320-4670

REFERENCES

Anselin, L., & Rey, S. (2014). *Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace, and PySAL*. Chicago, IL: GeoDa Press.

Arias, E., Escobedo, L. A., Kennedy, J., Fu, C., & Cisewski, J. (2018). U.S. small-area life expectancy estimates project: Methodology and results summary. National center for health statistics. *Vital Health Statistics*, *2*(181), 1–45. https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html

Besser, L., & Dannenberg, A. (2005). Walking to Public Transit: Steps to Help Meet Physical Activity Recommendations. *American Journal of Preventive Medicine*, *29*(4), 273–280.

Barney, J. (2020). *Exercise may protect against deadly COVID-19 complication, research suggests, UVA today*. University of Virginia. Retrieved from https://news.virginia.edu/content/exercise-may-protect-against-deadly-covid-19-complication-research-suggests?utm_source=UVAResearchDigest&utm_medium=email&utm_campaign=UVAResearchDigest_04-20

Booth, K. M., Pinkston, M. M., & Poston, W. S. (2005). Obesity and the built environment. *Journal of the Academy of Nutrition and Dietetics*, *105*(5), 110–117.

Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., & Pamuk, E. (2010). Socio-economic disparities in health in the United States: What the patterns tell us. *American Journal of Public Health*, *100*, S186–S196. https://doi.org/10.2105/AJPH.2009.166082

Case, A., & Deaton, A. (2020). *Deaths of despair and the future of capitalism*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctvpr7rb2

Center for Spatial Data Science (CSDS). (2020). *U.S. COVID-19 atlas*. University of Chicago. Retrieved from https://geodacenter.github.io/COVID/

Centers for Disease Control and Prevention (CDC). (2020a). Coronavirus Disease 2019 (COVID-19): How to get tested for current COVID-19 infection. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html

Centers for Disease Control and Prevention (CDC). (2020b). Coronavirus Disease 2019 (COVID-19): People who need extra precautions. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html

Cerin, E., Nathan, A., van Cauwenberg, J., Barnett, D. W., & Barnett, A. (2017). The neighbourhood physical environment and active travel in older adults: A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(15). https://doi.org/10.1186/s12966-017-0471-5

Chicago Department of Health (CDPH). (2020). Latest data. Retrieved from https://www.chicago.gov/city/en/sites/COVID-19/home/latest-data.html

CIA. (2017). The world factbook—country comparison: Life expectancy at birth. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Feng, J., Glass, T. A., Curriero, F. C., Stewart, W. F., & Schwartz, B. S. (2010). The built environment and obesity: A systematic review of the epidemiologic evidence. *Health & Place*, *16*(2), 175–190. https://doi.org/10.1016/j.healthplace.2009.09.008

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A social vulnerability index for disaster management. *Journal of Homeland Security and Emergency Management*, *8*(1), 1–22, Article 3.

Frank, L. D., Andresen, M. A., & Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventative Medicine*, *27*(2), 87–96. https://doi.org/10.1016/j.amepre.2004.04.011

Frank, L. D., Sallis, J. F., Conway, T. L., Chapman, J. E., Saelens, B. E., & Bachman, W. (2006). Many pathways from land use to health: Associations between neighbourhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association*, *72*(1), 75–87. https://doi.org/10.1080/01944360608976725

Goodnough, A., Thomas, K., & Kaplan, S. (2020). Testing falls woefully short as Trump seeks an end to stay-at-home orders. The New York Times, 15 April 2020. Retrieved from https://www.nytimes.com/2020/04/15/us/coronavirus-testing-trump.html

Homeland Infrastructure Foundation-Level Data (HIFLD). (2020). Hospitals [data file]. Retrieved from https://hifld-geoplatform.opendata.arcgis.com/datasets/6ac5e325468c4cb9b905f1728d6fbf0f_0/data?geometry=-90.118%2C41.479%2C-85.801%2C42.195

Illinois Department of Health (IDPH). (2020). Coronavirus disease 2019 (COVID-19): COVID-19 statistics. Retrieved from http://www.dph.illinois.gov/COVID19/COVID19-statistics

Johns Hopkins University (JHU). (2020). COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Retrieved from https://coronavirus.jhu.edu/map.html

Kamin, B. (2020). Column: Chicago's new 'shared street' tilts the balance in favor of pedestrians, bikers and social distancing. But it's tinkering, not structural change. *Chicago Tribune*, 11 June 2010. Retrieved from https://www.chicagotribune.com/columns/blair-kamin/ct-biz-shared-street-kamin-20200611-p5jruo57zjgsbetbaxxqrbleme-story.html

Kilgannon, C. (2018). This space available. *The New York Times*, 6 September 2018. Retrieved from https://www.nytimes.com/interactive/2018/09/06/nyregion/nyc-storefront-vacancy.html

Kolak, M., Bhatt, J., Park, Y. H., Padrón, N. A., & Molefe, A. (2020). Quantification of neighbourhood-level social determinants of health in the Continental United States. *JAMA Network Open*, *3*(1), 1–17. https://doi.org/10.1001/jamanetworkopen.2019.19928

Lindstrom, M. (2008). Means of transportation to work and overweight and obesity: A population-based study in Southern Sweden. *Preventive Medicine*, *46*(1), 22–28. https://doi.org/10.1016/j.ypmed.2007.07.012

National Center for Health Statistics (NCHS). (2015). US small-area life expectancy estimates project (USALEEP). National vital statistics system. Retrieved from https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html

New York City Department of Health (NYCHD). (2020a). *NYC coronavirus (COVID-19) data*. New York City Department of Health. Retrieved from https://github.com/nychealth/coronavirus-data/blob/master/deaths/deaths-by-race-age.csv

New York City Department of Health (NYCHD). (2020b). *NYC coronavirus (COVID-19) data*. New York City Department of Health. Retrieved from https://github.com/nychealth/coronavirus-data/blob/master/deaths/deaths-by-underlying-conditions.csv

New York City Department of Health (NYCHD). (2020c). *NYC coronavirus (COVID-19) data*. New York City Department of Health. Retrieved from https://github.com/nychealth/coronavirus-data/blob/master/README.md

Oppel, R. A., Gebeloff, R., Lai, K. K. R., Wright, W., & Smith, M. (2020). The fullest look yet at the racial inequity of coronavirus. The New York Times, 5 July 2020. Retrieved from https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html

Oran, D. P., & Topol, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection: A narrative review. *Annals of Internal Medicine*. https://doi.org/10.7326/M20-3012

Petrella, D., Munks, J., & Lourgos, A. L. (2020). Illinois expands COVID-19 testing to all and urges those who participated in protests to get tested. *Chicago Tribune*, 4 June 2020. Retrieved from https://www.chicagotribune.com/coronavirus/ct-coronavirus-illinois-testing-20200604-lql2idq3lnhhphokhfa7f3iuca-story.html

Rhone, A., Ver Ploeg, M., Dicken, C., Williams, R., & Breneman, V. (2017). *Low-income and low-supermarket-access census tracts, 2010-2015*. United States Department of Agriculture (USDA) Economic Research Service: Economic Information Bulletin Number 165. Retrieved from https://www.ers.usda.gov/webdocs/publications/82101/eib-165.pdf?v=5021.9

Saelens, B. E., Sallis, J. F., Black, J. B., & Chen, D. (2003). Neighbourhood-based differences in physical activity: An environment scale evaluation. *American Journal of Public Health*, *93*(9), 1552–1558. https://doi.org/10.2105/AJPH.93.9.1552

Silver, N. (2015). The most diverse cities are often the most segregated. FiveThirtyEight, 1 May 2015. Retrieved from https://fivethirtyeight.com/features/the-most-diverse-cities-are-often-the-most-segregated/

Smith, D. M. (1977). *Human geography: A welfare approach*. London: Arnold.

Smith, D. M. (1994). *Geography and social justice*. Oxford: Blackwell.

Solar, O., & Irwin, A. (2010). *A conceptual framework for action on the social determinants of health. Social determinants of health Discussion paper 2 (policy and practice)*. Geneva: World Health Organization (WHO). Retrieved from https://www.who.int/sdhconference/resources/conceptualframeworkforactiononsdh_eng.pdf

United States Census Bureau (US Census). (2020a). *QuickFacts: Chicago city, Illinois*. United States Census Bureau. Retrieved from https://www.census.gov/quickfacts/chicagocityillinois

United States Census Bureau (US Census). (2020b). *QuickFacts: New York city, New York*. United States Census Bureau. Retrieved from https://www.census.gov/quickfacts/newyorkcitynewyork

Wang, Y., Chau, C. K., Ng, W. Y., & Leung, T. M. (2016). A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighbourhoods. *Cities*, *50*, 1–15. https://doi.org/10.1016/j.cities.2015.08.004

Wikipedia. (2020). Midtown Manhattan. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Midtown_Manhattan

Wilkinson, R., & Pickett, K. (2009). *The spirit level: Why greater equality makes societies stronger*. New York: Bloomsbury.

Williams, D. R., Mohammed, S. A., Leavell, J., & Collins, C. (2010). Race, socio-economic status, and health: complexities, ongoing challenges, and research opportunities. *Annals of the New York Academy of Sciences.*, *1186*, 69–101. https://doi.org/10.1111/j.1749-6632.2009.05339.x

Williams, D. R., Priest, N., & Anderson, N. (2016). Understanding associations between race, socio-economic status and health: Patterns and prospects. *Health Psychology*, *35*(4), 407–411. https://doi.org/10.1037/hea0000242

Worldometer. (2020). COVID-19 coronavirus pandemic. Retrieved from https://www.worldometers.info/coronavirus/

Yan, Z. (2020). Extracellular superoxide dismutase, a molecular transducer of health benefits of exercise. *Redox Biology*, *32*. https://doi.org/10.1016/j.redox.2020.101508

How to cite this article: Credit K. Neighbourhood 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;12:1249–1271. https://doi.org/10.1111/rsp3.12321

APPENDIX A1. FULL REGRESSION RESULTS FOR TESTING RATE MODELS

Chicago

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
----------------------------------------
Data set: CHICITY_No0.csv
Weights matrix: File: CHICITY_No0.gal
Dependent Variable: LOGTR5_1 Number of Observations: 54
Mean dependent var: −3.4665 Number of Variables: 5
S.D. dependent var: 0.2761 Degrees of Freedom: 49
R-squared: 0.4562
Adjusted R-squared: 0.4118
Sum squared residual: 2.198 F-statistic: 10.2764
Sigma-square: 0.045 Prob(F-statistic):4.017e-06
S.E. of regression: 0.212 Log likelihood: 9.821
Sigma-square ML: 0.041 Akaike info criterion: −9.641
S.E of regression ML: 0.2017 Schwarz criterion: 0.304

---------------------------------------------------------------------------------
Variable Coefficient Std. Error t-Statistic Probability
---------------------------------------------------------------------------------
CONSTANT −4.0814024 0.1348018−30.2770710 0.0000000
BLKNH 0.1206940 0.0903995 1.3351190 0.1880073
CASER5_1 68.6853062 13.2877291 5.1690779 0.0000043
HISPNH -0.2070316 0.1080295−1.9164351 0.0611521
WNH 0.1611998 0.0946486 1.7031397 0.0948777
---------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 10.174

TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 12.057 0.0024

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 4 11.657 0.0201
Koenker-Bassett test 4 7.119 0.1297
```

REGRESSION

```
DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST MI/DF VALUE PROB
Lagrange Multiplier (lag) 1 0.535 0.4644
Robust LM (lag) 1 0.121 0.7280
Lagrange Multiplier (error) 1 0.415 0.5195
Robust LM (error) 1 0.001 0.9801
Lagrange Multiplier (SARMA) 2 0.536 0.7650
```

New York City

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)
-----------------------------------------------------
Data set: NYC.csv
Weights matrix: File: NYC.gal
Dependent Variable: LOGTR5_1 Number of Observations: 177
Mean dependent var: −3.2222 Number of Variables: 5
S.D. dependent var: 0.3251 Degrees of Freedom: 172
Pseudo R-squared: 0.8853
N. of iterations: 1 Step1c computed: No

-------------------------------------------------------------------------------
Variable Coefficient Std. Error z-Statistic Probability
-------------------------------------------------------------------------------
CONSTANT −4.0206343 0.0416183−96.6074403 0.0000000
BLKNH -0.0212400 0.0225906−0.9402120 0.3471088
CASER5_1 41.8813507 1.9240884 21.7668537 0.0000000
HISPNH -0.0793511 0.0273972−2.8963217 0.0037757
WNH 0.0624073 0.0197107 3.1661676 0.0015446
lambda 0.6735959 0.0494325 13.6265928 0.0000000
-------------------------------------------------------------------------------
```

APPENDIX A2. FULL REGRESSION RESULTS FOR COVID-19 CASE RATE BY RACIAL/ETHNIC NEIGHBOURHOOD TYPE

Chicago

REGRESSION

SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)

Data set: CHICITY_No0.csv
Weights matrix: File: CHICITY_No0.gal
Dependent Variable: LOGCR5_1 Number of Observations: 54
Mean dependent var: −4.9238 Number of Variables: 6
S.D. dependent var: 0.4468 Degrees of Freedom: 48
Pseudo R-squared: 0.6575
Spatial Pseudo R-squared: 0.6595
Sigma-square ML: 0.067 Log likelihood: −4.355
S.E of regression: 0.259 Akaike info criterion: 20.710
Schwarz criterion: 32.644

Variable Coefficient Std. Error z-Statistic Probability

CONSTANT −3.6557985 0.6892146−5.3042966 0.0000001
BLKNH -0.0719104 0.1218240−0.5902806 0.5550026
HISPNH 0.2516226 0.1310711 1.9197412 0.0548906
TESTR4_16 25.3363199 6.7565036 3.7499158 0.0001769
WNH -0.4388833 0.1078513−4.0693391 0.0000471
W_LOGCR5_1 0.3167159 0.1373560 2.3058025 0.0211217

New York City

REGRESSION

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)

Data set: NYC.csv
Weights matrix: File: NYC.gal
Dependent Variable: LOGCR5_1 Number of Observations: 177
Mean dependent var: −4.0561 Number of Variables: 5
S.D. dependent var: 0.4450 Degrees of Freedom: 172
Pseudo R-squared: 0.8025
N. of iterations: 1 Step1c computed: No

Variable Coefficient Std. Error z-Statistic Probability

CONSTANT −5.0220945 0.0752450−66.7432291 0.0000000
BLKNH 0.0687370 0.0327001 2.1020383 0.0355499
HISPNH 0.1538044 0.0315239 4.8789837 0.0000011
TESTR4_16 39.3288712 2.7573458 14.2633077 0.0000000
WNH -0.2002368 0.0334322−5.9893458 0.0000000
lambda 0.7309361 0.0527956 13.8446361 0.0000000

APPENDIX A3. FULL REGRESSION RESULTS FOR COVID-19 CASE RATE BASED ON UNDERLYING NEIGHBOURHOOD HEALTH/VULNERABILITY INDICATORS

Chicago

REGRESSION

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES

Data set: CHICITY_No0.csv
Weights matrix: File: CHICITY_No0.gal
Dependent Variable: LOGCR5_1 Number of Observations: 54
Mean dependent var: −4.9238 Number of Variables: 10
S.D. dependent var: 0.4468 Degrees of Freedom: 44
R-squared: 0.7766
Adjusted R-squared: 0.7309
Sum squared residual: 2.363 F-statistic: 16.9968

(Continues)

REGRESSION
Sigma-square: 0.054 Prob(F-statistic):1.241e-11
S.E. of regression: 0.232 Log likelihood: 7.858
Sigma-square ML: 0.044 Akaike info criterion: 4.284
S.E of regression ML: 0.2092 Schwarz criterion: 24.174

Variable Coefficient Std. Error t-Statistic Probability

CONSTANT −5.1559975 0.2932557−17.5819188 0.0000000
CROWDPER 8.5057481 1.7621848 4.8268196 0.0000170
FDTRTPER 0.2975711 0.2067477 1.4392953 0.1571436
MEDINC -0.0000036 0.0000025−1.4636324 0.1504024
PER65−0.9936237 0.9778590−1.0161216 0.3151265
PERHSRV 0.4826704 2.4362531 0.1981200 0.8438637
PERPEDB -0.4679333 0.4153484−1.1266043 0.2660168
POPDENS -26.6544149 13.9735316−1.9074931 0.0629973
TESTR4_16 22.8765437 7.1424753 3.2028874 0.0025301
WS__5−0.0001233 0.0019478−0.0633175 0.9498004

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 28.953
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 2.551 0.2794
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 9 13.670 0.1345
Koenker-Bassett test 9 13.915 0.1254
DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST MI/DF VALUE PROB
Lagrange Multiplier (lag) 1 0.023 0.8803
Robust LM (lag) 1 0.869 0.3512
Lagrange Multiplier (error) 1 0.419 0.5175
Robust LM (error) 1 1.265 0.2607
Lagrange Multiplier (SARMA) 2 1.288 0.5252

New York City

REGRESSION

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED LEAST SQUARES (HET)

Data set: NYC.csv

Weights matrix: File: NYC.gal

Dependent Variable: LOGCR5_1 Number of Observations: 177

Mean dependent var: −4.0561 Number of Variables: 10

S.D. dependent var: 0.4450 Degrees of Freedom: 167

Pseudo R-squared: 0.9033

N. of iterations: 1 Step1c computed: No

Variable Coefficient Std. Error z-Statistic Probability

CONSTANT −4.9083635 0.1530845−32.0630994 0.0000000

CROWDPER 1.3150069 0.2549007 5.1588989 0.0000002

FDTRTPER −0.0918100 0.1155435−0.7945924 0.4268506

MEDINC -0.0000026 0.0000011−2.3947939 0.0166297

PER65 0.2136352 0.2430527 0.8789665 0.3794194

PERHSRV 1.9074204 0.6186486 3.0832046 0.0020478

PERPEDB -0.5613131 0.2925620−1.9186122 0.0550334

POPDENS -1.0633426 1.3713242−0.7754130 0.4380956

TESTR4_16 35.8122688 2.2028704 16.2570932 0.0000000

WS__5−0.0003411 0.0008807−0.3872736 0.6985537

lambda 0.5473534 0.1059746 5.1649514 0.0000002

DOI: 10.1111/rsp3.12321

Resumen. Este artículo compara datos a nivel de código postal de pruebas y tasas de casos de COVID-19 observados entre la ciudad de Chicago y la ciudad de Nueva York con el fin de comprender mejor tanto: i) el alcance de la existencia de disparidades raciales y étnicas en las pruebas y tasas de casos de COVID-19 a nivel de vecindario; y ii) los factores más importantes de las causas de estas disparidades observadas a nivel de vecindario. A través de un mapeo espacial exploratorio y enfoques econométricos, el artículo encuentra que, en ambas ciudades, los barrios de mayoría hispana tienen tasas de pruebas significativamente más bajas que otros barrios raciales/étnicos, incluso después de haber controlado las tasas de infección observadas, que también son significativamente más altas para los barrios de mayoría hispana. Al mismo tiempo, los barrios de mayoría blanca tienen tasas de pruebas significativamente más altas y tasas observadas de infección más bajas. Dada esta disparidad observada, el artículo examina también una serie de factores subyacentes que potencialmente impulsan las tasas observadas de casos de COVID-19 a nivel de vecindario. Los hallazgos sugieren que el estatus socioeconómico y la provisión de entornos saludables diseñados así activamente están asociados significativamente de forma negativa con las tasas de infección de COVID-19, mientras que varios aspectos de la vulnerabilidad social son predictores positivos significativos de las tasas de infección de COVID-19. Estos hallazgos sugieren que los beneficios para la salud de los entornos construidos con una mayor densidad y transitables a pie pueden desempeñar un mayor papel de “protección” en las tasas de casos de COVID-19 observados a nivel de vecindario de lo que se suponía anteriormente y, al mismo tiempo, pueden indicar que la mayor prevalencia de COVID19 en barrios de mayoría hispana o afroamericana puede deberse en parte a su mayor riesgo de exposición ocupacional y la estructura multigeneracional del hogar (en particular para los barrios de mayoría hispana).

抄録: 本稿では、1)COVID-19の検査率と発症率の人種および民族的な差異はneighbourhood(近隣住区)レベルでどの程度存在するか、と2)1)で認められた差異のneighbourhoodレベルでの最も重要な要因は何かの以上についてよく理解するために、観察下でのCOVID-19検査に関するZIPコードレベルのデータと、シカゴ市とニューヨーク市のCOVID-19発症率との比較を行う。探索的空間マッピングと計量経済学的分析から、いずれの都市でも、ヒスパニック系が多数派を占める地区では、他の人種/民族が多数派の地区よりも検査率が有意に低かった。これは、ヒスパニック系多数派地区で有意に高いことが確認された感染率を調整した場合でも変わらなかった。一方で、白人が多数派を占める地区では、検査率が有意に高く、感染率が低かった。このような差異が認められたため、本稿では、neighbourhoodレベルのCOVID-19発症率を押し上げている可能性のある根底的な要因も検討した。その知見から、社会経済的に高い地位と健康的で活動的な建築環境にあることとCOVID-19の感染率には有意なマイナスの関連性があることが示唆される一方、社会的脆弱性のいくつかの側面がCOVID-19感染率の有意なプラスの予測因子であることが示唆される。これらの知見は、高密度で歩行可能な建築環境から得られる健康上の利益の、neighbourhoodレベルでのCOVID-19発症率に対する保護的役割は、以前に考えられていたよりも大きい可能性があること、同時にヒスパニックと黒人が多数派を占める地域におけるCOVID-19の高い有病率は、部分的には職業曝露と多世代世帯構造　(特にヒスパニック系住民の多い地区)による大きなリスクが原因である可能性があることを示している。