On the Spatial Structure of the Covid-19 Pandemic: Some Scale and Density Effects

A Research Note

Solly Angel, 14 June 2020

Abstract

The key finding reported here is that when we account for difference in city populations, average population density does not affect the number of Covid-19 cases and deaths in cities. On the contrary, in two cities with the same population, the city with the higher average population density can be expected to have less, not more, confirmed cases and deaths from the Covid-19 pandemic. Data on the total number of confirmed Covid-19 cases and deaths in U.S. metropolitan statistical areas (MSAs) on 7 June 2020 shows a *superlinear* relationship with their populations. Namely, a city with double the population of another one can be expected to have more than double the number of cases and deaths. This relationship is shown to hold in the case of the Covid-19 pandemic. We also know that larger cities have higher population densities, namely that average population density in cities is *superlinear* as well: A city with double the population of another one can be expected to have a density that is some 15% higher than the smaller one. Multiple regression models that use Covid-19 cases and death as dependent variables and metropolitan population and density as independent ones show a positive and significant effect of population and a negative and significant effect of density. Namely, when we account for the population effect on both cases and deaths and population density, the density effect becomes negative. This is important because there is an unfortunate tendency to 'blame' high urban densities on for the spread of the pandemic.

On the Scale of Social Interaction in Cities:

Geoffrey West, in his book *Scale: The Universal laws of Life, Growth and Deaths in organisms, Cities and Companies* (Penguin, 2017) suggests that the fundamental objective of cities is to facilitate physical contact among people:

> All socioeconomic activity in cities involves the interaction between people. Employment, wealth creation, innovation and ideas, the spread of infectious diseases, health care, crime, policing, education, entertainment, and indeed all of the pursuits that characterize modern *Homo Sapiens* and are emblematic of urban life are sustained and generated by the continual

> exchange of information, goods and money *between* people. The job of the city is to facilitate and enhance this process by providing the appropriate infrastructure such as parks, restaurants, cafés, sports stadiums, cinemas, theaters, public squares, plazas, offices buildings, and meeting halls to encourage and increase social connectivity (316).

West introduces us to Dunbar's Number, the number of people than an average individual can keep in touch with, estimated to be of the order of 150.[1] He points out that there is a hierarchy in these relationships: a few are more intense and more frequent than others and the intensity and frequency decreases as the number of acquaintances increases.

West provides many examples where the relationship between, say, infrastructure and energy use in cities is sublinear: a doubling of city size is associated with a 15% decline in infrastructure and energy use. In parallel, he postulates that, like economic activity or innovation, the spread of infectious disease—in our case, the spread of the Covid-19 pandemic—is *superlinear* with the population size of cities: a doubling of city size is associated with a 15% increase in cases or deaths from infectious disease. To quote:

> *The sublinearity of infrastructure and energy use is the exact inverse of the superlinearity of socioeconomic activity.* Consequently, the same 15 percent degree, the bigger the city the *more* each person earns, creates, innovates, and interacts—and the more each person experiences crime, disease, entertainment, and opportunity—and all of this at a cost that requires less infrastructure and energy for each of them. This is the genius of the city. No wonder so many people are drawn to them (323).

On the *Superlinearity* of Covid-19 Cases and Deaths:

How does the current pandemic relate to the population size of cities and metropolitan areas? The scaling laws formulated by West would suggest that larger cities—where more people interact with each other—will have more than their share of Covid-19 cases and deaths.

In this note, we focus on the relationship between Covid-19 cases and deaths and the populations of U.S. cities and metropolitan areas. We can examine this relationship using population and built-up area data for the 384 U.S. Metropolitan Statistical Areas (MSAs), defined by the U.S. Census Bureau and consisting of largely contiguous built-up areas in counties that are linked together by commuting patterns.

---

1 Dunbar, R.I.M., 1993. "Coevolution of neocortical size, group size and language in humans". *Behavioral and Brain Sciences*. 16 (4): 681–735.

We now focus on the postulated *superlinear* relationships between the total number of Covid-19 cases and deaths in U.S. MSAs as of 7 June 2020. Our first hypothesis is:

> Hypothesis I: The total number of confirmed Covid-19 cases in U.S. metropolitan areas is *superlinear* with their populations. More specifically, this number follows established scaling laws: A city with double the population of another one will have 1.15 times the number of confirmed cases per capita of the smaller city.

The relationship between the total population and the total number of confirmed cases of Covid-19 by 7 June 2020 in the 384 MSAs in logarithmic form is shown in figure 1 below.[2] The relationship is *superlinear* and highly significant ($R^2 = 0.63$). The slope of the Power Curve is 1.09 (which is higher than 1 but not as high as the postulated 1.15) and thus the relationship is *superlinear*. This translates to the observation that the expected number of confirmed cases per capita in a city with double the population of a smaller one will be 6.5% higher than that of the smaller city. In other words, larger cities have more than their share of confirmed Covid-19 cases than would be expected if cases were evenly distributed among the U.S. cities or among the U.S. as a whole. We can conclude, therefore that the first part of Hypothesis I is confirmed. The number of confirmed cases in MSAs is indeed *superlinear* with their populations. The second part of the hypothesis is only partially confirmed. The number of cases per capita does not increase by the expected 15% when city populations double but only by 6.5%.

---

2 A similar result was obtained by Stier, A., Berman, M. and Bettencourt, L., 2020. COVID-19 Attack Rate Increases with City Size, *Mansueto Institute for Urban Innovation Research Paper No. 19*, 31 March, Last revised: 9 June.

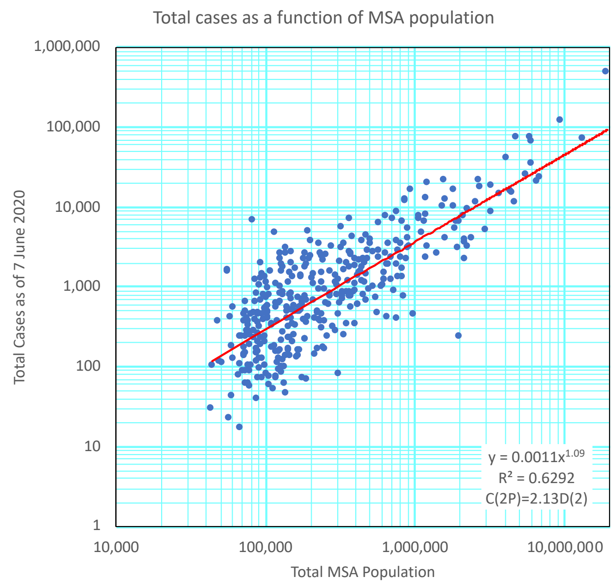

Figure 1: The total number of confirmed cases of Covid-19 on 7 June 2020 in MSAs in the United States as a function of their population.

Our second hypothesis is:

> Hypothesis II: The total number of Covid-19 deaths in U.S. metropolitan areas is *superlinear* with their populations. More specifically, this number follows established scaling laws: A city with double the population of another one will have 1.15 times the number of deaths per capita of the smaller city.

The relationship between the total population and the total number of deaths from Covid-19 by 7 June 2020 in the 384 MSAs in logarithmic form is shown in figure 2 below. The relationship is *superlinear* and highly significant (R^2 = 0.52). The slope of the Power Curve is 1.16 (which is higher than 1 and very close to the postulated 1.15) and the relationship is clearly *superlinear*. This translates to the observation that the expected number of deaths per capita from Covid-19 in a city with double the population of a smaller one will be 12% higher than that of the smaller city. In other words, larger cities have more than their share of Covid-19 deaths than would be expected if deaths were evenly distributed among the U.S. cities or among the U.S. as a whole. We can conclude, therefore that the first part of Hypothesis II is confirmed. The number of confirmed cases in MSAs is indeed *superlinear* with their populations. The second part of the hypothesis is also confirmed. The number of deaths per capita increases almost by the expected 15% when city populations double.

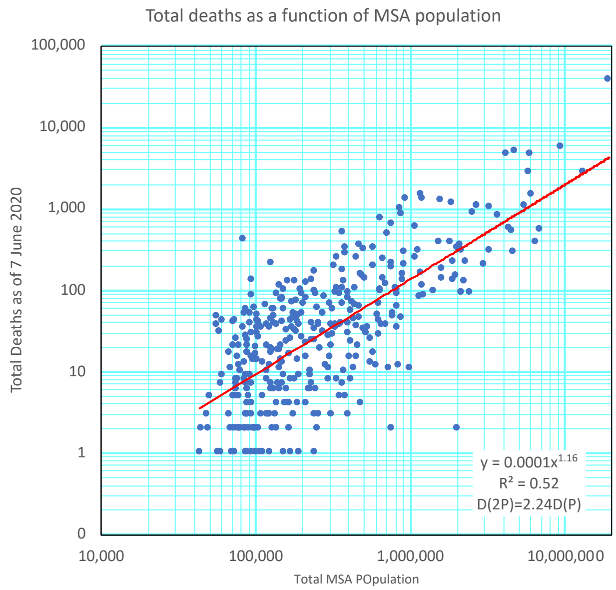

Figure 2: The total number of deaths from Covid-19 on 7 June 2020 in MSAs in the United States as a function of their population.

Comparing the results for Covid-19 cases and deaths for U.S. metropolitan areas, we observe that deaths are more *superlinear* than cases. Why that would be the case is unclear.

On the Relationship of the Pandemic to Average Population Densities:

Much of the discussion on the spread of the virus has focused on 'density', however defined, rather than on the populations of urban areas. We begin by noting that the total population of a city and its average density—defined as the ratio of its total population and its 'urbanized area (essentially its built-up area)—are not independent of each other: Other things being equal, when the population of a metropolitan area increases, its average population density increases in accordance West's scaling law. In other words, we need less than double the built-up area to accommodate a city with double the population.

The relationship between the population and built-up areas of the 384 MSAs in logarithmic form is shown in figure 3 below. As expected, the relationship is *sublinear* and highly significant ($R^2 = 0.93$). The slope of the Power Curve is 0.89, (which is less than 1 but not as low as the expected 0.85) and thus the relationship is *sublinear*. This translates to the observation that a city with double the population of a smaller one will have an area which is only 1.85 times the area of the smaller city.

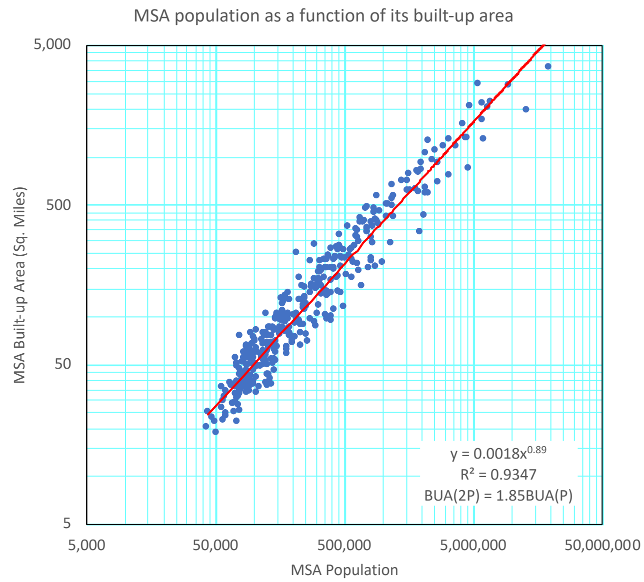

Figure 3: The built-up area of MSAs in the United States as a function of their population.

If we define the average population density of a city thus:

(1) Density = Population ÷ Built-up Area,

we can see than density increases with city population size. In fact, for a city with double the size of a smaller one it increases by a factor of 1.08 (which is more than 1 but not as high as 1.15). In other words, the density of cities is at least partially explained by their population size. Larger cities are more efficient in their use of land and therefore have higher average population densities.

Given the findings of the previous section, what is the expected relationship between Covid-19 cases and deaths and the average population densities of cities? We already know that all three increase when city populations increase. Clearly, given the result shown in figure 3, we can expect densities to be higher in larger cities. We know that both cases and deaths per capita are higher in larger cities as well. We can therefore expect them to be higher in higher-density cities as well. But that would not necessarily be a 'density effect'. It could be a 'population effect'. We can examine this question with multiple regression models that posit either the total number of cases or deaths in logarithmic form as dependent variables, and both population and density of American MSAs in logarithmic form as independent variables. If both the population and density coefficients are positive and significant, we would conclude that the number of Covid-19 cases and deaths are subject both population and density effects. Hypothesis III and IV below postulate this to be true:

> Hypothesis III: The total number of Covid-19 cases in U.S. metropolitan areas is *superlinear* with both their populations and their densities. Larger and denser cities will have more cases per capita than smaller ones.

Table 1 below presents the results of modeling the total number of cases on 7 June 2020 in U.S. metropolitan areas as a function of their populations and average densities. The model is robust (Adjusted R= 0.63) and all the coefficient are statistically significant (*p*-values<0.05). The coefficient of the population is 1.13 (which is higher than 1 and very close to the postulated 1.15) and the relationship is clearly *superlinear*. This translates to the observation that the expected number of cases per capita from Covid-19 in a city with double the population of a smaller one will be 9.5% higher than that of the smaller city.

Surprisingly, however, the coefficient of density in the model is negative. In other words, in a city with a given population, a higher density will translate into a smaller total number of cases. For example, other things being equal, a 10% increase in density will result in a 3% *decline* in the total number of cases. In fact, when we consider the effects of both population and density on the total number of cases, the population effect becomes stronger than the effect observed when density is not part of the model. Including density in the model strengthens the population effect. In light of this finding, we must reject Hypothesis III: The total number of cases in U.S. metropolitan area is not *superlinear* with respect to density. In fact, for a given population, it decreases when density increases.

Variable	Coefficient	*p*-value
Intercept	-4.55	0.00036
Log MSA Population	1.13	0
Log MSA Density	-0.36	0.04574
Observations	383	
Adjusted R-Squared	0.63	

Table 1: Modeling the total number of cases in U.S. metropolitan areas as a function of their populations and densities.

Similar results are observed in modeling the total number of deaths:

> Hypothesis IV: The total number of Covid-19 deaths in U.S. metropolitan areas is *superlinear* with both their populations and their densities. Larger and denser cities will have more cases per capita than smaller ones.

Table 2 below presents the results of modeling the total number of deaths on 7 June 2020 in U.S. metropolitan areas as a function of their populations and average densities. The model is robust (Adjusted R= 0.53) and all the coefficient are statistically significant (*p*-values<0.05). The coefficient of the population is 1.26 (which is higher than 1 and higher that the postulated 1.15 as well) and the relationship is clearly *superlinear*. This translates to the observation that the expected number of deaths per capita from Covid-19 in a city with double the population of a smaller one will be 19.5% higher than that of the smaller city.

Again, the coefficient of density in the model is negative. In other words, in a city with a given population, a higher density will translate into a smaller total number of deaths. For example, other things being equal, a 10% increase in density will result in a 8% *decline* in the total number of deaths. In fact, when we consider the effects of both population and density on the total number of deaths, the population effect becomes stronger than the effect observed when density is not part of the model. In other words, including density in the model strengthens the population effect. In light of this finding, we must reject Hypothesis IV as well: The total number of deaths in U.S. metropolitan area is not *superlinear* with respect to density. In fact, for a given population, it decreases when density increases.

Variable	Coefficient	*p*-value
Intercept	-5.88	0.00058
Log MSA Population	1.26	0
Log MSA Density	-0.84	0.00064
Observations	371	
Adjusted R-Squared	0.53	

Table 2: Modeling the total number of deaths in U.S. metropolitan areas as a function of their populations and densities.

Concluding remarks:

This leads us into a new and interesting discussion. Average population density—or, more precisely, the average population density of urban extents—is the correct metric for measuring density when it comes to urban sprawl or when it comes to estimating the amount of land that is needed to accommodate a given population. Even though it is a gross measure—a measure that lumps together all land uses—it is a precise measure because it tells us how much land a city consumes now or is likely to consume in the future.

The reciprocal of Average Population Density is Urban Land Consumption per Capita. Thus, when density increases land consumption per person declines and the city consumes less land. When density decreases, land consumption per person increases and the city consumes more land. Objections to sprawl—although often including other concerns like uniform land uses or unattractive suburban strip mall landscapes—are typically focused on low-density development. It is low-density development that increases the consumption of land for cities and the conversion of rural land into urban land. The fight against low-density sprawl is a fight to limit or reduce urban expansion into the rural periphery and thus to conserve rural land. This is all measured adequately by the average urban population density, even though this is a rather gross measure of density.

Interestingly enough, average urban extent density is also the kind of measure that is adequate to focus attention on the effect of urban form on climate change. When density increases, distances between locations decline. When density doubles, distances between random points decrease, on average, by $\sqrt{2}$. This means that increased density can lead to a considerable reduction of Vehicle Miles Traveled. In addition, higher urban density (greater than, say, 30-50 persons per hectare) makes the use of public transit viable. It also increases the number of locations that can be reached by walking or biking. This is why the average urban extent density of cities is such an important metric with clear and obvious uses for making, guiding, and measuring progress in on numerous policy fronts.

It is for this reason that the results reported on in this note are important. They mean that the key metric used to fight urban sprawl, to plan for urban expansion, or to mitigate global warming in cities is not a measure which is useful in detecting, explaining, and acting on managing the Covid-19 epidemic. If we want to speak of 'density' as something that matters for detecting, explaining, and acting on this virus, we need to talk about other kinds of density and other measures of density (e.g. the share of the population living at densities above 10,000 persons per square mile, or the share of commuters using public transport). Those used to combat sprawl or to mitigate global warming will not do.

On Measuring Social interaction—A Research Agenda:

We know that there is a hierarchy of social interactions: a few are more intense and more frequent than others and the intensity and frequency decreases as the number of interactions increases.

Imagine a single individual with a cell phone. Over a 24-hour period of a 7-day period, this individual interacts (comes in close physical contact with) a large number of people. Suppose we had cell phone data that would place this individual at a given 10ft.-by-10ft. grid cell at a certain moment in time. We can calculate how many *different* people that person came in close contact with (shared a grid cell) during the period. This would allow us to construct a map of contact intensity for the city, including its 'hotspots': the places that have the highest numbers of unique close physical contacts in a given period, and hence the places that would require attention and concomitant action.

For the purposes of studying the spatial structure of Covid-19, the metric that matters is the total number of *unique* close contacts in a city in a given period. The focus on the number of unique contacts in a given period takes the hierarchy of social interactions and eliminates more and more if its core. The persons you know and interact with frequently—say, everyone you were in close contact with (i.e. occupied the same cell at the same time) during the past week—will only be counted once when constructing the map of the intensity of social interaction. The focus would shift to close contact with strangers, which is equivalent to new close contacts. What matters, in the final analysis, in modeling the spread of the Covid-19 virus is the number of distinct social interactions in close contact in a given time period in a given city. West postulates that this number would be *superlinear* with city population size and would help explain—in a much more direct way than 'density'—the relationship between city populations and the prevalence of cases and deaths.

* * *