



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

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



# 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

Shima Hamidi<sup>a,\*</sup>, Reid Ewing<sup>b</sup>, Sadegh Sabouri<sup>b</sup>

<sup>a</sup> Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD, 21205, USA

<sup>b</sup> Department of City and Metropolitan Planning, College of Architecture + Planning, University of Utah, 375S 1530 E, Salt Lake City, UT, 84112, USA

## ABSTRACT

This longitudinal study aims to investigate the impacts of development density on the spread and mortality rates of COVID-19 in metropolitan counties in the United States. Multilevel Linear Modeling (MLM) is employed to model the infection rate and the mortality rate of COVID-19, accounting for the hierarchical (two-level) and longitudinal structure of the data. This study finds that large metropolitan size (measured in terms of population) leads to significantly higher COVID-19 infection rates and higher mortality rates. After controlling for metropolitan size and other confounding variables, county density leads to significantly lower infection rates and lower death rates. These findings recommend that urban planners and health professionals continue to advocate for compact development and continue to oppose urban sprawl for this and many other reasons documented in the literature, including the positive relationship between compact development and fitness and general health.

## 1. Introduction

For decades, urban planners have been mostly advocating development of dense, mixed use, walkable and transit accessible community design in compact and polycentric regions due to their environmental, social, economic, and climate change benefits. Compact development has been empirically linked to a decrease in private-vehicle commute distances and times (Ewing et al., 2003; Zolnik, 2011), reduction in teenage driving (McDonald and Trowbridge, 2009), reduction in traffic fatalities (Ewing et al., 2016), greater physical activity and less likelihood of obesity, heart disease, cancer prevalence (Ewing et al., 2014), improved air quality (Stone, 2008), reduction in extreme heat events (Stone et al., 2010), reduction in residential energy use (Ewing and Rong, 2008), greater social capital (Nguyen, 2010), faster emergency response times (Trowbridge et al., 2009), better access to healthy food (Hamidi, 2020), and increased life expectancy (Hamidi et al., 2018). The planning literature suggests that the benefits of compact development far outweigh the costs (such as higher housing prices and more racial segregation) (Ewing and Hamidi, 2015a).

That professional consensus in favor of compact development is challenged by the emergence of COVID-19. This is evident from academic conversations on platforms such as Planetnew, a closed listserv shared by planning faculty in the U.S and internationally, referring to

the recent turn away from the “back to the city” movement, similar to talk about fears of terrorism following 9/11 or of nuclear attack during the Cold War. These concerns are not limited to the academic context, there are already signs of impacts on public opinion and planning practice. News outlets blame density for the rapid spread of COVID-19 in New York City and refer to suburban living as the United States’ “secret weapon” against coronavirus. Accordingly, the State of California’s pioneering state-wide plan for infill and transit-oriented housing development is being increasingly criticized for potentially facilitating the spread of future viruses (Kahn, 2020). These reactions and concerns call for an evidence-based investigation to help planners and policy-makers make more informed decisions that would shape the future of our cities. This is the main purpose of this longitudinal study.

According to a survey in May 2020, about 27% percent of adults and 43% of millennials who were surveyed in the U.S are considering moving homes to suburban and exurban areas. Businesses are following them with the suburban office relocations. The United States Commercial Real Estate Services (CBRE) reports that, in the first quarter of 2020, nine out of 10 largest office markets in the US recorded an increase in downtown vacancy rates of 30 basis points as compared the 10-basis points in suburban areas.

The role of development density on the spread of pandemics is the subject of an ongoing debate. Compact areas facilitate closer human

\* Corresponding author.

E-mail addresses: [shamidi2@jhu.edu](mailto:shamidi2@jhu.edu) (S. Hamidi), [ewing@arch.utah.edu](mailto:ewing@arch.utah.edu) (R. Ewing), [sadegh.sabouri@utah.edu](mailto:sadegh.sabouri@utah.edu) (S. Sabouri).

contacts and could lead to higher exposure to the infection, which makes them the potential epicenter of the pandemic crisis (Glaeser, 2011; Eubank et al., 2004). At the same time, dense areas have superior health and educational systems that are more prepared to handle pandemics, leading to higher rates of recovery and lower rates of mortality (Dye, 2008). Compact areas also have the infrastructure to more effectively put in place measures that foster social distancing, thus reducing actual rates of infection. Density also could make it easier to provide services for citizens in-need at the time of social distancing orders (Bell et al., 2009). Also, it is possible that denser environments make it easier for people to stay somewhat connected with neighbors, families, and friends, while they are sheltering in place.

Equally, the empirical evidence on the relationship between development density and the spread of pandemics is mixed (Chandra et al., 2013; Kao et al., 2012; Garrett, 2010; Chowell et al., 2008; Nishiura and Chowell, 2008; Mills et al., 2004). While a number of studies point to a positive and significant relationship between state-level population density and the death rate from the 1918 Influenza in the U.S (Garrett, 2010), others report no significant association between population density and death rate during the same 1918 pandemic in the U.K (Chowell et al., 2008), Japan (Nishiura and Chowell, 2008) and 45 large U.S cities (Mills et al., 2004).

There exists one major shortcoming that accounts for these mixed and unconvincing findings.

Previous studies fail to control for the key determinants that may confound the relationship between density and pandemic outbreaks. Among confounding factors, research points to demographic variables (Valeri et al., 2016; Lowcock et al., 2012) including gender, men are more likely to be infected by COVID-19 than women, and age, seniors 65+ year-old are more likely to die from COVID-19 than other age groups (Center for Disease Control and Prevention, 2020). Several studies have also related race, ethnicity and income to the morbidity and mortality rates from pandemics. Minority and low-income populations are reported to have higher levels of exposure to the pandemic due to the lack of access to resources that would enable social distancing such as workplace policies, paid sick days and job security (Blumenshine et al., 2008; Kumar et al., 2012; Quinn et al., 2011; Quinn and Kumar 2014). In addition, access to health care facilities for both testing, measured in terms of the primary care physician rate and treatment measured in terms of the number of ICU beds, contribute to the infection and mortality rates of pandemics. Finally, social distancing strategies are one of the most effective ways to slow the spread of pandemics. Such strategies include, but are not limited to, closure of schools, bars, restaurants and social or sporting events, sick leave, work from home policies, splitting of shifts to reduce workplace interactions, shelter-in-place and travel/trade bans (Chu et al., 2017; Katz et al., 2019; Halloran et al., 2008; Blendon et al., 2008).

Another confounding factor could be the larger context of metropolitan areas in which counties are located. Large metropolitan areas tend to have higher peak and average densities due to higher land rents, though with many exceptions. Contrast Boston with Atlanta, San Francisco and Houston. Large metropolitan areas (and mega regions) with a higher number of counties tightly linked together through economic, social and commuting relationships are more likely to exchange tourists, businesspeople, and commuters with each other and with other parts of the world (Neiderud, 2015), thus increasing the risk of cross border infections (Yashima and Sasaki, 2014; Alirol et al., 2011).

This longitudinal study addresses the major gap in the literature by employing a natural experimental research design to investigate the relationship between county density, metropolitan population size and the *spread and deadliness* of COVID-19 in 1165 and 913 U.S. metropolitan counties, respectively, controlling for confounding factors. The COVID-19 pandemic is a perfect case study for this natural experiment.

This novel Coronavirus outbreak was first recorded in Wuhan, China in December 2019 (Cascella et al., 2020). In less than four months, it has spread to more than 212 countries and territories around the world, and

was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 (WHO b, 2020). As of June 3rd, there are more than 1, 841,629 confirmed cases of COVID-19 in the US, with 106,696 confirmed COVID-19 related deaths.

The mortality rate of COVID-19 is about 3.4 percent as compared to 0.1 percent for seasonal flu. It is at least two times more contagious than the flu, and it takes up to 14 days for people with COVID-19 to develop symptoms as compared to 2 days for the flu (Resnick and Animashaun, 2020). Meanwhile, people with the novel coronavirus can unknowingly be infecting others. Finally, its rate of hospitalization is about 19 percent as compared to 2 percent for the flu which makes it a threat to overwhelm the health care system (Alirol et al., 2011).

## 2. Methods

### 2.1. Data and variables

Table 1 presents a list of outcome and independent variables, data sources, and descriptive statistics. All variables were computed for the 1165 metropolitan counties in the sample (for which one or more virus infections have occurred).

The two outcomes variables represent the confirmed infection and mortality rates of COVID-19 per 10,000 population for every day since the first confirmed case by county. They were computed using data from Johns Hopkins which is based on reports from state and local health agencies. An infection is confirmed when an individual tests positive for the virus using an approved test administered and reported by a health professional. A death is confirmed when an individual who has tested positive for the virus dies of a complication or, in some counties, has COVID symptoms and dies of a complication.

While this was the best available time-series data on the infection and mortality counts of COVID-19 at the county level as of May 25, 2020, it is subject to limitations. First, coronavirus tests have become increasingly available, but actual infection rates, particularly in January and February 2020, were grossly underreported over most or all of the U.S. Perhaps rates of undertesting and underreporting were relatively uniform, in which case this fact should create little bias in relative rates. Also, it seems likely that the most severe and life-threatening 20 percent of cases were already being tested and diagnosed, thereby also minimizing bias. Second, as noted by the New York Times in their data collection efforts, due to highly fragmented public health system and variations in reporting methods across the state and local health departments in the U.S, there might be sources of inconsistency in reporting confirmed cases.

The explanatory variable of greatest interest is a measure of county compactness. This study measured compactness somewhat differently than in earlier transportation-related studies (Ewing and Hamidi, 2014, 2015, 2017; Ewing et al., 2018; Hamidi et al., 2015; Hamidi and Ewing, 2015; Hamidi and Zandiatashbar, 2019). In earlier studies, compactness indices (or their opposite, sprawl indices) were a composite of variables measuring four dimensions of the built environment, all related to travel: density, mixed use, activity centering, and street connectivity. The degree of land use mixing or the degree of street connectivity has an obvious link to transportation, but not to the spread of infectious diseases such as COVID-19. So as in one earlier study (Ewing et al., 2018), the measure of the built environment in this study is related to only one of the four compactness dimensions, density. County population and employment were summed and divided by the land area to obtain activity density that accounts for both employment and population concentrations in the county. The county population data came from the American Community Survey (5-year estimates), and the county employment data came from the Longitudinal Employer-Household Dynamics (LEHD) 2017, both data sources released by the U.S. Census.

The models controlled for the degree of flight connectivity. The raw data for passenger (enplanement) rate came from the Air Carrier Activity Information System, a Federal Aviation Administration database

**Table 1**  
Variables, data sources and descriptive statistics (based on sample for the virus infection model)<sup>a</sup>.

Variable	Description	Data Sources	Mean (SD)
<b>Outcome Variables</b>			
	ln of number of deaths per 10,000 (each day)	(Dong et al., 2020) <sup>b</sup>	Varies by day
	ln of number of COVID-19 cases per 10,000 (each day)	(Dong et al., 2020) <sup>b</sup>	Varies by day
<b>Independent Variables Level 1 (Day Level)</b>			
	ln of day 1 (first death) till day X (as of May 25)	(Dong et al., 2020) <sup>b</sup>	3.88 (1.09)
	ln of day 1 (first confirmed case) till day X (as of May 25)	(Dong et al., 2020) <sup>b</sup>	4.21 (0.34)
	shelter in place order (dummy) for each day	New York Times <sup>c</sup>	41.48 (18.40)
	% of staying at home (i.e., no trips with a destination more than one mile away from home) since the first case was confirmed until May 25	(Maryland Transportation, 2020)	Varies by day
<b>Independent Variables Level 2 (County Level)</b>			
	ln of metropolitan population	ACS 5-year estimates (U.S. Census Bureau, 2020a)	13.42 (1.36)
	% of Black population	ACS 5-year estimates (U.S. Census Bureau, 2020a)	11.12 (13.57)
	% of male population	ACS 5-year estimates (U.S. Census Bureau, 2020a)	49.49 (1.74)
	% of population aged 60 and over	ACS 5-year estimates (U.S. Census Bureau, 2020a)	22.26 (4.80)
	% of adults with education beyond high school	ACS 5-year estimates (U.S. Census Bureau, 2020a)	56.43 (10.60)
	enplanements in metropolitan area per 10,000 population	(Federal Aviation Administration, 2018)	19,768 (24,415)
	primary care physicians per 10,000 population	(RWJF, 2020)	6.21 (3.83)
	% of adults currently smoking	(RWJF, 2020)	16.77 (3.20)
	% of adults who are overweight	(RWJF, 2020)	32.11 (5.27)
	ln of activity density (population + employment per square mile)	ACS, 2017; LEHD 2017 (U.S. Census Bureau, 2020b)	5.48 (1.47)
	state-wide number of COVID-19 testing per 10,000 population	The COVID Tracking Project <sup>d</sup>	432.48 (162.55)
	ICU beds per 10,000 population	(Kaiser Health News, 2019)	1.96 (2.12)

<sup>a</sup> Means and standard deviations for level 2 variables are for 1150 counties that make up our dataset for the virus rate equation. Values for the 913 counties that make up the dataset for the death rate equation are available upon request.

<sup>b</sup> <https://github.com/CSSEGISandData/COVID-19>, accessed June 1, 2020.

<sup>c</sup> <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html> accessed June 3, 2020.

<sup>d</sup> The COVID Tracking Project. (2020), Retrieved from <https://covidtracking.com/>.

that contains revenue passenger boarding data. The enplanements at all commercial service airports within a metropolitan area were summed to compute the rate per 10,000 population. This study also accounted for the number of people tested for COVID-19 in each state (and hence, computing testing rate) from the COVID Tracking Project (CTP) website. For most states, the data was obtained directly from state public health authorities. A few states such as California and New York, where the state-wide information was not available, the CTP used other reporting tools such as trusted news sources, interviewing officials, and news conferences. In addition, three variables were borrowed from the County Health Rankings & Roadmaps project developed by the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. The first two variables are the percentage of population currently smoking and the percentage of population who are overweight, based on the Behavioral Risk Factor Surveillance System (BRFSS) from 2017. The third variable, the rate of primary care physicians came from The Area Health Resource File which is a collection of data from more than 50 sources, including: the American Medical Association, American Hospital Association, US Census Bureau, Centers for Medicare & Medicaid Services, Bureau of Labor Statistics, and National Center for Health Statistics (RWJF, 2020).

Other independent variables were recommended by previous empirical research and control for the major contributors to a pandemic outbreak, such as socioeconomic characteristics, healthcare infrastructure, and unhealthy behaviors. The other independent variable of greatest interest is metropolitan size measured in terms of metropolitan population in thousands with data from the American Community Survey 2017 (5-year estimate).

## 2.2. Analytical method

Multilevel Linear Modeling (MLM) was employed for the two longitudinal analyses in this study to model the infection rate and the mortality rate. The MLM in this study accounts for two levels of data structure. Level-1 is the repeated observations (days) within each county in the sample and Level-2 is the county level, with its own set of

variables. The MLM models were estimated using HLM software.<sup>1</sup>

Multilevel modeling has several advantages that make it the best choice for these analyses. First, it accounts for the dependency among individual observations which in this case it would be the observed days

<sup>1</sup> The nesting design of our data set is technically a four-level structure - measurements nested within counties nested within MTAs nested within states. We evaluated the necessity of modeling the data as a four-level model to assess the potential bias in model parameter estimates by comparing relative fit statistics and specific parameter estimates in two- and three-level models. Our primary concern is the introduction of bias in parameter estimates in a two-level model relative to a three or a four-level model (e.g., Maas & Hox, 2005). The primary variable of concerns in this regard are 1) the reliability of the intercepts across levels and 2) the natural log of metropolitan population; this variable is measured at the MSA level and is shared for counties within the same MSA. In a three-level model, this variable is included at level three; in a two-level model, this variable is included at level two. Examining the reliability of the intercepts for the two- and three-level models revealed a substantially smaller reliability coefficient for the three-level model as compared with the two-level model. This result indicates difficulty with estimation of the level three intercept and, by extension, variability about the intercept at level three. This result suggests that the two-level model is preferable to the three-level model. Likewise, parameter estimates for the natural log of metropolitan population in the two and three level models revealed very similar standard errors of the regression coefficient but markedly different regression coefficients. Methodological work on collapsing level and introduction of bias indicates that if bias is introduced to specific regression estimates, it is most of concern in a reduction of the size of the standard error of the regression coefficient in the two-level model relative to the three-level model. In this case, the standard errors of the regression coefficients for the same variable was very similar (.031 vs. .025) in the three-level model compared to the two-level model. However, the regression coefficients are markedly different (.008 vs. .12) across the two models. These findings are consistent with those of the reliability coefficients in indicating that difficulty in estimating level three effects, likely due to the reliability of the level three intercept. For this collection of reasons, we are confident that the results of the two-level model do not include bias due to the two-level structure and that the results of the three-level model should not be interpreted because of model estimation issues.

for each county (Raudenbush and Bryk, 2002). This dependence violates the independence assumption of ordinary least squares (“OLS”) regression. Standard errors of OLS regression coefficients would be underestimated, and OLS coefficient and standard error estimates would be inefficient. Multilevel modeling overcomes these limitations, accounting for the dependence among cases and producing more accurate estimates.

Second, within a multilevel model, each level in the dataset (e.g., repeated observations within counties) is represented by its own sub-model. In other words, MLM can be used to estimate individual growth curves for each county and the regression parameters such as intercepts, slopes or both are treated as random variables to explain the variation at Level-2 (county level) (Kwok et al., 2007; Ewing et al., 2006).

These are known as growth curves because the dependent variable grows over time. Basically, a separate curve was estimated for each county, with a different intercept and slope. The main Level-1 variables were the rates of confirmed infections and deaths, cumulatively, the dependent variables, and the number of days since the first infection or death was reported as the main independent variable. The dependent and independent variables were logged so the resulting curves are power functions. A power function was chosen based on plots of infection and death rates versus days (see Fig. 1).

The infection and death rates increase slowly at first, then accelerate with time, and then level off. Given the initial explosive growth of infections, it would be expected that the number of days since first infection is raised to a power greater than 1.0 for most counties. Deaths do not increase as fast so the exponent of days in this model is less clear. Days were group mean centered, meaning that they were measured as deviations from their group means. Another control variable, included in the models, is a dummy variable equal to 1 if a shelter-in-place order or advisory was in effect in that particular place (mostly that particular state) on that particular day. An alternative to this variable was the percentage of the population staying at home developed and released by the University of Maryland based on location data from multiple sources such as smartphones and vehicle sensors to capture person and vehicle movements (see Table 1). Both variables were tested and showed similar findings in terms of the sign and relative magnitude of the relationships.

Then, at Level-2, the intercepts from Level-1 are modeled in terms of the full range of county-specific variables. County variables were uncentered, as is typically the case in growth curve models. These are random coefficient (random slope) models, as the slopes of the power

functions were found to vary significantly from county to county. To keep the models simple, the authors did not attempt to model variations in slopes (exponents of the power functions) from county to county, only including a random effect term in each Level 2 equation.

The intercepts would be expected to vary positively with some county-level variables, and negatively with others. Of greatest interest to us are the signs and significance levels of two county level variables, the metropolitan population of the county and the activity density of the county. Based on earlier cross-sectional modeling by the authors (Hamidi et al., 2020), the intercepts (group means) would be expected to increase with metropolitan population as connections and interactions increase with the size of the metropolitan area. Counter-intuitively, based on the earlier modeling (Hamidi et al., 2020), the intercept would be expected to be unrelated to activity density in the virus infection rate equation, and be negatively related to activity density in the virus death rate equation. An explanation for these counterintuitive results hinges on the likelihood of greater adherence to social distancing advisories or orders in compact counties (countering other effects of density on infections—see above), and the likelihood of better intensive health care infrastructure in compact counties (again, see above).

### 3. Results and discussion

#### 3.1. Density and the COVID-19 confirmed virus rate

Results for the confirmed virus rate model are presented in Table 2. As expected, the confirmed virus rate increases with time since the first infection in days raised to the 1.39 power. For the period under study, the infection rate increases with time, as expected, and accelerates with time. At some point, the curve levels off and another functional form will be required to represent the growth rate over time in future studies.

Counter-intuitively, the shelter-in-place variable has a positive sign and is highly significant in the virus rate equation. Shelter in place (stay at home) should ultimately flatten the curve of new infections, but in the early days of the virus, it may more of an effect of high infection rates than a cause of lower infection rates. Also at play is a time lag between a shelter-in-place order and a slowing down of infections, and a further time lag between a slowing down of infections and a slowing down of deaths. This again may change as the pandemic progresses. Rates are changing so fast right now (data are updated twice daily) that it is hard to predict when a change in the functional form of the models will be required. The recent lifting of shelter-in-place orders, and the measured

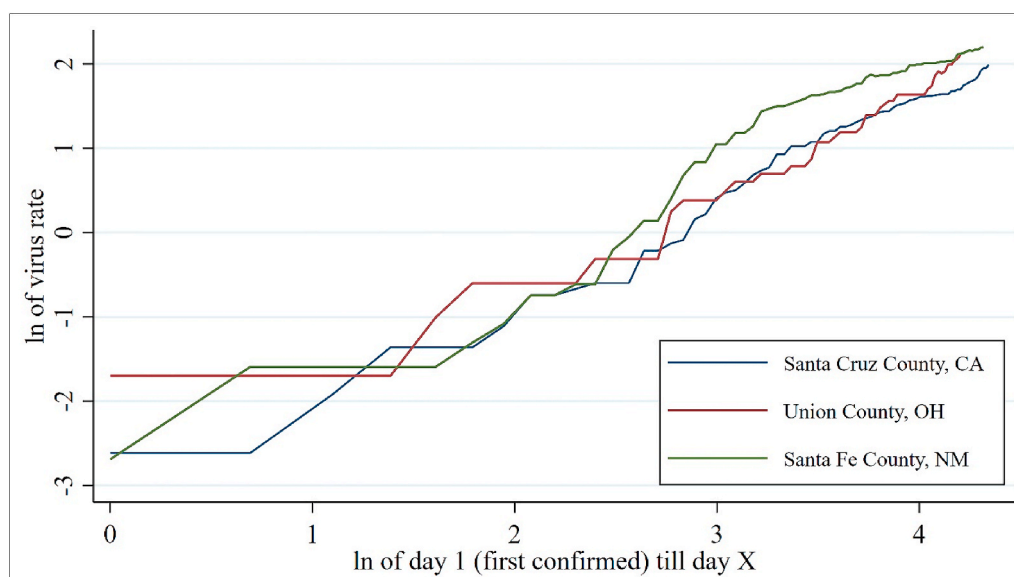


Fig. 1. The natural log of virus rate with respect to the natural log of days (since the first infection) in three metropolitan counties.

**Table 2**

Results of the Random Slope Model (Outcome Variable = Natural Log of Number of Confirmed Infections Per 10,000 population).

Variable	coefficient	standard error	t-ratio	p-value
intercept	-0.578	0.796	-0.73	0.468
ln of metropolitan population	0.122	0.023	5.12	<0.001
ln of activity density	-0.143	0.0241	-5.92	<0.001
enplanements in metropolitan area per 10,000	-1.0E-06	1.0E-06	-1.18	0.240
primary care physicians per 10,000	0.023	0.0086	2.73	0.007
% of population aged 60 and over	0.0078	0.0045	1.73	0.083
% of male population	0.0133	0.0133	1.00	0.318
% of Black population	0.0164	0.0016	10.03	<0.001
% of adults with education beyond high school	-0.0097	0.0028	-3.45	0.001
number of COVID-19 testing per 10,000 population	0.0014	0.0001	11.19	<0.001
ln of day 1 (first confirmed) till day X	1.392	0.0161	86.44	<0.001
shelter in place order (dummy)	0.255	0.0233	10.92	<0.001
<b>Random Effect Parameters</b>				
Random Effect	S.D.	variance component	chi-square	p-value
intercept	0.899	0.810	96259.9	<0.001
ln of day 1 (first confirmed) till day X	0.528	0.279	65982.3	<0.001
shelter in place order (slope)	0.729	0.532	28510.5	<0.001
n	1165			
log likelihood (fit)	-43258.34			
log likelihood (null)	-142842.80			
McFadden R-squared	0.697			

opening up of the economy, make predictions of future trends in infections and deaths especially difficult.

The two independent variables of greatest interest have the expected signs and are highly significant. Larger metropolitan areas have higher infection rates for reasons indicated above. It all has to do with connections and interactions, both of which increase with metropolitan area size. The more interesting result is the negative relationship between the infection rate and the density of population and employment. On a listserv devoted to dialogue among academic urban planners (PLAN-ETNEW), it is universally opined that dense places will experience faster spread of COVID-19, the reason seeming to be that social distancing is harder to maintain in cramped quarters (compact counties). Planners, who have been advocating compact development for decades (over sprawl), are now forced to defend that position. But a high rate of infections in compact counties does not appear to be the case. Perhaps the heightened attention to social distancing requirements (and advisories and orders to that effect) has led to enlightened social distancing behavior and hence lower infection rates. Whatever the reason, this result is different from that found in our earlier cross-sectional analysis of infection rates for a similar sample of metropolitan counties (Hamidi et al., 2020). The earlier study found no relationship between density and the virus rate. This earlier study used a structural equation model to relate exogenous variables (a similar set to this study's independent variables) to infection rates, and then infection rates and exogenous variables to death rates from COVID-19.

As for the control variables in the confirmed virus rate model, most have signs that are consistent with expectations based on common sense, theory, and early reports on the incidence of the virus (CDC, 2020). Most relationships are also highly significant. The infection rate increases with the percentage of the population over 60 years of age, though not quite at a significant level, which also seems reasonable given their weaker immune systems and early reports of higher incidence of the virus among the elderly. The infection rate increases with the statewide testing rate per 10,000 population. Note that a large share of persons with the virus has not been tested due to a shortage of test kits (by some reports as high as 80 percent of those infected have not been confirmed). So states and their counties that do more testing would identify and confirm more cases of the virus. This phenomenon will become less problematic as testing becomes more available. Also, one would assume that a high and fairly uniform percentage of those with severe symptoms

are even now being tested, so this model will understate the virus rate and the coefficients of the independent variables. But it may not affect the sign or significance of the independent variables.

The infection rate is not significantly related with enplanements per capita at regional airports. Global connectivity may determine the advent of the pandemic but, given the number of travelers relative to the total population of the metropolitan area, not the rate of spread after the first case (Neiderud, 2015). The New York Times reports that most early genetic samples of the new coronavirus taken across the country carry distinct mutations that can be traced back to viruses introduced into New York, the most globally connected city in the US and internationally, while only a fraction of them are from a line associated with the outbreak in Washington State. The same article, featuring several studies and interviews with researchers, concluded that while New York has served as a national hub for the COVID-19 spread in the early days, now that it has spread all around the country, it no longer contributes to the progression of the pandemic (Carey and Glanz, 2020). In addition, once travel restrictions were adopted, infections due to international travel may have dropped off. The enplanement rate does not account for these effects.

The infection rate is higher in counties with a higher percentage of African American population (CDC, 2020a). This is consistent with early reports that minorities, particularly African Americans, are hard hit by the virus (CDC, 2020a). The percentage of males in the population has the expected sign (positive) but is not statistically significant. This is also consistent with early reports of higher incidence of the virus among males than females (CDC, 2020a). The lack of significance may be due to the small variance in the percentage of males and females from county to county. Also positive is the relationship between virus infection rates and the rate of primary care physicians per 10,000 population. Most testing is not done by primary care physicians, but they make referrals. The physician rate may be a proxy for the general quality of health care in a county, and hence the rate of testing. Finally, the percentage of adults with education beyond high school in a county is significantly and negatively associated with the infection rates. There may be many reasons for this, but one is the greater ability of white collar workers to work from home.

### 3.2. Density and the COVID-19 mortality rate

Results for the COVID-19 mortality rate model are presented in Table 3. Since you must have the virus to die from it (though there is a small chance that those who die from the virus weren't formally tested and don't show up in confirmed cases), you would expect the mortality rate model to have many of the same variables and relationships as the virus rate model, even after controlling for the virus rate in the mortality rate equation. In our earlier cross-sectional paper (Hamidi et al., 2020), the relationship between the confirmed virus infection rate and the mortality rate (by metropolitan county) was highly significant with a critical ratio of 35.4. In a structural equation model, the virus rate (not unexpectedly) had by far the most significant effect on the mortality rate. It also has the most significant effect on the mortality rate in this study.

There are four important distinctions between the two outcome variables and the corresponding models. First, the mortality rate is subject to smaller errors due to the fact it is not subject to under-testing (undercounting) as is the infection rate. In this respect, the mortality rate is actually a more valid and reliable dependent variable than the confirmed infection rate. Second, and on the other hand, the number of confirmed cases of the virus in the dataset exceeds the number of deaths from the virus by a factor of 16.52 (or the inverse, 0.0604 deaths per infection), and this affects the sampling error. While 1150 counties out of 1165 counties in the sample have one or more infections, only 913 counties have one or more fatalities, and 120 of these have only one fatality (as of May 25). Third, the fatality rate lags behind the virus infection rate by a reported two weeks or more, which also weakens the correlation between the two. Fourth, slightly different sets of variables are logically related to the two different rates. There is no reason, for example, why the infection rate would be related to the number of ICU beds per 10,000 population, but there is every reason that the number of ICU beds would affect the mortality rate. As a result of these factors, the simple correlation between the two outcome variables is only 0.65, which means that the virus rate only explains 42.2 percent of the variation in the mortality rate.

The coefficient of natural logarithm of days since the first death is 0.155, much lower than the comparable coefficient in the confirmed virus rate model. In fact, it suggests that the mortality rate is increasing over time (as it must) but is slowing down as time passes. As noted

above, the mortality rate lags behind the virus rate by a couple of weeks so this result could change by the time this paper is published. This surprising result may reflect the increasing ability to detect and treat infected persons early, and perhaps ply them with more treatment options such as experimental drugs.

The variables of main interest, metropolitan population and activity density of the county, are still significant with the expected signs. Metropolitan population is positively related to the mortality rate (at the 0.001 level) and the density variable is negatively related to the mortality rate (at the 0.073 level). If you leave the virus rate out of the mortality rate equation, the significance of the density variable rises to the 0.001 level because density is inversely related to the virus rate. The common assumption that compact development increases the spread and deadliness of the virus is simply not supported by the evidence.

Most other variables have the same signs as in the virus infection model, and are still significant. This is true of the percentage of adults with education beyond high school (-), the percentage of African American population (+) and the percentage of the population 60 years and over (+). In fact, the significance of the latter has gone way up. The death rate is now strongly related to the percentage 60 years and over. When the elderly get the virus, they are much more likely to die from it than are other age groups.

One variable that was not included in the virus rate model (because it has no obvious theoretical relationship to the infection rate and no statistical relationship) is strongly and significantly related to the death rate is the ICU-bed rate per 10,000 population (-). The ICU-bed rate may be taken literally as a capacity measure for handling COVID-19 infected persons, or may be taken as a proxy measure for the health system's ability to handle critically ill patients. The per capita ICU bed rate is crucial to a successful response to the COVID-19 outbreak and according to recent studies it may at least partially explain the different COVID-19 mortality rates in European countries (Furlong and Hirsch, 2020). Germany, by far, is the best equipped with 2.92 beds per 10,000 population and it also has one of the lowest death rates among European counties. On the other hand, Italy and Spain with only 1.25 and 0.97 ICU beds per 10,000 population have reported some of the highest death tolls in Europe (Rhodes et al., 2012). In the U.S, even though the per capita ICU bed capacity (3.47 ICU beds per 10,000) is higher than its European counterparts, there exist geographic disparities in the distribution of ICU beds across the country. Note that all ICU bed rates are

**Table 3**  
Results of the Random Slop Model (Outcome Variable = Natural Log of Number of Deaths per 10,000 per day).

Variable	coefficient	standard error	t-ratio	p-value
intercept	-1.589	0.402	-3.95	<0.001
ln of metropolitan population	0.121	0.025	4.802	<0.001
ln of activity density	-0.069	0.038	-1.79	0.073
number of ICU beds per 10,000 population	-0.057	0.017	-3.26	0.002
% of adults with education beyond high school	-0.017	0.0037	-4.57	<0.001
% of population aged 60 and over	0.025	0.0064	3.95	<0.001
% of Black population	0.014	0.0022	6.23	<0.001
ln of virus rate	0.783	0.033	23.85	<0.001
ln of day 1 (first death) till day X	0.155	0.023	6.86	<0.001
shelter in place order (dummy)	-0.119	0.017	-6.73	<0.001
<b>Random Effect Parameters</b>				
Random Effect	S.D.	variance component	chi-square	p-value
intercept	1.006	1.012	95762.3	<0.001
shelter in place order (slope)	0.411	0.169	6707.5	<0.001
ln of day 1 (first death) till day X	0.655	0.429	8764.8	<0.001
ln of virus rate	0.905	0.818	13025.6	<0.001
n	913			
log likelihood (fit)	-7967.24			
log likelihood (null)	-65957.88			
McFadden R-squared	0.879			

actual as opposed to estimated ICU bed rates as of 2012). This study confirms that counties with higher per capita ICU capacity are better equipped to handle the pandemic and have reported significantly lower mortality rates.

#### 4. Conclusion

The findings of this study indicate that county density results in a reduction in both the infection rate and death rate of COVID-19 pandemic, after controlling for the larger context of the metropolitan area and other confounding factors. This is in contrast to public assumptions and the U.S news media that have been largely based on the simple correlation between density and the infection rates of COVID-19 and also based on the observations about the spread and severity of the COVID-19 pandemic in New York City. These findings suggest that the relationship between density and the pandemic is more complex than a simple correlation. Density increases the concentration of people and facilitates person-to-person contacts, but density also could lead to better health care infrastructure that is more prepared to respond to pandemics. More research is needed to investigate further the impacts of density on the success of public health measures in fighting pandemics.

It may still be too early to draw a definitive conclusion since the COVID-19 pandemic is rapidly evolving at the time of this writing. However, the findings of this study put a pause on the current and emerging practical impacts of the COVID-19 pandemic on the future landscape of community design and planning in the U.S and internationally. The concerns about density being a risk factor or an “enemy” that could cause emerging pandemics to become more severe, is dominating conversations in urban planning, epidemiology and public health. These concerns could lead to a population shift from urban centers to suburban and exurban areas. This is evident from California’s statewide legislation (AB-7100) that promotes infill and transit-oriented development. It has been gaining more opposition recently. This study offers a new perspective in this dialogue. This study calls for the continued support for compact and walkable communities while also recommends that decisionmakers be proactively cautious about the vulnerability of large metropolitan areas during the pandemics.

#### Research support

This research was supported by the Bloomberg American Health Initiative at the Johns Hopkins Bloomberg School of Public Health.

#### References

Alirol, E., Getaz, L., Stoll, B., Chappuis, F., Loutan, L., 2011. Urbanization and infectious diseases in a globalized world. *Lancet Infect. Dis.* 11 (2), 131–141.

Bell, D.M., Weisfuse, I.B., Hernandez-Avila, M., Del Rio, C., Bustamante, X., Rodier, G., 2009. Pandemic influenza as 21st century urban public health crisis. *Emerg. Infect. Dis.* 15 (12), 1963.

Blendon, R.J., Koonin, L.M., Benson, J.M., Cetron, M.S., Pollard, W.E., Mitchell, E.W., et al., 2008. Public response to community mitigation measures for pandemic influenza. *Emerg. Infect. Dis.* 14 (5), 778.

Blumenshine, P., Reingold, A., Egerter, S., Mockenhaupt, R., Braveman, P., Marks, J., 2008. Pandemic influenza planning in the United States from a health disparities perspective. *Emerg. Infect. Dis.* 14 (5), 709.

Carey, B., Glanz, J., 2020. Travel from New York City Seeded Wave of U.S. Outbreaks. *The New York Times*. Accessed on. <https://www.nytimes.com/2020/05/07/us/new-york-city-coronavirus-outbreak.html>. (Accessed 30 May 2020).

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C., Di Napoli, R., 2020. Features, evaluation and treatment coronavirus (COVID-19) [Internet]. In: StatPearls Publishing.

Center for Disease Control and Prevention, 2020a. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *Morbidity and Mortality Weekly Report*. <https://doi.org/10.15585/mmwr.mm6912e2>.

Centers for Disease Control and Prevention, 2020b. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html>.

Centers for Disease Control and Prevention, 2020c. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. <https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm>.

Chandra, S., Kassens-Noor, E., Kuljanin, G., Vertalka, J., 2013. A geographic analysis of population density thresholds in the influenza pandemic of 1918–19. *Int. J. Health Geogr.* 12 (1), 9.

Chowell, G., Bettencourt, L.M., Johnson, N., Alonso, W.J., Viboud, C., 2008. The 1918–1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact. *Proc. Biol. Sci.* 275 (1634), 501–509.

Chu, S., Swarup, S., Chen, J., Marathe, A., 2017. A comparison of targeted layered containment strategies for a flu pandemic in three US cities. In: *Proceedings of the 16th Conference on Autonomous Agents and Multi Agent Systems*, pp. 1499–1501.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20 (5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).

Dye, C., 2008. Health and urban living. *Science* 319 (5864), 766–769.

Eubank, S., Guclu, H., Kumar, V.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429 (6988), 180–184.

Ewing, R., Hamidi, S., 2014. *Measuring Urban Sprawl and Validating Sprawl Measures*. National Institutes of Health and Smart Growth America, Washington, DC.

Ewing, R., Hamidi, S., 2015a. Compactness versus sprawl: a review of recent evidence from the United States. *J. Plann. Lit.* 30 (4), 413–432.

Ewing, R., Hamidi, S., 2015b. Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities – update and refinement. *Transport. Res. Rec.: Journal of the Transportation Research Board* (2513), 40–47.

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

Ewing, R., Rong, F., 2008. The impact of urban form on US residential energy use. *Housing Policy Debate* 19 (1), 1–30.

Ewing, R., Pendall, R., Chen, D., 2003. Measuring sprawl and its transportation impacts. *Transport. Res. Rec.: Journal of the Transportation Research Board* 1831 (1), 175–183.

Ewing, R., Brownson, R.C., Berrigan, D., 2006. Relationship between urban sprawl and weight of United States youth. *Am. J. Prev. Med.* 31 (6), 464–474.

Ewing, R., Hamidi, S., Tian, G., Proffitt, D., Tonin, S., Fregolent, L., 2018. Testing Newman and Kenworthy’s theory of density and automobile dependence. *J. Plann. Educ. Res.* 38 (2), 167–182.

Ewing, R., Meakins, G., Hamidi, S., Nelson, A.C., 2014. Relationship between urban sprawl and physical activity, obesity, and morbidity—Update and refinement. *Health Place* 26, 118–126.

Ewing, R., Hamidi, S., Grace, J.B., 2016. Urban sprawl as a risk factor in motor vehicle crashes. *Urban Stud.* 53 (2), 247–266.

Federal Aviation Administration, 2018. *Enplanements at All Commercial Service Airports (by Rank)*. Retrieved from. [https://www.faa.gov/airports/planning\\_capacity/passenger\\_allcargo\\_stats/passenger/](https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/).

Furlong, A., Hirsch, C., 2020. Charting Europe’s capacity to deal with the coronavirus crisis. *Politico*. Accessed. <https://www.politico.eu/article/charting-europes-capacity-to-deal-with-the-coronavirus-crisis/>. (Accessed 30 May 2020).

Garrett, T.A., 2010. Economic Effects of the 1918 Influenza Pandemic: Implications for a Modern-Day Pandemic. Working paper CA0721. 2007. Federal Reserve Bank of St. Louis. Accessed. [https://www.stlouisfed.org/~media/files/pdfs/community-development/research-reports/pandemic\\_flu\\_report.pdf](https://www.stlouisfed.org/~media/files/pdfs/community-development/research-reports/pandemic_flu_report.pdf). (Accessed 23 March 2020).

Glaeser, E., 2011. Cities, productivity, and quality of life. *Science* 333 (6042), 592–594.

Halloran, M.E., Ferguson, N.M., Eubank, S., Longini, I.M., Cummings, D.A., Lewis, B., et al., 2008. Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci. Unit. States Am.* 105 (12), 4639–4644.

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

Hamidi, S., Ewing, R., 2015. Is sprawl affordable for Americans? Exploring the association between housing and transportation affordability and urban sprawl. *Transport. Res. Rec.* 2500 (1), 75–79. <https://doi.org/10.3141/2500-09>.

Hamidi, S., Ewing, R., Preuss, L., Dodds, A., 2015. Measuring sprawl and its impacts: an update. *J. Plann. Educ. Res.* 35 (1), 35–50.

Hamidi, S., Sabouri, S., Ewing, R., 2020. Does Density Aggravate the COVID-19 Pandemic? Early Findings and Lessons for Planners? *J. Am. Plann. Assoc.* <https://doi.org/10.1080/01944363.2020.1777891>.

Hamidi, S., Ewing, R., Tatalovich, Z., Grace, J.B., Berrigan, D., 2018. Associations between urban sprawl and life expectancy in the United States. *Int. J. Environ. Res. Publ. Health* 15 (5), 861.

Hamidi, S., Zandiatashbar, A., 2019. Does urban form matter for innovation productivity? A national multi-level study of the association between neighbourhood innovation capacity and urban sprawl. *Urban Stud.* 56 (8), 1576–1594. <https://doi.org/10.1177/0042098018767002>.

Kahn, D., 2020. California Saw Dense Housing Near Transit as Its Future. What Now? *Politico* retrieved from. <https://www.politico.com/states/california/story/2020/03/27/california-saw-dense-housing-near-transit-as-its-future-what-now-1269263>.

Kao, C.L., Chan, T.C., Tsai, C.H., Chu, K.Y., Chuang, S.F., Lee, C.C., et al., 2012. Emerged HA and NA mutants of the pandemic influenza H1N1 viruses with increasing epidemiological significance in Taipei and Kaohsiung, Taiwan, 2009–10. *PLoS One* 7 (2).

Kaiser Health News, 2019. Millions of older Americans live in counties with no ICU beds as pandemic intensifies: Analysis of hospital cost reports filed to the Centers for Medicare & Medicaid Services. <https://khn.org/news/as-coronavirus-spreads-wide-ly-millions-of-older-americans-live-in-counties-with-no-icu-beds/>.

Katz, R., Vaught, A., Simmens, S.J., 2019. Local decision making for implementing social distancing in response to outbreaks. *Publ. Health Rep.* 134 (2), 150–154.

Kumar, S., Quinn, S.C., Kim, K.H., Daniel, L.H., Freimuth, V.S., 2012. The impact of workplace policies and other social factors on self-reported influenza-like illness incidence during the 2009 H1N1 pandemic. *Am. J. Publ. Health* 102 (1), 134–140.



- Kwok, O.M., West, S.G., Green, S.B., 2007. The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: a Monte Carlo study. *Multivariate Behav. Res.* 42 (3), 557–592.
- Lowcock, E.C., Rosella, L.C., Foisy, J., McGeer, A., Crowcroft, N., 2012. The social determinants of health and pandemic H1N1 2009 influenza severity. *Am. J. Publ. Health* 102 (8), e51–e58.
- Maas, C.J., Hox, J.J., 2005. Sufficient sample sizes for multilevel modeling. *Methodology* 1 (3), 86–92.
- Maryland Transportation Institute, 2020. University of Maryland COVID-19 impact analysis platform. <https://data.covid.umd.edu>.
- McDonald, N., Trowbridge, M., 2009. Does the built environment affect when American teens become drivers? Evidence from the 2001 National Household Travel Survey. *J. Saf. Res.* 40 (3), 177–183.
- Mills, C.E., Robins, J.M., Lipsitch, M., 2004. Transmissibility of 1918 pandemic influenza. *Nature* 432 (7019), 904–906.
- Neiderud, C.J., 2015. How urbanization affects the epidemiology of emerging infectious diseases. *Infect. Ecol. Epidemiol.* 5 (1), 27060.
- Nguyen, D., 2010. Evidence of the impacts of urban sprawl on social capital. *Environ. Plann. Plann. Des.* 37 (4), 610.
- Nishiura, H., Chowell, G., 2008. Rurality and Pandemic Influenza: Geographic Heterogeneity in the Risks of Infection and Death in Kanagawa, Japan (1918–1919).
- Quinn, S.C., Kumar, S., 2014. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur. Bioterrorism Biodefense Strategy, Pract. Sci.* 12 (5), 263–273.
- Quinn, S.C., Kumar, S., Freimuth, V.S., Musa, D., Casteneda-Angarita, N., Kidwell, K., 2011. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am. J. Publ. Health* 101 (2), 285–293.
- Raudenbush, S.W., Bryk, A.S., 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, vol. 1. Sage Publication.
- Resnick, B., Animashaun, C., March 18, 2020. Why Covid-19 is worse than the flu, in one chart. *VOX*. Accessed. <https://www.vox.com/science-and-health/2020/3/18/21184992/coronavirus-covid-19-flu-comparison-chart>. (Accessed 23 March 2020).
- Rhodes, A., Ferdinande, P., Flaatten, H., Guidet, B., Metnitz, P.G., Moreno, R.P., 2012. The variability of critical care bed numbers in Europe. *Intensive Care Med.* 38 (10), 1647–1653.
- Robert Wood Johnson Foundation, 2020. Community Health Rankings Project. Retrieved from. <https://www.countyhealthrankings.org/explore-health-rankings/measure-s-data-sources/2020-measures>.
- Stone Jr., B., 2008. Urban sprawl and air quality in large US cities. *J. Environ. Manag.* 86 (4), 688–698.
- Stone, B., Hess, J.J., Frumkin, H., 2010. Urban form and extreme heat events: are sprawling cities more vulnerable to climate change than compact cities? *Environ. Health Perspect.* 118 (10), 1425.
- Trowbridge, M.J., Gurka, M.J., O'connor, R.E., 2009. Urban sprawl and delayed ambulance arrival in the US. *Am. J. Prev. Med.* 37 (5), 428.
- U.S. Census Bureau, 2020a. American community survey 5- year estimates, 2013–2017 tables. <https://data.census.gov/cedsci>.
- U.S. Census Bureau, 2020b. LEHD origin-destination employment statistics data (2002–2017). <https://lehd.ces.census.gov/data/#lodes>.
- Valeri, L., Patterson-Lomba, O., Gurmu, Y., Ablrh, A., Bobb, J., Townes, F.W., Harling, G., 2016. Predicting subnational ebola virus disease epidemic dynamics from sociodemographic indicators. *PLoS One* 11 (10).
- World Health Organization, 2020. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 – 11 March 2020. WHO, Geneva. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>.
- Yashima, K., Sasaki, A., 2014. Epidemic process over the commute network in a metropolitan area. *PLoS One* 9 (6), e98518.
- Zolnik, E.J., 2011. The effect of sprawl on private-vehicle commuting outcomes. *Environment and Planning-Part A* 43 (8), 1875.