NYU | Marron Institute of Urban Management

4 August 2020 Working Paper

COVID-19 THRIVES IN LARGER CITIES, NOT DENSER ONES

+SHLOMO (SOLLY) ANGEL AND ALEJANDRO M. BLEI

ABSTRACT

The findings introduced in this longitudinal study of cumulative Covid-19 cases and deaths are based on reported data for 384 U.S. Metropolitan Statistical Areas (MSAs) for 20 weeks, starting on 6 March 2020 and ending on 23 July 2020. We look at the variation in the 7-day averages of the cumulative numbers of reported cases and deaths in each city at the end of every week as a function of its total population, its 'urbanized area,' and its average population density (the ratio of its population and its urbanized area). We find that during the last 10 weeks, the numbers have tended to converge: (1) a city with double the population of a smaller one can be expected to have 17% more cases per capita and 28% more deaths per capita than the smaller city; (2) a city with double the urbanized area of a smaller one can be expected to have 19% more cases per capita and 38% more deaths per capita than the smaller city; and, finally, (3) a city with double the population density of a smaller one can be expected to have 4.1% fewer cases per capita and 7.4% fewer deaths per capita than the smaller city. Larger cities have more than their share of cases and deaths in part because the larger the city, the larger the number of possible interactions among its inhabitants. And it is this larger number, rather than the overall average proximity of people to each other—expressed by the average density in the city—that accounts for that larger share. In fact, when it pertains to Covid-19 cases and deaths, denser metropolitan areas appear to be better able to contain their numbers than more spread out ones.

COVID-19 Thrives in Larger Cities, Not Denser Ones[1]

Shlomo Angel and Alejandro M. Blei
The Marron Institute of Urban Management at New York University

Preamble—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 that an average individual can keep in touch with, estimated to be of the order of 150.[2] 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,

---

1 This research note follows an earlier one titled "The Spatial Structure of the Pandemic: Some

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

> 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 motivation for this study:

Because the Coronavirus spreads from one person to another when they are in close proximity, we have come to expect that dense places—where people are, on the whole, closer to each other—will have more than their share of Covid-19 cases and deaths. It is quite clear that places within cities where people are crowded together, be they crowded homes, crowded workplaces, crowded bars and restaurants, crowded stadiums and theaters, crowded trains, and crowded streets and public open spaces bring people into close proximity to each other and accelerate the spread of the pandemic. But crowding is localized, it is not evenly distributed throughout the city. The mere existence of some crowded places in a city does not tell us about the level of crowdedness in the city as a whole. That level is measured by the city's average population density, defined as the ratio of its total population and its total urbanized area. This density, which is easily measurable, is an important metric for assessing a number of public policy concerns: the sprawling of the city into the countryside, the length of its infrastructure lines, the amount of vehicles miles traveled, and the resulting greenhouse gas emissions from these vehicles, to take a few examples.[3] There have thus been repeated calls for increasing this density in cities, or at least for halting its decline. The Covid-19 pandemic has brought urban density into the fore as one of the possible explanations for the spread of the virus. It now threatens to compromise efforts at densification in future policy debates, unless we are able to show that urban density, as defined above does not, at the very least, affect the spread of the pandemic and, better yet, that cities with higher densities experience fewer cases and deaths than lower density ones.

As we shall see below, the cumulative numbers of Covid-19 cases and deaths are *superlinear* with metropolitan populations: larger cities have more than their share of cases and deaths, i.e. they have more cases and deaths per capita than smaller cities. We can also show than urban population density is *sublinear* with metropolitan populations: larger cities are denser than smaller ones. What we aim to show is that once we account for the population effect on Covid-19 cumulative cases and deaths and once the pandemic becomes

---

3 For a comprehensive review of studies showing the benefits of higher urban densities, see Boyko, C.T. and Cooper, R., 2011. Clarifying and Re-conceptualizing Density, *Progress in Planning* 76, 1–61; and Ahlfeldt, G. and Pietrostefani, E., 2018. "Demystifying compact urban growth: Evidence from 300 studies from across the world", *OECD Regional Development Working Papers*, 2018/03, OECD Publishing, Paris.

firmly established,[4] the effect of density becomes negative and we can see that the relationship between cumulative Covid-19 cases and deaths and density is sublinear: Denser cities have less than their share of cases and deaths i.e. they have fewer cases and deaths per capita than more sprawling cities. A similar insight was investigated by Hamidi, Sabouri and Ewing (2020)[5] and they have arrived at similar findings. Their study was focused on the effect of the gross density of *counties* within MSAs—defined as the sum of their population and employment divided by their area[6]—on cumulative Covid-19 cases and deaths by 25 May 2020 period. Essentially, their study contrasted dense central-city counties with suburban counties and outlying ex-urban counties within U.S. metropolitan areas. Our study, because it focuses on MSAs rather than on counties—contrasting dense MSAs with more sprawling ones—and because it is longitudinal—covering a longer study period, 6 March 2020-23 July 2020, rather than focusing on a single date—amplifies and substantiates their findings.

These are important findings because they highlight the role of urban populations in the spread of the virus, while minimizing or negating the role of urban density in that spread, freeing urban planners and policy makers to continue to pursue densification policies in the future with a clear conscience, policies that have numerous benefits, not the least of which is their potential mitigating effect on climate change.

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

---

4 An earlier study by Wheaton and Thompson using county data for 31 March 2020, found that, within U.S. metropolitan areas, counties with higher gross density has significantly more cumulative numbers of Covid-19 cases and deaths. See: Wheaton, W.C. and Thompson, A.K., 2020. The Geography of Covid-19 growth in the US: Counties and Metropolitan Areas, April 6. Available at SSRN: https://ssrn.com/abstract=3570540 or http://dx.doi.org/10.2139/ssrn.3570540.

5 Hamidi, S., Sabouri, S. an Ewing, R., 2020. Does Density Aggravate the COVID-19 Pandemic?, *Journal of the American Planning Association*, online at: 10.1080/01944363.2020.1777891

6 We note here that using the county area in the denominator when calculating urban density tends to exaggerate the difference in density between central counties in MSAs, which tend to be fully built-up with urban areas and peripheral counties that tend to have large areas which are not built-up (see figure 4 below). Measuring density with the 'urbanized area' within a county (defined in footnote 9 below) in the denominator, rather than the county's entire area, yields a more comparable measure of urban density, one that does not rely on arbitrary decisions on the administrative boundaries of counties. We also note that using both the residential population and the number of jobs in a county in the nominator tends to exaggerate the difference in density between central counties in MSAs, which tend to have more than their share of jobs and peripheral counties that tend to have fewer jobs.

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. Stier, Berman and Bettencourt observed such a *superlinear* relationship between the growth rate in the number of Covid-19 cases in U.S. cities and their populations very early in the current pandemic using data for 14-18 March 2020.[7] This is not just saying that the *number* of cases is *superlinear* with the population of cities, but that the rate of growth of that number is *superlinear* with the population of cities. This may have been true during the early stage of the Covid-19 pandemic, but it is certainly no longer true. There was no statistically significant relationship between the weekly growth rate of the number of cases and the population of U.S. metropolitan areas during any of the ten weeks ending on 23 July 2020.[i]

We therefore avoid focusing on the rate of growth of cases and deaths and begin by focusing instead on the cumulative numbers of cases and deaths in U.S. metropolitan areas and their *superlinear* relationships with their populations. We also shy away from examining these relationship at one point in time and focus on weekly observations during a period of 20 weeks ending on 23 July 2020, a period at the end of which these relationships seem to have stabilized. We present this formally as our first hypothesis thus:

> Hypothesis I: The cumulative numbers of confirmed Covid-19 cases and deaths in U.S. metropolitan areas, once the pandemic has been firmly established, is *superlinear* with their populations. More specifically, these numbers follow established scaling laws: A city with double the population of a smaller one will have 1.15 times the number of confirmed cases and deaths per capita of the smaller city.

We begin by looking at the data for the week ending on 23 July 2020. The relationship is shown in figures 1 and 2 for data on the 7-day average of cumulative number of Covid-19 confirmed cases and deaths on 23 July 2020 in the 384 MSAs in the United States. On that date, counties that were parts of MSAs accounted for 90.1% of Covid-19 cases and 93.7% of deaths in the US. The data is presented in logarithmic form.

---

7 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.

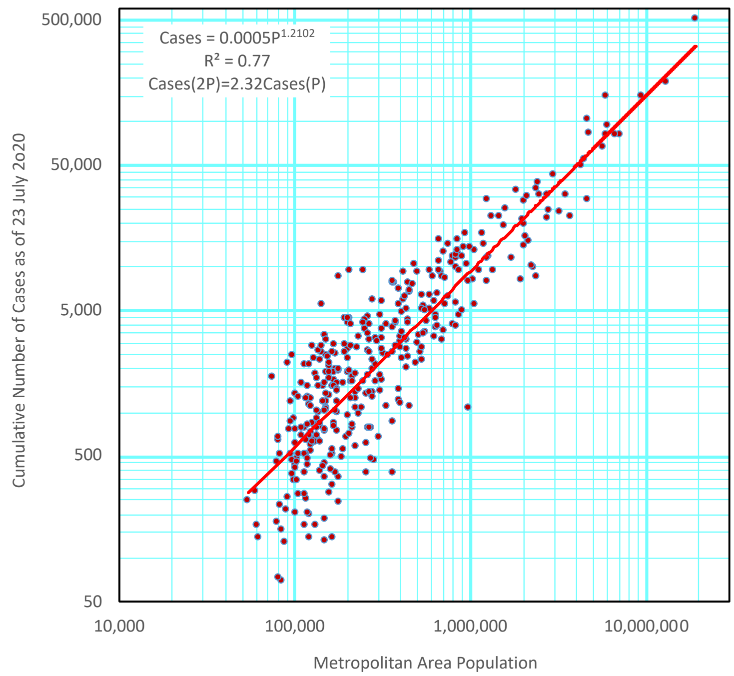


Figure 1: Cumulative number of Covid-19 cases in 384 U.S. Metropolitan Statisticasl Areas (MSAs) as a function of their populations, shown in logarithmic form.

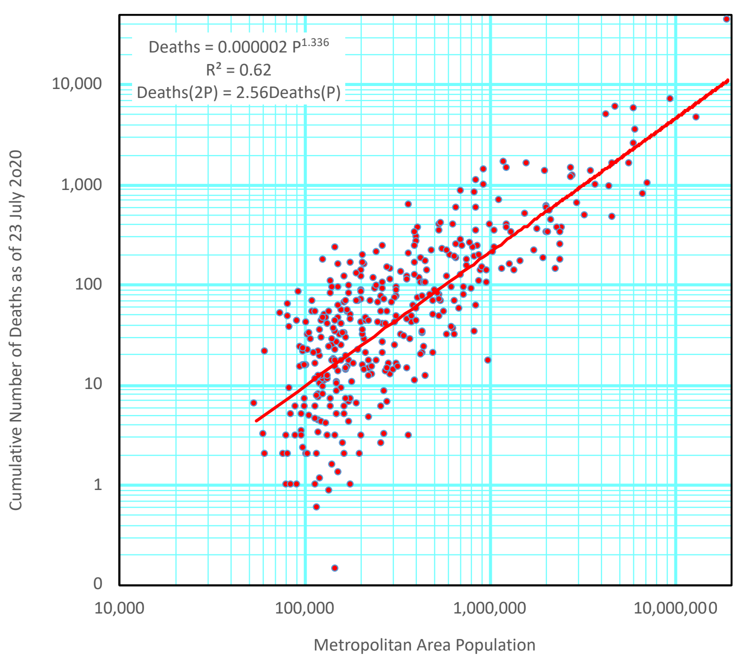


Figure 2: Cumulative number of Covid-19 deaths in 384 U.S. Metropolitan Statistical Areas (MSAs) as a function of their populations, shown in logarithmic form.

The Log-Log relationship between the cumulative number of cases and metropolitan populations shown in figure 1 is clearly *superlinear* and highly significant ($R^2 = 0.77$). The straight trend line is a power curve of the form [Cases] = A×[Population]$^{\beta}$, or, in logarithmic form Log[Cases] = LogA + βlog[Population], which is a linear relationship that can be investigated with linear regression models. The slope of the power curve on that date was 1.21, which implies that when the population of a city doubles, the cumulative number of Covid-19 cases increases by a factor of 1.16 (which is very similar to the postulated 1.15).[8]

The same relationship holds for Covid-19 deaths. The Log-Log relationship between the cumulative number of deaths and metropolitan populations shown in figure 2 is also clearly *superlinear* and highly significant ($R^2 = 0.62$). The slope of the Power Curve is 1.37, which implies that when the population of a city doubles, the cumulative number of Covid-19 deaths per capita increases by a factor of 1.28 (which is almost double the postulated 1.15).

The reader is reminded that the relationship between the cumulative numbers of Covid-19 cases and deaths and the populations or areas of U.S. cities can and does change over time. As several months have now passed since the onset of the pandemic, we can ask ourselves whether this relationship is by now stable, namely whether the parameter of the power functions characterizing them and the scaling coefficients derived from them have converged into stable averages. Instead of exploring these relationships at a single date—like the date 23 July 2020, which is the latest date for which we have data—we have chosen to focus on weekly observations during the past 20 weeks, from 6 March 2020 to 23 July 2020. For the last day of each of these 20 weeks, we obtained a 7-day average value for the cumulative numbers of cases and deaths. We then ran simple regression models in logarithmic form, one for cumulative cases and one for cumulative deaths as dependent variables and the MSA population as an independent variable for each of the 20 weeks, i.e. a total of 40 models. The results of the first set of 20 models—for cumulative Covid-19 cases—are summarized in table 1, and the results for the second set of 20 models—for Covid-19 deaths—are summarized in table 2.

8 This requires some explanation. The scaling coefficient S is calculated from $S = 2^{\beta}/2 = 2^{\beta-1}$. Therefore, $S = 2^{1.21-1} = 2^{0.21} = 1.16$.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Cases in Week Ending at Date as a Dependent Variable:										
Population Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	-4.264	-11.271	-15.027	-13.465	-11.897	-11.107	-10.498	-10.074	-9.678	-9.409
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	0.339	0.934	1.350	1.347	1.300	1.281	1.261	1.249	1.236	1.23
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.190	0.496	0.656	0.718	0.736	0.722	0.700	0.678	0.665	0.658
Population Scaling Factor	0.632	0.955	1.275	1.272	1.231	1.215	1.198	1.188	1.178	1.173
Population Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-9.207	-8.991	-8.844	-8.681	-8.573	-8.482	-8.380	-8.166	-7.926	-7.675
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	1.226	1.221	1.219	1.217	1.219	1.223	1.229	1.228	1.223	1.217
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.656	0.656	0.66	0.667	0.681	0.703	0.727	0.747	0.760	0.769
Population Scaling Factor	1.170	1.166	1.164	1.162	1.164	1.167	1.172	1.171	1.167	1.162

Table 1: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 cases as a dependent variable and the logarithm of the MSA population as an independent variable for 20 weeks between 6 March 2020 and 23 July 2020.

Examining table 1, the reader can verify that during the early weeks, the regression models in the table are weaker and have a lower R-squared value, and that the R-squared value increases systematically over time. One can also notice that the population coefficient fluctuates during the early weeks before stabilizing during the last ten weeks. We therefore focus our attention on the pooled results from the last ten weeks, from the week ending on 21 May 2020 to the week ending on 23 July 2020 and we look at the average weekly values for these ten weeks. The Log-Log relationship between the average cumulative number of cases and metropolitan populations is clearly *superlinear* and has high explanatory power (the 10-week average R^2 = 0.70±0.03, where 0.03 is the 95% confidence interval). The 10-week average slope of the power curve is 1.222±0.003 and the resulting 10-week average population scaling factor is 1.167±0.003: When the population of a city doubles, the cumulative number of Covid-19 cases increases by a factor of 1.167±0.003 (which is very similar to the postulated 1.15). This translates to the observation that the expected cumulative number of confirmed cases per capita in a city with double the population of a smaller one will be 16.7% 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 uniformly distributed among the population of cities or among the population of the U.S. as a whole.

We next look at the relationships of city populations and Covid-19 deaths, summarized in table 2. Again, we notice that during the early weeks, the regression models in the table are weaker and have a lower R-squared value, and that the R-squared value increases systematically over time. We also notice that the population coefficient fluctuates during the early weeks before stabilizing during the last ten weeks. We again focus our attention on the last ten weeks, from the week ending on 21 May 2020 to the week ending on 23 July 2020 and we look at the average weekly values for these ten weeks. The Log-Log relationship between the average cumulative number of Covid-19 deaths and metropolitan populations is

clearly *superlinear* and highly significant (R^2 = 0.57±0.01). The average slope of the power curve is 1.354±0.004 and the resulting average population scaling factor is 1.278±0.004, which implies that when the population of a city doubles, the cumulative number of Covid-19 deaths increases by a factor of 1.278±0.003 (which is almost double the postulated 1.15). This translates to the observation that the expected number of confirmed deaths per capita in a city with double the population of a smaller one will be 27.8% higher than that of the smaller city. In other words, larger cities have a lot more than their share of confirmed Covid-19 deaths than would be expected if Covid-19 deaths were evenly distributed among the U.S. cities or among the U.S. as a whole.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Deaths in Week Ending at Date as a Dependent Variable:										
Population Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	0.309	-0.283	-3.956	-9.274	-13.089	-13.983	-14.331	-14.307	-14.387	-14.333
p-value Intercept	0.038	0.306	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	-0.025	0.018	0.316	0.771	1.120	1.236	1.295	1.319	1.344	1.354
p-value Population	0.029	0.402	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.010	-0.001	0.193	0.456	0.578	0.606	0.590	0.596	0.582	0.568
Population Scaling Factor	0.491	0.506	0.622	0.853	1.087	1.178	1.227	1.247	1.269	1.278
Population Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-14.329	-14.165	-13.997	-13.812	-13.739	-13.631	-13.608	-13.568	-13.472	-13.232
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	1.365	1.362	1.357	1.350	1.350	1.348	1.352	1.354	1.355	1.345
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.564	0.556	0.555	0.557	0.558	0.566	0.574	0.585	0.596	0.611
Population Scaling Factor	1.288	1.285	1.281	1.275	1.275	1.273	1.276	1.278	1.279	1.270

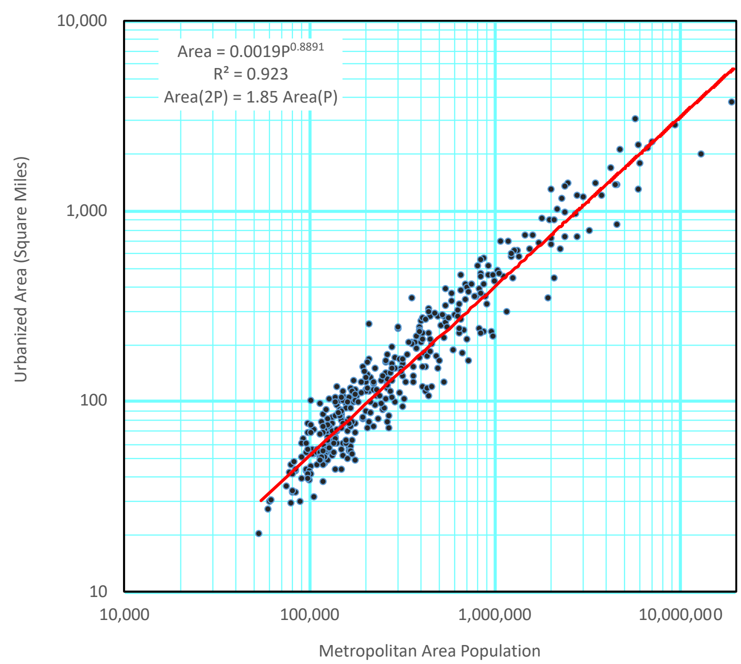
Table 2: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 deaths as a dependent variable and the logarithm of the MSA population as an independent variable for 20 weeks between 6 March 2020 and 23 July 2020.

We can conclude, therefore, given the evidence, we fail to reject Hypothesis I. Once the pandemic has become firmly established, the number of confirmed cases and deaths in U.S. MSAs has become firmly *superlinear* with their populations. The second part of the hypothesis is confirmed for cases: The number of cases per capita does increase by the expected 15%—it increases by 16.7%, to be precise—when city populations double. The number of deaths per capita increases by almost twice the expected 15%—it increases by 28%, to be precise—when city populations double.

This in an important finding. It tells policy makers that Covid-19 cases and deaths are not evenly distributed among the U.S. population. To the extent that federal or any other resources for identifying and treating Covid-19 cases are distributed evenly according to a formula that uses population share to allocate these resources, larger metropolitan areas will receive too little resources, while rural areas and smaller cities and towns will receive too much. Similarly, to the extent that larger metropolitan areas lean more towards the Democratic party than smaller cities and towns, we can safely say that Democratic voters are bearing a heavier burden of Covid-19 cases, and an even heavier burden of Covid-19 deaths than Republican voters.

On the *Superlinearity* of Urban Densities with City Populations:

Our main interest in this research note is to investigate the relationship between urban density on the one hand and Covid-19 cases and deaths in U.S. metropolitan areas on the other. This is not as straightforward as it seems. Data on the populations and the urbanized area[9] within all 384 Metropolitan Statistical Areas (MSAs) in the U.S.[10] is readily available, and we can thus calculate their average population density as the ratio of the former and the latter. The problem is—and it has been noted before—that urban density itself is *superlinear* with city population: Larger cities are denser than smaller ones. This is demonstrated clearly by graphing the urbanized areas of U.S. MSAs against their populations, as shown in figure 3 below.



9 Urbanized Areas within Metropolitan Statistical Areas (MSAs) are defined by the U.S. Census Bureau as the total area of contiguous or near contiguous census tracts and/or blocks with population densities in excess of 1,000 persons per square mile and a combined population of at least 50,000 inhabitants. Urbanized Area criteria for the 2010 Census are described in the Federal Register of August 24, 2011.

10 U.S. Metropolitan Statistical Areas (MSAs) are defined by the U.S. Census Bureau and consist of counties that are linked together by commuting patterns.

Figure 3: The urbanized areas of 384 MSAs in the United States as a function of their populations, presented in logarithmic form.

The relationship between the urbanized areas and the population of U.S. metropolitan areas is *sublinear* and very strong (R^2=0.92). Figure 2 shows that when the population of a city is double that of a smaller one, its area does not double by increases by only 85%. Since its average density is the ratio of its population and its urbanized area, its density increases by 8% (2.00 ÷ 1.85= 1.08).

Given this finding, we cannot look directly at the relationship between urban density and the numbers of cumulative cases and deaths in U.S. cities without taking their populations into account, because some of the effect of density on these numbers will be due to population size. We have to account for the effect of population on the cumulative numbers of cases and deaths when we look at the effect of urban density on these numbers. We can do this in a purely statistical way by examining multiple regression models with the cumulative numbers of cases or deaths as dependent variables and the populations and densities of MSAs as independent variables. We have done that, as we shall show below, but it hides more than it reveals. A better way is to look first at the *superlinear* relationship between the cumulative numbers of Covid-19 cases and deaths and the urbanized areas of cities, and then infer the effect of density on these numbers from the two *superlinear* relationships, those of the populations and those of urbanized areas of cities.

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

The ‘urbanized areas’ within Metropolitan Statistical Areas (MSAs) are defined by the U.S. Census Bureau as the total area of contiguous or near-contiguous census tracts and/or blocks with population densities in excess of 1,000 persons per square mile and a combined population of at least 50,000 inhabitants . The urbanized areas within the Chicago and Houston MSAs, are shown in figure 4 below. We now focus on the cumulative numbers of cases and deaths in U.S. metropolitan areas and their *superlinear* relationships with their urbanized areas (or ‘areas’, for short).

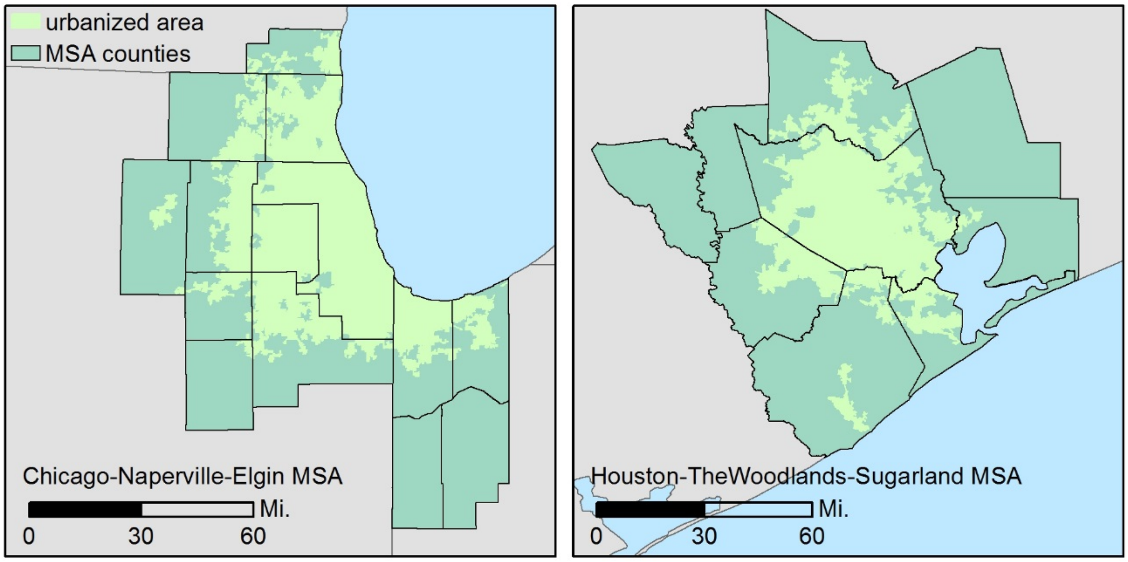


Figure 4: The urbanized area (light green) within the 14 counties that comprised the Chicago MSA (left) and the 9 counties that comprised the Houston MSA (right). The Chicago MSA contained

The Log-Log relationship between the cumulative number of cases and urbanized areas is clearly *superlinear* and quite strong ($R^2 = 0.73$). The straight trend line is a power curve of the form [Cases] = A×[Area]$^{\beta}$, or, in logarithmic form Log[Cases] = LogA + βlog[Area], which is a linear relationship that can be investigated with linear regression models. The slope of the power curve is 1.26, which implies that when the area of a city doubles, the cumulative number of Covid-19 cases increases by a factor of 1.20 (which is somewhat higher than the postulated 1.15).

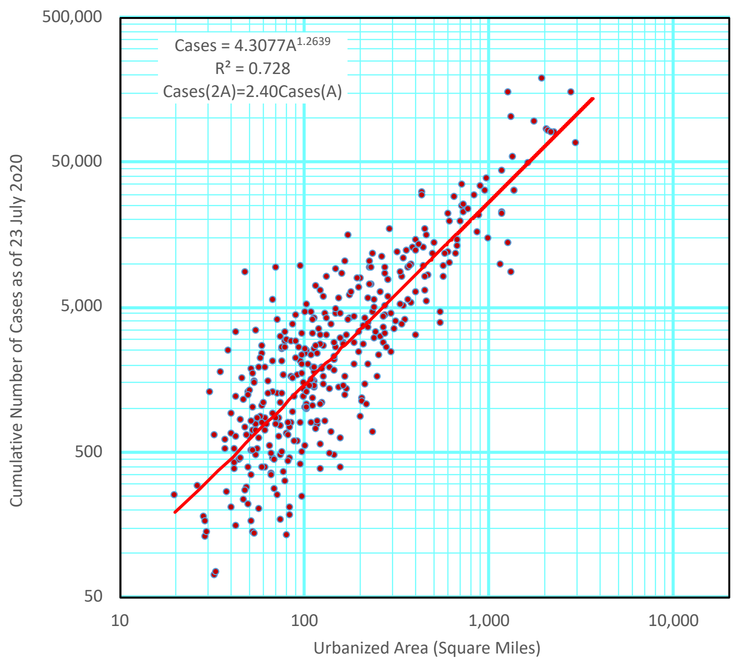


Figure 5: Cumulative number of Covid-19 cases (left) and deaths (right) in 384 U.S. Metropolitan Statistical Areas (MSAs) as a function of their areas, shown in logarithmic form.

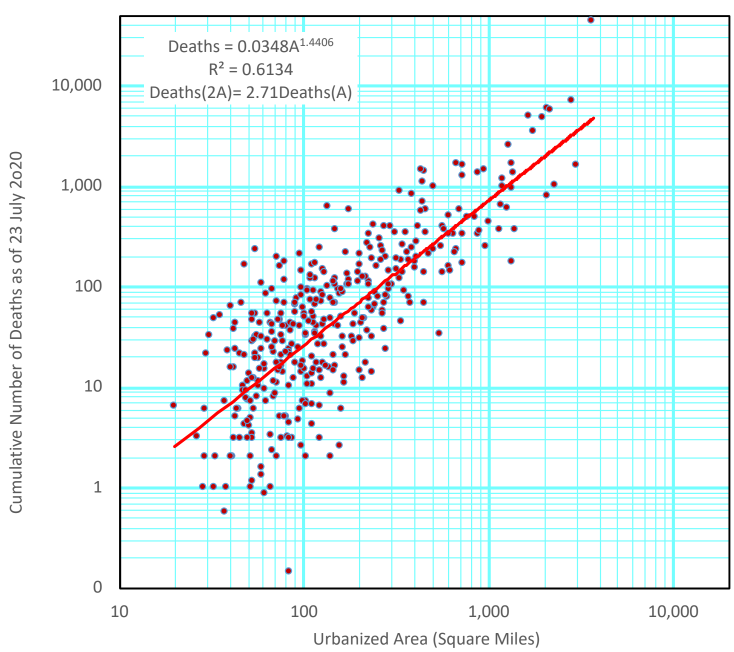


Figure 6: Cumulative number of Covid-19 cases (left) and deaths (right) in 384 U.S. Metropolitan Statisticasl Areas (MSAs) as a function of their areas, shown in logarithmic form.

The same relationship holds for Covid-19 deaths. The Log-Log relationship between the cumulative number of deaths and the urbanized areas of cities (shown on the right) is also clearly *superlinear* and quite strong ($R^2 = 0.61$). The slope of the Power Curve is 1.44, which implies that for a city with double the area of a smaller one the cumulative number of Covid-19 deaths increases by a factor of 1.36 (which is more than double the postulated 1.15).

As several months have now passed since the onset of the pandemic, we can ask ourselves whether the cumulative numbers of Covid-19 cases and deaths have a stable *superlinear* relationship with urbanized area as well. Namely, whether the parameter of the power functions characterizing them and the scaling coefficients derived from them have converged into stable averages. We again focus on the past 20 weeks, from 6 March 2020 to 23 July 2020. For the last day of each of these 20 weeks, we obtained a 7-day average value for the cumulative numbers of cases and deaths. We then ran simple regression models in logarithmic form, one for cumulative cases and one for cumulative deaths as dependent variables and the MSA urbanized area as an independent variable for each of the 20 weeks, i.e. a total of 40 models. The results of the first set of 20 models—for cumulative Covid-19 cases—are summarized in table 3, and the results for the second set of 20 models—for Covid-19 deaths—are summarized in table 4.

The reader can again notice that during the early weeks, the regression models in table 3 are weaker and have a lower R-squared value, and that the R-squared values increase systematically over time. One can also notice that the urbanized area coefficients fluctuate during the early weeks before stabilizing during the last ten weeks. We again focus our attention on the pooled results of the last ten weeks, and look at the average weekly values for these ten weeks. The Log-Log relationship between the average cumulative number of cases and the urbanized areas of cities is clearly *superlinear* and quite strong highly significant (the 10-week average $R^2 = 0.67 \pm 0.03$). The 10-week average slope of the power curve is 1.285 ± 0.003 and the 10-week average area scaling factor is 1.219 ± 0.003, which implies that when the area of a city doubles, the cumulative number of Covid-19 cases increases by a factor of 1.219 ± 0.003. This translates to the observation that the expected number of confirmed cases per capita in a city with double the urbanized area of a smaller one will be 22% higher—namely, almost one-half higher than the postulated 15%—than that of the smaller city. In other words, cities with larger areas have considerably more than their share of confirmed Covid-19 cases than would be expected if cases were distributed among U.S. cities or U.S. counties in proportion to their populations.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Cases in Week Ending at Date as a Dependent Variable:										
Urbanized Area Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	-1.530	-4.099	-4.823	-3.379	-2.232	-1.639	-1.183	-0.848	-0.532	-0.291
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.039	0.265
Log Urbanized Area Coefficient	0.314	0.935	1.383	1.401	1.366	1.357	1.336	1.323	1.308	1.297
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.139	0.425	0.590	0.665	0.696	0.693	0.673	0.652	0.637	0.627
Urbanized Area Scaling Factor	0.622	0.956	1.304	1.320	1.289	1.281	1.262	1.251	1.238	1.229
Urbanized Area Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-0.109	0.063	0.197	0.345	0.464	0.602	0.76	0.974	1.182	1.378
p-value Intercept	0.676	0.808	0.444	0.174	0.061	0.012	0.001	0.000	0.000	0.000
Log urbanized Area Coefficient	1.292	1.286	1.285	1.282	1.284	1.287	1.291	1.287	1.282	1.277
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.623	0.624	0.627	0.634	0.647	0.666	0.686	0.703	0.715	0.726
Urbanized Area Scaling Factor	1.224	1.219	1.218	1.216	1.218	1.220	1.223	1.220	1.216	1.212

Table 3: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 cases as a dependent variable and the logarithm of the MSA urbanized area as an independent variable for 20 weeks between 6 March 2020 and 23 July 2020.

We next look at the relationships of the urbanized areas of cities and Covid-19 deaths, summarized in table 4. Again, we notice that during the early weeks, the regression models in the table are weaker and have a lower R-squared value, and that the R-squared values increase systematically over time. We can also notice that the urbanized area coefficients fluctuate during the early weeks before stabilizing during the last ten weeks. We again focus our attention on the last ten weeks, and we look at the average weekly values for these ten weeks.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Deaths in Week Ending at Date as a Dependent Variable:										
Urbanized Area Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	0.106	-0.108	-1.415	-3.467	-4.796	-4.910	-4.877	-4.706	-4.636	-4.504
p-value Intercept	0.099	0.366	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Urbanized Area Coefficient	-0.024	0.011	0.294	0.795	1.184	1.321	1.395	1.426	1.459	1.469
p-value Population	0.057	0.633	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.007	-0.002	0.142	0.415	0.553	0.593	0.586	0.597	0.588	0.573
Urbanized Area Scaling Factor	0.492	0.504	0.613	0.868	1.136	1.249	1.315	1.344	1.375	1.384
Urbanized Area Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-4.406	-4.275	-4.132	-4.003	-3.931	-3.829	-3.778	-3.706	-3.600	-3.425
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log urbanized Area Coefficient	1.479	1.477	1.470	1.463	1.464	1.459	1.463	1.464	1.463	1.451
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adjusted R-squared	0.567	0.560	0.558	0.560	0.562	0.568	0.576	0.585	0.595	0.609
Urbanized Area Scaling Factor	1.394	1.392	1.385	1.378	1.379	1.375	1.378	1.379	1.378	1.367

Table 4: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 deaths as a dependent variable and the logarithm of the MSA urbanized area as an independent variable for 20 weeks between 6 March 2020 and 23 July 2020.

The Log-Log relationship between the average cumulative number of Covid-19 deaths and the urbanized areas of cities is clearly *superlinear* and quite strong (R^2 = 0.57±0.01). The average slope of the power curve is 1.465±0.006 and the average population scaling factor is 1.381±0.006, which implies that when the urbanized area of a city doubles, the cumulative number of Covid-19 deaths increases by a factor of 1.381±0.006 (which is 2.5 times the postulated 1.15). This translates to the observation that the expected number of confirmed deaths per capita in a city with double the area of a smaller one will be 38.1%

higher than that of the smaller city. In other words, cities with larger areas have a lot more than their share of confirmed Covid-19 deaths than would be expected if Covid-19 deaths were distributed among U.S. cities or U.S. counties in proportion to their populations.

We can conclude, therefore, given the evidence, we fail to reject Hypothesis II. Once the pandemic has become firmly established, the number of confirmed cases and deaths in U.S. MSAs has become firmly *superlinear* with their urbanized areas. The second part of the hypothesis is confirmed too, but with higher values than the postulated 15%: The number of cases per capita does increase by one-half more than the expected 15%—it increases by 22%, to be precise—when city populations double. The number of deaths per capita increases by 2.5 times the expected 15%—it increases by 38%, to be precise—when city areas double.

On the *Sublinearity* of Covid-19 Cases and Deaths with Urban Density:

A cursory glance at the differences between the scaling coefficients of population and urbanized area and their change over time presented in the previous two sections suggests that the effect of the urbanized area on the cumulative numbers of Covid-19 cases and deaths in cities is stronger—i.e. significantly larger in statistical terms—than the effect of population. To wit, for cumulative Covid-19 cases, the average population scaling coefficient for the past 10 weeks is 1.167±0.003, while the average urbanized area scaling coefficient for cases for the past 10 weeks is 1.219±0.006. For Covid-19 deaths, the average population scaling coefficient for the past 10 weeks is 1.278±0.003, while it is 1.381±0.006 for urbanized area. This, in turn, suggests that in two cities with the same population, the one having a larger area than the other will have more cases and deaths. This is interesting, because the one with the larger area will also have a lower average population density. That suggests that the effect of population density on the cumulative numbers of Covid-19 cases and deaths can be expected to be negative: The higher the density, the lower these numbers can be expected to be. We thus formulate our third hypothesis:

> Hypothesis III: The cumulative number of Covid-19 cases and deaths in U.S. metropolitan areas is *sublinear* with their densities. Accounting for their populations, denser cities will have fewer cases and deaths per capita than less dense ones. More specifically, these numbers follow established scaling laws: A city with double the density of another one will have 0.87 times the number of confirmed cases and deaths per capita of the less dense city.

We examine the evidence for this hypothesis with data for the 20 weeks beginning on 6 March 2020 and ending in 23 July 2020 in two complementary ways. We first ask whether multiple linear regression models with cumulative cases or deaths as dependent variables and both MSA population and density as independent variables yields statistically significant coefficients, a positive one for population and a negative one for density. This is explored in tables 5 and 6 below.

Looking at table 5, the reader can again ascertain that during the early weeks, the regression models are weaker and have lower R-squared values. The R-squared values then increase to a peak in Week 5, decline during Weeks 6-11, and then increase slowly yet systematically over time during Weeks 12-20. One can also notice that the density coefficient becomes negative in Week 4 and remains negative thereafter. We again focus our attention on the last ten weeks, and look at the average weekly values for these ten weeks. The Log-Log relationship between the average cumulative number of cases and the average population density of cities is clearly *sublinear* but it is not statistically significant: The average value of the density coefficient during this period is -.214 and its average *p*-value is 0.182, which is above 0.05, implying that the density coefficient is not statistically significant at the 95% confidence level. This translates to the observation that the expected cumulative number of confirmed cases per capita in a city with double the density of another one will not be significantly larger; it is likely to be smaller, but we cannot say with confidence that it will be significantly smaller.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Cases in Week Ending at Date as a Dependent Variable:										
Population & Density Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	-7.393	-14.087	-16.474	-13.266	-10.552	-8.93	-8.303	-7.884	-7.715	-7.778
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	0.284	0.884	1.324	1.351	1.324	1.32	1.299	1.288	1.271	1.259
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Density Coeficient	0.498	0.448	0.231	-0.032	-0.214	-0.347	-0.350	-0.349	-0.313	-0.260
p-value Density	0.0003	0.016	0.236	0.85	0.167	0.028	0.033	0.041	0.073	0.141
Adjusted R-squared	0.216	0.502	0.656	0.718	0.736	0.725	0.703	0.681	0.667	0.659
Population & Density Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-7.691	-7.456	-7.286	-7.160	-7.015	-7.108	-7.172	-7.163	-6.935	-6.489
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	1.253	1.248	1.247	1.244	1.246	1.247	1.251	1.245	1.241	1.236
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Density Coeficient	-0.241	-0.244	-0.248	-0.242	-0.248	-0.219	-0.192	-0.160	-0.158	-0.183
p-value Density	0.171	0.164	0.154	0.156	0.135	0.167	0.200	0.261	0.249	0.166
Adjusted R-squared	0.657	0.657	0.661	0.668	0.682	0.704	0.727	0.748	0.760	0.771

Table 5: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 cases as a dependent variable and the logarithms of the MSA population and density as independent variables for 20 weeks between 6 March 2020 and 23 July 2020.

In the case of Covid-19 deaths, we *can* say that the expected number of confirmed deaths per capita in a city with double the density of another one will be significantly smaller. Looking at table 6, the reader can again notice that during the first five weeks, the regression models are weaker and have lower R-squared values. The R-squared values then stabilize in Week 6 and remain at approximately the same value during the remaining weeks. One can also notice that the density coefficient becomes negative in Week 5 and remains negative thereafter. We again focus our attention on the last ten weeks and we look at the average weekly values for these ten weeks. The Log-Log relationship between the average cumulative number of deaths and the average population density of cities is clearly *sublinear* and, in the cases of Covid-19 deaths, it is statistically significant: The average value of the density coefficient during this period is -.778±0.026 and its average *p*-value is 0.001±0.000, which is above 0.05, implying that the density coefficient is statistically significant at the 95% confidence level. This translates to the observation that the expected cumulative number of confirmed deaths per capita in a metropolitan area with double the density of another one will not be larger; it can be expected to be significantly smaller.

The logarithm of the 7-Day average Cumulative Number of Covid-19 Deaths in Week Ending at Date as a Dependent Variable:										
Population & Density Models Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Intercept	0.503	-0.909	-6.751	-9.739	-11.379	-10.858	-10.229	-9.656	-9.159	-9.130
p-value Intercept	0.119	0.130	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	-0.022	0.007	0.267	0.763	1.151	1.291	1.367	1.401	1.436	1.446
p-value Population	0.083	0.762	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Density Coeficient	-0.031	0.100	0.445	0.074	-0.272	-0.498	-0.653	-0.741	-0.833	-0.829
p-value Density	0.498	0.239	0.000	0.658	0.152	0.012	0.002	0.001	0.000	0.000
Adjusted R-squared	0.008	0.000	0.216	0.455	0.579	0.611	0.599	0.607	0.596	0.581
Population & Density Models Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Intercept	-9.283	-9.019	-9.017	-8.786	-8.634	-8.733	-8.723	-8.854	-8.877	-8.764
p-value Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Population Coefficient	1.454	1.452	1.445	1.439	1.44	1.434	1.438	1.438	1.436	1.424
p-value Population	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log Density Coeficient	-0.804	-0.819	-0.793	-0.801	-0.813	-0.780	-0.778	-0.751	-0.732	-0.712
p-value Density	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Adjusted R-squared	0.575	0.568	0.567	0.569	0.57	0.577	0.586	0.596	0.606	0.621

Table 6: Linear regression models with the logarithm of the 7-day average of the cumulative number of Covid-19 deaths as a dependent variable and the logarithms of the MSA population and density as independent variables for 20 weeks between 6 March 2020 and 23 July 2020.

Given these results, we can confirm the first part of Hypothesis III: The cumulative number of Covid-19 cases in U.S. metropolitan areas is indeed *sublinear* with their densities, but it is significantly sublinear only in regard to deaths, but not in regard to cases. Accounting for their populations, denser cities are likely to have fewer cases per capita than less dense ones and can be expected to have fewer deaths per capita than less dense ones.

Finally, we can derive the scaling factors for density from the scaling factors for population and urbanized area using a simple formula. Denote the population, urbanized area, density, and the number of cumulative Covid-19 cases at a given time period as *P, U, D,* and *C* respectively. Denote the scaling factors for population, urbanized area and density as S_P, S_U, and S_D respectively. We know that the number of cases in a city with double the population of another one will be double the number of cases multiplied by the population scaling factor:

$$(1) \qquad C(2P) = 2S_PP.$$

Similarly,

$$(2) \qquad C(2U) = 2S_UU.$$

By definition, we know that density is the ratio of a city's population and is urbanized area:

$$(3) \qquad D = P/U.$$

We can divide equation (1) by equation (2):

$$(4) \qquad C[2(P/U)] = (S_P/S_U)(P/U), \text{ or } C(2D) = (S_P/S_U)D.$$

In other words, the case scaling factor for density is simply the ratio of the scaling factors for population and urbanized area. The same is, of course, true for Covid-19 deaths as well. We used this finding to construct tables 7 and 8 below, as well as figures 4 and 5 that summarizes the data in these two tables.

7-Day average Cumulative Number of Covid-19 Cases in Week Ending at Date as a Dependent Variable:										
Density Scaling Factor Estimate Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Population Scaling Factor	0.632	0.955	1.275	1.272	1.231	1.215	1.198	1.188	1.178	1.173
Urbanized Area Scaling Factor	0.622	0.956	1.304	1.320	1.289	1.281	1.262	1.251	1.238	1.229
Density Scaling Factor	1.017	0.999	0.977	0.963	0.955	0.949	0.949	0.950	0.951	0.955
Density Scaling Factor Estimate Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Population Scaling Factor	1.170	1.166	1.164	1.162	1.164	1.167	1.172	1.171	1.167	1.162
Urbanized Area Scaling Factor	1.224	1.219	1.218	1.216	1.218	1.220	1.223	1.220	1.216	1.212
Density Scaling Factor	0.955	0.956	0.955	0.956	0.956	0.957	0.958	0.960	0.960	0.959

Table 7:Deriving the density scaling factor for the cumulative number of Covid-19 cases as a ratio of the population and the urbanized area scaling factors.

The weekly population scaling factors for the cumulative number of cases in table 7 were given in table 1, and the weekly urbanized area scaling factors for the cumulative number of cases were given in table 3. The density scaling factor is their ratio. Similarly, he weekly population scaling factors for the cumulative number of deaths in table 8 were given in table 2, and the weekly urbanized area scaling factors for the cumulative number of cases were given in table 4.

7-Day average Cumulative Number of Covid-19 Deaths in Week Ending at Date as a Dependent Variable:										
Density Scaling Factor Estimate Weeks 1-10	3/12/20	3/19/20	3/26/20	4/2/20	4/9/20	4/16/20	4/23/20	4/30/20	5/7/20	5/14/20
Population Scaling Factor	0.491	0.506	0.622	0.853	1.087	1.178	1.227	1.247	1.269	1.278
Urbanized Area Scaling Factor	0.492	0.504	0.613	0.868	1.136	1.249	1.315	1.344	1.375	1.384
Density Scaling Factor	0.999	1.005	1.015	0.984	0.957	0.943	0.933	0.929	0.923	0.923
Density Scaling Factor Estimate Weeks 11-20	5/21/20	5/28/20	6/4/20	6/11/20	6/18/20	6/25/20	7/2/20	7/9/20	7/16/20	7/23/20
Population Scaling Factor	1.288	1.285	1.281	1.275	1.275	1.273	1.276	1.278	1.279	1.270
Urbanized Area Scaling Factor	1.394	1.392	1.385	1.378	1.379	1.375	1.378	1.379	1.378	1.367
Density Scaling Factor	0.924	0.923	0.925	0.925	0.924	0.926	0.926	0.927	0.928	0.929

Table 8:Deriving the density scaling factor for the cumulative number of Covid-19 deaths as a ratio of the population and the urbanized area scaling factors.

Again, we can see that the scaling factors gradually stabilize. This can be most clearly seen in figures 4 and 5 below. Figure 4 shows that the scaling factors for population and urbanized area in the case of Covid-19 cumulative cases are closer in value to each other, and their ratio is therefore closer to 1. Indeed, the scaling factor for density in the case of Covid-19 cumulaive cases converged to 0.957±0.001 during the last ten weeks, while the scaling factor for density in the case of Covid-19 deaths, shown in figure 5, converged to 0.926±0.001 during this period. In other words, when density doubles, the cumulative number of cases can be expected to decline by 4.1% and the cumulative number of deaths can be expected to decline by 7.4%.

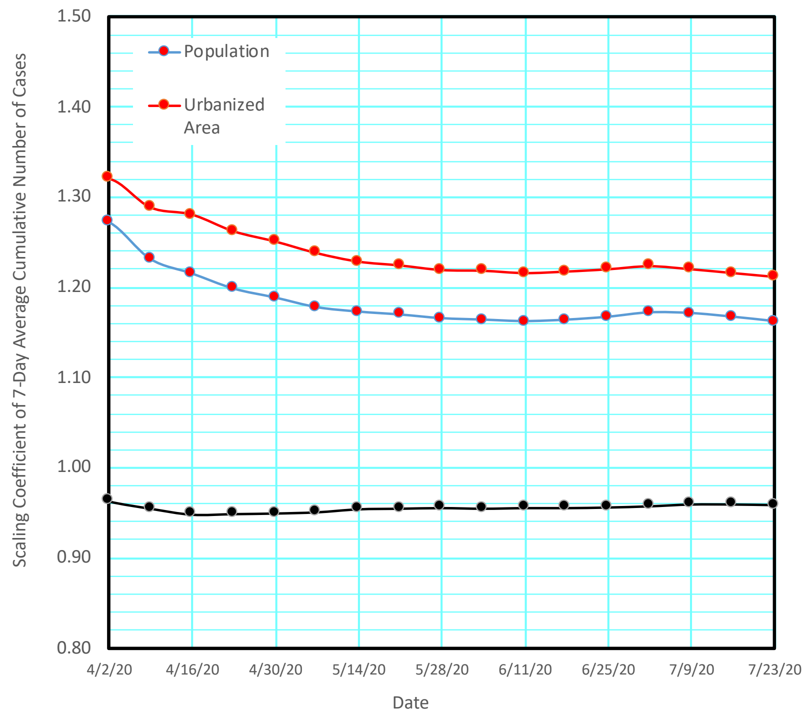


Figure 4: The convergence of the population, land area, and density scaling coefficients for the cumulative numbers of Covid-19 cases for the 17 weeks from 27 March 2020 to 23 July 2020

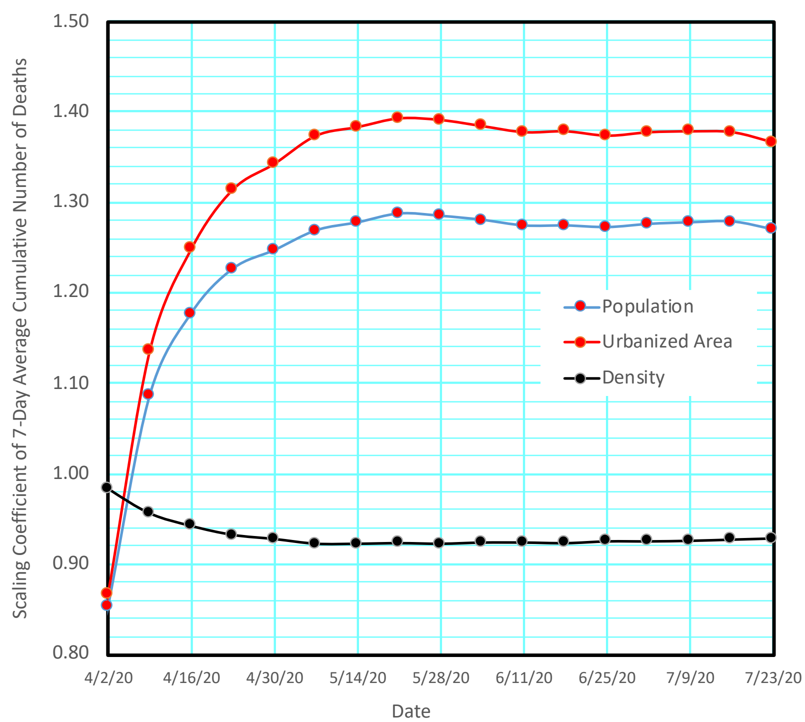


Figure 5: The convergence of the population, land area, and density scaling coefficients for the cumulative numbers of Covid-19 deaths for the 17 weeks from 27 March 2020 to 23 July 2020

Given these results, we can confirm the second part of Hypothesis III: The cumulative numbers of Covid-19 cases in U.S. metropolitan areas are indeed *sublinear* with their densities, but their scaling factor converges to only 96%, not as low as the expected 87%. The cumulative numbers of Covid-19 deaths in U.S. metropolitan areas are also *sublinear* with their densities. Its scaling factor, 93% is lower, but not as low as the expected 87%.

Conclusion:

To our knowledge, this is the first longitudinal study of the scaling relationships between the cumulative numbers of Covid-19 cases and deaths with the population, area, and density of U.S. Metropolitan Statistical Areas (MSAs). We are able to show that as the pandemic stabilized—while still far from being under control—these scaling relationships stabilized as well. And when we look at these stabilized relationships, we can confirm several important results.

First, cities and metropolitan areas with larger populations have significantly higher cumulative numbers of Covid-19 cases and deaths than their share in the overall population of the country. In fact, a city or metropolitan area with double the population of another one can be expected to have 16.7±0.3% more cases per capita and 27.8±0.3% more deaths per capita.

Second, cities and metropolitan areas with larger urbanized areas have significantly higher cumulative numbers of Covid-19 cases and deaths than their share in the overall population of the country. In fact, a city or metropolitan area with double the area of another one can be expected to have 21.9±0.6% more cases per capita and 38.1±0.6% more deaths per capita.

Third, denser cities and metropolitan areas have significantly lower cumulative numbers of Covid-19 cases and deaths than their share in the overall population of the country. In fact, a city or metropolitan area with double the average population density of another one can be expected to have 4.3±0.1% fewer cases per capita and 7.4±0.6% fewer deaths per capita.

The effect of the population density of cities on the cumulative numbers of Covid-19 cases and deaths is brought to light when we account for the effects of the population and urbanized area first. That makes it possible to see that density has a positive, rather than a negative effect on the occurrence of both Covid-19 infections and deaths. This result in important in shifting the 'blame' for the prevalence of cases in cities to their overall population size, rather than to their densities, weakening the case for lowering urban densities—or, at the very least, not calling for their increase—in light of the current pandemic. Our conclusion is that the data on the current pandemic does not implicate urban density and does not, therefore, weaken the case for further densification of cities in the years to come.

* * *

i The graph for the week ending on 23 July 2020 is shown in figure A1 below.

Figure A1: The daily growth rate of Covid-19 cases in 384 U.S. MSAs during the week of 16-23 July 2020 as a function of the population of these MSAs, shown in logarithmic scale.