DISCUSSION PAPER SERIES

DP14842
(v. 2)

UNDERSTANDING SPATIAL VARIATION IN COVID-19 ACROSS THE UNITED STATES

Klaus Desmet and Romain Wacziarg

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

UNDERSTANDING SPATIAL VARIATION IN COVID-19 ACROSS THE UNITED STATES

Klaus Desmet and Romain Wacziarg

Discussion Paper DP14842
First Published 02 June 2020
This Revision 14 July 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

UNDERSTANDING SPATIAL VARIATION IN COVID-19 ACROSS THE UNITED STATES

Abstract

We analyze the correlates of COVID-19 cases and deaths across US counties. We consider a wide range of correlates - population density, public transportation, age structure, nursing home residents, connectedness to source countries, etc. - finding that these variables are important predictors of variation in disease severity. Many of the effects are persistent - even increasing - through time. We also show that there are fewer deaths and cases in counties where Donald Trump received a high share of the vote in 2016, partly explaining the emerging political divide over lockdown and reopening policies, but that this correlation is reversed when controlling for shares of minority groups. The patterns we identify are meant to improve our understanding of the drivers of the spread of COVID-19, with an eye toward helping policymakers design responses that are sensitive to the specificities of different locations.

JEL Classification: I18, R1

Keywords: COVID-19, spatial variation, US counties, Determinants

Klaus Desmet - kdesmet@smu.edu
Southern Methodist University and CEPR

Romain Wacziarg - wacziarg@ucla.edu
UCLA, NBER, University of California

Understanding Spatial Variation in COVID-19 across the United States*

Klaus Desmet
SMU, NBER and CEPR

Romain Wacziarg
UCLA and NBER

July 2020

Abstract

We analyze the correlates of COVID-19 cases and deaths across US counties. We consider a wide range of correlates - population density, public transportation, age structure, nursing home residents, connectedness to source countries, etc. - finding that these variables are important predictors of variation in disease severity. Many of the effects are persistent - even increasing - through time. We also show that there are fewer deaths and cases in counties where Donald Trump received a high share of the vote in 2016, partly explaining the emerging political divide over lockdown and reopening policies, but that this correlation is reversed when controlling for shares of minority groups. The patterns we identify are meant to improve our understanding of the drivers of the spread of COVID-19, with an eye toward helping policymakers design responses that are sensitive to the specificities of different locations.

*Desmet: Department of Economics and Cox School of Business, Southern Methodist University, 3300 Dyer, Dallas, TX 75205, kdesmet@smu.edu; Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, wacziarg@ucla.edu.

This is an updated version of the paper with the same title released on June 8, 2020 as *NBER Working Paper* #27329. No RAs were harmed in the writing of this paper. We thank Ricardo Perez-Truglia for useful comments.

1 Introduction

By April 8, 2020, 80% of US counties were covered by stay-at-home orders issued in response to the COVID-19 pandemic. Yet 50% of US counties had experienced five or fewer documented cases of the disease, and 72% of counties had experienced no deaths attributable to COVID-19. What is the source of heterogeneity in cases and deaths across US counties? Should policies be sensitive to such spatial variation? There are, we think, two legitimate views on these questions.

Under the first view, spatial variation in disease severity only reflects differences in timing. As the disease spreads, ultimately every location in the US will have similar infection rates, similar death rates, and similar rates of hospitalization. This view would justify uniform lockdown policies. Such policies would slow down disease spread to allow the health care infrastructure to cope with the disease burden.

Under the second view, spatial variation in cases and deaths reflects underlying fundamental differences across locations - population density, modes of transportation, housing arrangements, the age distribution, health conditions, weather, etc. At any point in time, locations will continue to differ according to these characteristics. They will differ no matter the number of days since onset, and the differences will persist, perhaps even increase over time. This provides a foundation for policies that are sensitive to local specificities, where less affected places can have less stringent lockdowns or earlier reopenings because their health care systems are less likely to become overwhelmed.

In this paper, we pinpoint the determinants of heterogeneity in COVID-19 cases and deaths, and provide evidence strongly consistent with the second view. We document substantial spatial heterogeneity across US counties, and identify novel and interesting correlates of variation in the number of cases and the number of deaths across US counties. We also analyze the persistence of these effects over time, finding that many of them have stable or even increasing effects as the disease spreads.[1]

We examine a broad set of correlates of disease severity. We pay particular attention to population density, using a variety of approaches to carefully measure dimensions of population density that have been hypothesized to affect the spread and severity of COVID-19. For instance, we look at the role of public transportation, living arrangements, housing density, and the distribution of the population at a high level of spatial resolution. We also consider the age distribution, racial composition, underlying health conditions, inequality and poverty, political orientation, among many other variables. A strength of our approach, unlike others that study putative determinants of COVID severity one at a time, is that we consider many potential correlates all at once.

Our analysis examines the role of these factors at various points in time, starting on March 15,

[1] An emerging literature examines the determinants of local variation in COVID-19 severity, also uncovering substantial spatial heterogeneity. Knittel and Ozaltun (2020) exploit cross-county variation in the US, like us, but only look at deaths and do not correct for differential timing in disease onset. Leamer (2020) studies cross-county variation within California, finding a significant effect of population density. McLaren (2020) looks more specifically at the relationship between COVID severity and racial composition, arguing that racial differences are partly related to differential prevalence of public transit at the county level in the US. Other papers study spatial variation for other countries, such as Belgium (Verwimp, 2020), France (Ginsburgh, Magerman and Natali, 2020) and England and Wales (Sá, 2020).

2020 and ending on June 29, 2020. We examine variation in COVID-19 cases and deaths on a daily basis using two approaches. The first approach looks at the cross-section of US counties at a given date, providing snapshots of the correlates of disease severity at particular moments in time. The second approach looks at the cross-section putting all counties at the same stage in terms of days since cases and deaths reached a certain threshold per capita. This allows us to correct for differences in the timing of disease onset, to better assess if spatial variation reflects variation in the timing of disease onset or fundamental differences between locations.

Our paper documents five major sets of facts. First, there is substantial variation in cases and deaths across counties. Second, this variation is associated with differences in a range of variables that capture population density, modes of transportation, urbanicity, the age structure of the population, the proportion of the population living in nursing homes, and distance to major airports with direct flights to countries where COVID-19 was prevalent early on. Third, the effects of these correlates persist through time, especially for variables that capture density, the presence of elderly individuals and nursing home population. Fourth, a deeper analysis uncovers additional correlates of disease severity: counties with many members of minority groups (especially African-Americans and Hispanics) are disproportionately impacted, as are counties with many poor people and a higher proportion of people with a bachelor's degree or more. Counties that imposed stay-at-home orders early on tend to have fewer deaths.

Fifth and finally, we document interesting facts when it comes to the political orientation of counties differentially affected by the disease. We find that the severity of the disease is politically patterned: in a simple specification controlling only for log population, counties with a high proportion of Trump voters in the 2016 general election have lower cases and deaths. These results may help explain the growing political divide over policies to ease stay-at-home orders. At the same time, the Trump effect on severity weakens and is even reversed when including additional controls for the shares of minority groups. Taking into account these factors, Trump's vote share in 2016 positively predicts cases and deaths. This could reflect differences in attitudes, policies and behaviors across counties on either side of the political divide.

2 The Correlates of COVID-19 Severity

In this section, we relate our empirical specification to standard epidemiological models, provide a brief overview of the data, and report our findings on the correlates of COVID-19 cases and deaths across U.S. counties.

2.1 Specification

Specification consistent with the SIRD model. Standard epidemiological models, such as the SIRD model, posit laws of motion of the number of susceptible people, infectious people, recovered people and deceased people for a given population and a given infectious disease. These laws of motion are governed by a few key parameters: the rate of infection, the rate of recovery and the rate of mortality. Together, they determine, for a given population, the evolution of the number of cases and deaths over time.

To fix ideas, denote by C_{it} the cumulative number of cases and by D_{it} the cumulative number of deaths from COVID-19 in county i at time t. The rate of infection, β_i, and the rate of death, δ_i, are likely to be, to an extent, county-specific. For example, we would expect counties with higher population density, where individuals are more likely to run into each other, to have a higher rate of infection β_i. Similarly, we would expect counties with a larger share of elderly to experience higher death rates δ_i. Differences in these parameter values across counties imply differences in the paths of C_{it} and D_{it} across counties. For example, a county with a higher β_i will have higher cumulative cases and deaths at any point time, compared to a similar county with a lower β_i. This is related to the well-known result that a higher expected number of infections from an infected individual (i.e., a higher basic reproduction number R_0) generates in the limit more cumulative cases and more cumulative deaths. Some of these insights are illustrated with simulations in the recent work by Fernández-Villaverde and Jones (2020).

The objective of this paper is to explore the importance of county-specific factors that affect β_i and δ_i. These parameters affect the dynamic paths of C_{it} and D_{it}, and therefore their levels at every point in time. We are interested in accounting for differences in levels of cumulative cases and deaths at a given point in time in the cross-section of counties. Hence we run, for each time period t, county-level regressions of the logarithm of cases or deaths on a set of potential determinants of β_i and δ_i:

$$\log(C_i) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij} + \varepsilon_i \tag{1}$$

and

$$\log(D_i) = \gamma_0 + \sum_{j=1}^{k} \gamma_j x_{ij} + \nu_i \tag{2}$$

where x_{ij} are county-level regressors that potentially affect β_i and δ_i (and hence C_{it} and D_{it}) and ε_i and ν_i are county-level disturbance terms. These k regressors, indexed by j, include variables such as a county's density, age structure and health conditions.

Note that these period-by-period regressions are able to capture any functional form for the path of the number of cumulative cases and deaths over time. As such, they are consistent with the functional forms generated by standard epidemiological models. Indeed, to allow for maximum flexibility in the changing relation between the county-level determinants and the disease severity, we choose a parsimonious period-by-period cross-sectional regression framework over a more structural empirical model that explicitly estimates the SIRD model.

The standard SIRD model assumes that individuals have equal probabilities of interacting with each other. In that sense, it does not really capture spatial features that make some individuals (or groups) more or less likely to interact with others. Bisin and Moro (2020) introduce a spatial SIR model with behavioral responses that explicitly incorporates these spatial concerns. When people are no longer matched randomly with the entire population, but are more likely to interact with people in their vicinity, local herd immunity becomes a possibility. In this model, spatial heterogeneity in disease severity can be magnified due to differences in modes of interaction and the spatial scale of interaction. As such, this model suggests that differences in population density are not sufficient to predict variation in disease severity, and that a proper analysis should take into account both modes of interaction as well as local effective density. This is precisely what we do below, by measuring a range of variables that capture the intensity of local interactions.

Timing and the definition of cross-sectional samples. We take two approaches to define the sample used in the cross-county analysis. The first approach is to carry out the analysis date by date. In this case, a time period t refers to a calendar date d, and we simply run regressions (1) and (2) day by day, from March 15, 2020 to June 29, 2020. A potential issue with this approach is that part of the cross-county variation in disease severity may be related to timing factors. To address this concern, we control for certain factors that could affect the timing of the arrival of COVID-19 to a particular county. For instance, we control for the distance to an airport with direct international flights to high-severity countries.

The second approach more directly addresses differential timing of onset by considering each county at the same time elapsed since onset. Here we refer to onset as the day when a county reached a certain threshold, either in terms of cases per capita or deaths per capita. To formally define days elapsed since onset, start by denoting, for each county i, an indicator variable I_{id}^{C} that takes a value of 1 if county i has reached at least 1 case per $100,000$ population on day d. For each county i and day d, the number of days since it reached that threshold is then:

$$s_{id}^{C} = \sum_{v=1}^{d} I_{iv}^{C}.$$

For the choice of each cross-county sample, we then set s_{id}^{C} to a fixed number t.[2] That is, the first sample consists of all counties one day after reaching the threshold, the second sample consists of all counties two days after reaching the threshold, and so on. Since each regression compares counties that all have passed the same threshold of per capita cases a fixed number of days before, this limits the effect of differential timing of onset across locations.

Similarly, we define the time elapsed since reaching the threshold of 0.5 deaths per $100,000$ population. For each county i and day d, the number of days since it reached that threshold is $s_{id}^{D} = \sum_{v=1}^{d} I_{iv}^{D}$, where I_{id}^{D} is an indicator variable taking a value of 1 if county i has reached at least 0.5 deaths per $100,000$ population on day d. Here as well, each regression compares counties that have passed the deaths per capita threshold a fixed number of days before.

Treatment of zeros. Counties with zero cases and zero deaths are particularly prevalent early in the sample period. Taking logs of cases and deaths amounts to ignoring the extensive margin. To address this shortcoming, we consider the log of one plus cases or deaths (resulting in a balanced sample of $3,137$ counties). For June 29, 2020, for instance, there were $3,045$ counties with strictly positive cases, and $1,934$ counties with strictly positive deaths. Including the extensive margin gives

[2] For instance, when fixing $t = 5$, the sample consists of each county on the specific calendar date d when it reached $s_{id}^{C} = 5$.

us the following specifications:

$$\log(1 + C_i) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij} + \varepsilon_i \tag{3}$$

$$\log(1 + D_i) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij} + \varepsilon_i \tag{4}$$

State fixed effects. Other policy choices and certain omitted variables may affect cumulative cases and deaths. To partly address this concern, in some specifications we include state fixed effects. In addition to picking up differences across states that go beyond the other variables we are already controlling for, we are also interested in the magnitude of these effects *per se*. However, we do not include state fixed effects in all specifications, as they absorb a lot of variation that we would prefer to explicitly capture.

Summary of specifications. To summarize, we have eight specifications. There are two outcomes: cases and deaths. There are two ways to construct the sample: by calendar date, using the log of one plus cases or deaths as dependent variables; or placing each county at the same time since onset for both deaths and cases (the latter excludes counties with zero deaths and cases by construction, and additionally excludes counties where the threshold defining onset has not been crossed).[3] Finally, there is another specification choice: whether we include state fixed-effects or not.

2.2 Data

We use daily data on COVID-19 reported cases and deaths collected at the county level by the *New York Times*. Appendix Table A1 (Panel A) contains summary statistics for various metrics of cases and deaths constructed from these data, revealing substantial variation across counties. To our knowledge these are the best data available at the county level, yet it is important to acknowledge several possible data challenges. These are particularly acute for cases, and early in the period, since reported cases depend on testing, and testing was initially far from uniformly and widely prevalent. Data issues are not absent from deaths data either, as reporting standards vary across jurisdictions and adjudicating whether a death was caused by COVID-19 involves an element of judgment. An alternative would be to use data based on excess mortality, but these are not available at the county level on a daily basis.[4]

Regarding measurement error, we note the following: First, if errors are random, they will raise the standard error of the regression without creating bias. However, if both testing and the reporting of deaths are systematically correlated with the included explanatory variables, we will need to interpret the corresponding estimates carefully as reflecting effects on both underlying severity and on reporting

[3] In the Appendix, we also consider a sample based on calendar dates, using the log of cases or deaths, i.e. only the intensive margin.

[4] The National Vital Statistics System of the National Center for Health Statistics reports weekly excess deaths at the state level: `https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm`. For other examples of excess deaths estimates, see New York City Department of Health and Mental Hygiene COVID-19 Response Team (2020) and Banerjee et al. (2020).

of cases and deaths. Second, to the extent that testing capacity varies at the state level, including state fixed effects may in part correct for systematic measurement error due to uneven testing intensity. Third, early in the spread of the disease, testing may also be more strongly targeted toward individuals showing symptoms, resulting is artificially high case fatality rates (CFR =deaths/cases). To address this possibility we reran our baseline regressions removing from the sample observations with $CFR >$ 0.1 - the upper tail of the distribution of CFR, most likely to be severely affected by selection in testing (Section 2.3 discusses the results). Fourth, testing and reporting regimes improve through time, so the passage of time should make measurement error in cases and deaths less relevant, as locations ramp up testing and fine tune the reporting of deaths.

We also gathered a wide range of county-level indicators to be used as independent variables. Variable definitions and sources are provided in the Data Appendix, summary statistics are in Appendix Table A1 (Panel B) and most of the variables are displayed in map form in Appendix Figure A1.

2.3 The Correlates of Spatial Variation in COVID-19 Severity

Tables 1 and 2 report estimates of all eight specifications outlined above. Table 1 considers a cross-section of counties as of June 29, 2020 (the last date in our sample). Table 2 reports estimates synchronizing the sample in terms of days since onset. For cases, we use 70 days since onset as the baseline and for deaths we choose 60 days since onset. These choices are motivated by a trade-off: by choosing a small number of days since onset, we would obtain a large cross-section of counties, less likely to be selected, but we would consider counties very close to onset, where the effect of fundamental determinants may not yet have emerged. Instead, by choosing a larger number of days since onset we would limit the number of counties in the sample in ways that are potentially selected, since only early onset counties are likely to appear. Our choice reflects this trade-off, and leads to a relatively large sample for both cases and deaths (respectively $2,755$ and $1,446$ counties).

We consider a set of eleven baseline correlates. The first is log population, which acts as a scaling variable. Its inclusion implies that the other estimates can be interpreted as the determinants of cases and deaths in per capita terms.

Density measures. A first group of regressors relates to population density, since living in closer proximity is likely to imply a higher infection rate β. Given the potential importance of density, we use several variables. One is simply population density as measured by the county's population divided by its land area. This may not adequately capture effective density, since some counties may have extensive land areas, in spite of most people living in fairly dense areas. We therefore complement simple density with variables that indicate whether a county is classified by the National Center for Health Statistics as a large metro area or as a medium or small metro area. In addition, we also include the share of the population that commutes by public transit, a factor that has been argued to be an important spreader of the virus (Harris, 2020).

Results are consistent across all specifications in showing the importance of density as a determinant of severity: all four density measures are jointly statistically highly significant and positively associated with the number of cases in all specifications. Looking at variables individually, we find that counties with a higher proportions of individuals using public transit have significantly higher severity, with large standardized magnitudes particularly for deaths ($12-16\%$). Magnitudes are sometimes reduced when including state fixed effects, but remain broadly consistent. Deaths are higher in large metro counties than in medium or small metro counties, which in turn tend to be higher than in the excluded category of non-urban counties. The effect of log population density itself tends to be positive, but is not consistently significant across specifications. This finding highlights the importance of properly measuring effective density using a variety of metrics, a task we further pursue in Section 4.1, where we include additional measures of effective density, based on housing arrangements and on the density experienced by an average individual in the square kilometer grid cell where they live.

Age and nursing homes. A second group of regressors relates to the age structure of the population. Given the much higher mortality rate among the elderly, we control for the share of the population aged 75 and above. It is important to note that the age gradients of cases and deaths may be quite different from each other (Hay et al., 2020, report data on the age gradient of infections rather than deaths). As is often observed, the elderly living in nursing home may be particularly susceptible (Barnett and Grabowski, 2020). We therefore also include a county-level measure of nursery home residents divided by population.

We find interesting results. Cases are negatively associated with the percentage of people aged 75 and older. This may reflect differences in lifestyles between counties with different age structures. For instance, places with a large share of retired individuals may feature fewer places (bars, stadiums) where the disease spreads rapidly. On the other hand, we find no consistent pattern regarding the correlation between age structure and deaths. It is well established that deaths from COVID-19 disproportionately occur among older individuals, but this does not necessarily imply that counties with a greater proportion of elderly persons experience a higher number of deaths, after controlling for other determinants. Indeed, as discussed, counties with a greater share of people aged 75 and above have a *lower* number of cases. When it comes to the share of the population in nursing homes, we find positive and economically large partial correlations especially for deaths, and especially when isolating the intensive margins of the disease. For instance in columns (3) of Tables 2 and A2, the standardized betas on the share of nursing home residents are respectively 15.4% and 13.6%. This finding is consistent with the idea that once a county is affected by the pandemic, its nursing homes can quickly become powder kegs, and account for large shares of county-wide deaths.

Other correlates. A third group of regressors include other factors that have been hypothesized to affect the onset and severity of the pandemic. Early reports suggested that temperature may play a role in the spread of the disease, so we include a county-level measure of the average temperature in February, March and April (using data from China, Qi et al., 2020, suggested that higher temperatures slowed the disease, but Xie and Zhu, 2020, find a flatter temperature gradient). We find evidence that locations with higher temperatures in those months experienced more cases and deaths, with sometimes large standardized magnitudes. The implications for the evolution of the disease in the summer months are unclear, since both the absolute level of temperature and its spatial distribution will change.

The onset of the pandemic in specific locations in the US may have been related to connectivity with high-severity countries (Wells et al., 2020). We construct a measure of the distance to any airport with direct flights to one of the top-5 countries with coronavirus cases on March 15, 2020 (China, South Korea, Iran, Italy and Spain). This variable bears a consistently negative relationship with cases and deaths.

Among the remaining correlates, we first include median household income, a standard metric to capture differences in economic well-being across counties. We do not find a robust effect of median income across specifications. Second, a measure of social capital from Rupasingha, Goetz and Freshwater (2006), bears a positive relationship with cases and deaths in Table 1, but this result does not hold up when looking at the intensive margin only (Table 2).[5]

State fixed-effects. Tables 1 and 2 report results with or without state fixed effects. Appendix Figures A2 and A3 graphically display estimates on the state fixed effects, ordered by size, for the specifications of columns (2) and (4) of Table 1. These plots reveal that, after controlling for the eleven baseline set of correlates of disease severity, some states have lower or higher cases or deaths. We find that counties in Hawaii and California, for instance, have lower severity than expected, while counties in Louisiana, Connecticut or New Jersey have higher severity than expected. These differences could reveal idiosyncrasies that are hard to capture using additional regressors varying at the county level (for instance the fact that Hawaii is an island, or that New Jersey is close and tightly integrated with New York, a major center of the disease in the US). They could also capture some omitted factors excluded from our parsimonious specification. At any rate the inclusion of state fixed-effects does not seem to greatly affect the patterns uncovered regarding the measured determinants of disease severity.

Isolating the intensive margin. In Table A2, we consider the determinants of log cases and log deaths as of June 29, 2020, i.e. counties with zero cases or deaths drop from the sample. The results confirm the findings discussed above, indeed Table A2 resembles Table 1. Results differ a little more for deaths than for cases, because by June 29 most counties in the US reported positive cases, whereas over $1,000$ counties still did not report a single death. Differences between Tables 1 and A2 are expected to fade as time goes by as the samples will overlap more and more.

Incidence of high CFR counties. Some counties in our sample exhibit very high case fatality rates (CFR), especially early in the period. This is perhaps because testing was limited, and selected to apply mostly to individuals showing severe COVID-19 symptoms. As testing became more widespread, this source of bias was likely reduced. To examine the robustness of the results to the inclusion of counties where testing was biased in this manner, we rerun our baseline regressions removing observations with $CFR > 10\%$. This also implies removing counties with zero cases. Comparing Tables A2 and 2 to Tables A3 and A4 (the sample restriction applies to the latter), we find only very minor differences in the estimates. These results mitigate the concern that bias in testing only symptomatic individuals

[5] For a further investigation of the ambiguous role of social capital as a determinant of social distancing, see Ding et al. (2020), who find a negative effect of community activities but a positive effect of voter turnout. Durante, Guiso and Gulino (2020), across Italian provinces, find that mobility declined more in areas with higher civic capital.

drives our results. Moreover, as time goes by and testing becomes less and less selected, the concern should also be alleviated.

3 Persistence in the Determinants of COVID-19 Heterogeneity

The foregoing discussion concerned the cross-section of disease severity at a specific date (June 29) or at a constant time since onset. These effects offer a snapshot of spatial variation, but do not describe how the partial correlations that we calculated evolve over time. As the disease progresses, do these sources of heterogeneity in severity persist?

To examine this question, we estimate our model daily and plot estimated coefficients and their confidence bounds through time. It is important to emphasize that this also represents a time-slice of the effects. Indeed, we do not know how they will further change past the last date in our dataset (currently June 29) but we will update the results as more data becomes available.

Evolution between March and June 2020. Figure 1 displays coefficient estimates from the specifications of equations (3) and (4), with 95% confidence intervals. The sample of counties is the same over time ($3,137$ counties) and the dates run from March 15 to June 29, 2020. In most cases, we see an initial period where coefficient magnitudes move away from zero. This is natural since there is not much variation to explain early on, and there is randomness in locations that got the virus early. One important exception is the variable capturing distance to international airports with connections to the top-5 COVID incidence countries as of March 15, 2020. This variable predicts the cross-section of cases from the get-go, as we would expect.

Many of the 11 regressors display increasing absolute effects over time. When focusing on the effect of density on cases, it grew over time until about day 80, and has been slightly weakening since. The other three measures of density display weakening effects. Public transit usage and urban categories display persistent positive effects of deaths. Looking at all four measures of density jointly, we see that there is so far no indication that density is disappearing as a predictor of the cumulative number of cases and deaths.

Turning to the elderly population, our results echo what we found previously: the share of the population aged 75 and above is negatively and persistently correlated with cases, but the earlier positive correlation with deaths has faded since early May. As for the share of the population living in nursing homes, its impact is positive and rising over time, both for cases and for deaths.

Other correlates deserve a brief mention. The distance to the closest international airport with direct flight connections to high-incidence countries is negatively correlated with both cases and deaths, and those correlations are stable over time, showing the persistent effect of initial conditions. Median household income initially bore a slight positive correlation with cases, which disappeared gradually since early May, but it is uncorrelated with deaths. A last correlate worth discussing is log population. We observe that the elasticity of cases to population rises over time and reaches one by late June, suggesting no scale effects. For deaths, the elasticity does not reach one by the end of the period, suggesting that there still exists a negative scale effect on per capita deaths.

Overall, many of the location-specific characteristics that affect the rate of infection and the rate of death, such as population density and age composition, display persistent and sometimes increasing (absolute) correlations with cumulative cases and deaths. As such, the evidence so far suggests that the severity of COVID-19 is unlikely to equalize across space. Whether these findings hold up as the pandemic further unfolds remains an open question.

Evolution since onset. One possible issue with Figure 1 is that the coefficients may partly pick up the differential timing of onset across different types of counties. For example, if low-density counties are hit later by COVID-19 than high-density counties, then their cumulative cases or deaths will tend to be lower on any given date. Of course, if timing were the main difference between low and high density counties, then the coefficient on density should be declining over time, as disease severity in low density counties catches up with high density counties. Since many of the regressors display increasing absolute effects over time, it is unlikely that differential timing is an important driver of our results.

However, to limit any impact of differential timing, we fix the sample in terms of days since onset. Figure A4 displays how coefficient estimates evolve as a function of days since onset. To grasp how to read these graphs, a concrete example may help. The public transit graph in Figure A4A plots the coefficients on public transportation from 90 different regressions, one for each of the different time lags since a county reached the threshold of 1 case per $100,000$. Increasing the number of days since onset decreases the sample size because fewer counties meet the criterion for passing the threshold early on. We illustrate this changing sample size among the graphs displayed in Figure A4. As can be seen, there are over $3,000$ counties in the sample of counties one day after passing the case threshold, but there are about $2,200$ in the sample of counties 90 days after onset.

As before, we find strong evidence of persistence regarding many determinants of cases and deaths. For example, the importance of density for cases grows as the pandemic runs its course in a given location, and public transit shows a persistent effect on both cases and deaths. As for nursing home residents, its correlation with cases and deaths is also persistent and increasing in the days since onset. The only determinants of both cases and deaths that seem persistently insignificant are median income (a variable that did not bear a robust relationship with cases and deaths in Section 2) and social capital. As we would expect, in the early days since onset coefficients on the different regressors tend to be close to zero.[6] In sum, whether defining the sample by calendar dates or by days since onset, we find substantial persistence in the determinants of spatial variation in disease severity.

4 Further Investigation of Specific Correlates

4.1 Effective Density

Our baseline results indicate an important role for density in determining the severity of COVID-19. This should come as no surprise: as with any other infectious disease, contact between susceptible and

[6] In the limit, on the first day of reaching the threshold, we are comparing counties that are identical in terms of the variable we are trying to explain. In the absence of any cross-sectional variation, we would not expect any of the regressors to explain anything.

infected individuals is a key determinant of the spread of the disease. However, the actual degree of contact between people is not straightforward to measure. The four indicators already included in the baseline specification may not fully capture relevant dimensions of density.

In Table A5, we continue to control for the baseline set of 11 determinants, but add three additional measures aimed at better capturing the likely intensity of contact between people. Two of these relate to housing and living arrangements: the share of individuals living in multi-unit housing structures and the number of people per household. A third measures the average density a random individual of a county experiences in the square kilometer around him. We refer to this variable as a county's "effective local density". Columns (1) and (4) of Table A5 report coefficient estimates for specifications where we add the controls for living arrangements. We see that multi-unit housing and the size of households are positively associated with both cases and deaths.[7] Columns (2) and (5) add effective local density: its correlation with cases is statistically insignificant, whereas its correlation with deaths is negative and significant. For reasons of further comparison, columns (3) and (6) drop housing arrangements and public transportation, and only maintain simple density and effective local density. As can be seen, effective local density now displays a positive and statistically very significant relation with both cases and deaths. Overall, this suggests that a county's effective local density matters, but that its effect may operates through dense housing and public transit.

4.2 Other Factors

Race. Table A6 explores the possible role of race. It reports four different specifications: columns (1) and (3) report regressions for cases and deaths, based on a cross-section of counties as of June 29, whereas columns (2) and (4) also report regressions for cases and deaths, but now based on a cross-section of counties 70 days after onset (for cases) and 40 days after onset (for deaths). To the baseline regressors, we add measures of the racial composition of a county by controlling for the shares of African Americans, Hispanics, American Indians and Asians, with the excluded category being the share of Whites and others. The results display a strong and consistent positive correlation between the share of African Americans and the share of Hispanics with both the number of cases and the number of deaths. The share of American Indians exhibits a positive correlation with deaths, but not with cases, whereas the share of Asians shows a weaker correlation with COVID-19 incidence. In terms of magnitudes, the share of African Americans stands out with large standardized β coefficients between 25% and 33%. Overall these results confirm concerns that the COVID-19 pandemic has a disparate effect on various racial groups.

Education. Table A7 analyzes whether the level of education may be a source of heterogeneity in disease severity across counties. We take the same four specifications as in the previous table with the same baseline regressors, and add two controls for the level of education: the share of a county's population that has a high school degree or more and the share of a county's population that has a bachelor's degree or more (the excluded variable is the share of people with less than a high school

[7] Public transit continues to be highly significant for log(1+deaths), but its coefficient in the regression for log(1+cases) is sensitive to the inclusion of the percentage of housing units in multi-unit structures. Indeed, the percentage of people who use public transportation to commute is highly correlated with dense housing arrangements ($\rho = 0.6$).

degree). We find a non-monotonic relationship between average educational attainment, and disease severity. Counties with large proportions of high school graduates fare best, followed by counties with a large share of individuals without a high school degree. Places with many college graduates fare the worst. Hence, we find little evidence that more disadvantaged locations (measured by education) fare worse. These correlations, while informative, remain open for interpretation.

Health. Table A8 investigates whether underlying health conditions or the quality of health care have an impact on outcomes. As measures of underlying health issues, we take the share of the population that smokes and the share of the population that is obese. As measures of quality of health care, we take the risk-adjusted 30-day mortality rates for heart attacks, heart failure and pneumonia. The share of smokers and obese people does not seem to be a consistent driver of heterogeneity in COVID-19 incidence across counties, though we find some evidence that the share of smokers is associated with lower cases and the share of obese persons is positively correlated with deaths. Turning to risk-adjusted mortality rates, we find some evidence that risk-adjusted mortality from pneumonia is positively correlated with deaths, suggesting a role for the quality of the health infrastructure (on the other hand the signs of the correlations on risk-adjusted mortality from heart attacks and heart failures often have the opposite signs from what is expected, and are small in magnitude). These results tend to be sensitive to the inclusion of more controls, as we show using a more comprehensive specification discussed below. In sum, we find only weak evidence that often-hypothesized health drivers of COVID-19 severity - either the prevalence of underlying health conditions or the quality of the health care infrastructure - are first-order determinants of cross-county variation in cases and deaths.

Inequality and Poverty. Table A9 reports results of an in-depth investigation of the role of inequality and poverty. In the baseline regressions we already included median household income. We add three measures that capture inequality and poverty: the Gini index within the bottom 99%, the poverty rate, and the top 1% income share. The share of top incomes is insignificant, the Gini index among the lower 99% is positive in column (1) but not in the other specifications, and poverty positively predicts severity measured both by deaths and cases. The results are quantitatively meaningful: for example, the poverty rate shows standardized coefficients in the range of 16% to 28% when considering its impact on deaths. In sum, we find evidence that poverty (but not inequality) is a significant determinant of deaths.

4.3 Comprehensive Specification

In Tables A5-A9 we entered new categories of variables one by one. Do the main results hold up when all the putative determinants of cases and deaths are entered jointly? To answer this question we examine results from a comprehensive specification that includes not only the baseline set of 11 correlates of disease severity, but also most of the additional regressors considered in Tables A5-A9 (two exceptions are the share of people who are obese and the share of people who smoke, because their inclusion would result in the loss of close to one third of the sample). The results are reported in Table A10. The estimates broadly conform to the results obtained earlier: our various measures of density are positively related to severity, as are the share of county residents living in nursing homes and the shares of African-Americans and Hispanics.

4.4 Stay-At-Home Orders

So far, we focused on time-invariant county determinants of the incidence of COVID-19. Some determinants may change over time. The prime example here is stay-at-home orders. These are aimed at reducing the rate of infection, and hence slowing down the increase in cases and deaths. Needless to say, identifying the causal effect of stay-at-home orders is fraught with difficulty, since the local severity of the disease is likely to prompt earlier policy intervention.[8] Arguably, such endogeneity concerns are somewhat mitigated when fixing the sample in terms of days since onset.[9] In that case, we are comparing counties with identical initial conditions in terms of cases or deaths per capita, but possibly different dates at which stay-at-home orders were imposed. Table A11 focuses only on specifications where the sample is chosen based on reaching a specific threshold of cases and deaths, as defined previously. We consider specifications with or without state fixed-effects, to exploit within-state variation in the timing of stay-at-home orders. We include a variable describing the number of days since the first stay-at-home order applied to a particular county. As of April 30, all but 631 counties were under stay-at-home orders. Among those, the average number of days since the order was issued was 26, and extended up to 44 days. We find no significant effect of this variable on cases, but a statistically significant and economically meaningful negative effect on deaths.[10]

Figure A5 depicts the coefficient estimates of the stay-at-home orders, defined as the number of days since the first stay-at-home order was implemented in a county. The regression specifications are identical to those in Figure A4, the difference being that we control for stay-at-home orders. There is a slight positive correlation between the length of stay-at-home orders and the number of cases, but it is only statistically significant during the first ten days after reaching the threshold of 1 case per $100,000$. In contrast, there is a negative correlation between the duration of stay-at-home orders and the number of deaths, and it remains statistically significant during much of the time period. The correlation fades to zero past day 70 or so, because the relatively small set of counties that had an early onset of deaths also tended to adopt stay-at-home orders early on. Thus, there is not much variation in days since stay-at-home-orders for that small and selected sample of counties.[11]

[8] Indeed, optimal lockdown policies are likely to differ across locations. For an in-depth investigation of optimal spatial lockdown policies, see Fajgelbaum et al. (2020).

[9] Of course, the concern is not eliminated. For two counties with identical days since onset, some unobserved factor may drive both disease severity and the decision to issue stay-at-home orders. Since the policy is not randomly assigned, the endogeneity concern is hard to fully address.

[10] Several studies look at the effectiveness of lockdown policies. Jinjarak et al. (2020) show that countries with stricter policies to limit social contact had later and less pronounced disease peaks. Dave et al. (2020) adopt an event study approach finding large effects especially among early adopters and dense locations. Kapoor et al. (2020) use rainfall shocks to identify the effect of staying at home, finding that social distancing has a persistent negative effect on cases and deaths across US counties. Lin and Meissner (2020) find that stay-at-home orders reduced mobility but not COVID-19 cases, looking at counties on either side of state borders.

[11] For instance, when the number of days since reaching 0.5 deaths per $100,000$ population is 80, there are only 959 counties in the sample, only 90 of which have no stay-at-home orders in place.

5 Spatial Patterns and Political Orientation

Many commentators have observed that there exists a political divide over attitudes toward the COVID-19 pandemic (see for instance Pew Research Center, 2020). In turn, these disagreements may reflect underlying differences in disease severity across locations with different political orientations. Weniger and Ou (2020) and Kolko (2020a, 2020b) observe that the disease is more severe in Democratic-leaning states and counties than in Republican-leaning locations.[12] Does severity indeed vary according to local political orientation? In this subsection, we try to better understand the political divide in disease severity.

We start by observing that indeed, locations that voted for Donald Trump in the 2016 presidential election had lower cases and deaths on June 29, 2020. Column (1) of Table 3 Panel A reports that the coefficient on the Trump vote share is negative, statistically significant and large in magnitude (with a standardized beta of 11%) in a regression explaining log(1+cases), controlling only for log county population. Column (1) of Table 3 Panel B finds the same for deaths (with a standardized beta of 15.1%). The left-side panels of Figure 2 confirm these findings and extend them over time by plotting the average residuals from regressions of log(1+cases) and log(1+deaths) on log population, since March 15, for jurisdictions with different political orientations.[13] We see a large political divide for both cases and deaths, which persists over time and even increases when it comes to deaths. Obviously, these patterns do not represent a causal effect of political orientation on disease severity. Rather, they simply suggest that disease severity is geographically patterned according to political orientation. These results may help explain the observed political fault lines over the desirability of lockdown policies, with Republican-leaning locations seemingly much more eager to reopen early and suspend the lockdowns as compared to Democratic-leaning locations.

What might explain this spatial pattern? The remaining columns in Panels A and B of Table 3 investigate this question using the cross-section of counties as of June 29.[14] By including our baseline set of control variables (those in Tables 1 and 2), the second column of Table 3 displays a statistically unchanged effect of Trump vote share. However, the third column adds the shares of various racial groups (African-Americans, Hispanics, Asians, Native Americans), leading to a sign flip in the Trump vote share effect. Finally the fourth column adds all the variables included in the comprehensive specification discussed in Section 4.3, leading still to a positive effect of the Trump vote share. To further explore these patterns, the right-side panels of Figure 2 display the average residuals from

[12] Of course, preferences for lockdown policies are not solely determined by spatial patterns of disease severity. Ideological predilections and media influence may also play a role in the emerging political divide over the response to COVID-19. See for instance Bursztyn et al. (2020) and Allcott et al. (2020).

[13] Red counties are defined as those with a 2016 Trump vote share greater than 55%, blue counties are those with a Trump vote share smaller than 45%, and purple counties represent the balance.

[14] Table A12 carries out similar regressions using the set of counties 70 days from onset (for cases - Panel A) and 60 days from onset (for deaths - Panel B). Column 4 of this table additionally includes a regressor representing the number of days a stay-at-home order has been in place. We find results very similar to those in Table 3. The inclusion of length of stay-at-home orders does not reduce the coefficient on Trump vote share, suggesting that it is not because Trump-leaning counties put in place stay-at-home orders at different times that they experienced more cases and deaths. Beyond policies, it could be individual behaviors to avoid infection that differ across political orientations.

the comprehensive specification, by county political orientation (red, purple and blue). We uncover interesting patterns. We first confirm that at the end of the sample period, Democratic counties experience lower disease severity than Republican counties, consistent with the regressions in the fourth column of Table 3. We also show that this lower severity is the result of a reversal: even after controlling for an exhaustive set of determinants of disease severity, the average residual in the regression for log(1+cases) is higher in Democratic-leaning counties than in Republican-leaning ones until about April 15, and in the regression for log(1+deaths) until May 15. This pattern reversed after these dates in such a way that Republican areas, controlling for other determinants of cases and deaths, now experience worse disease severity.

We can only speculate as to why the sign of the Trump effect flipped when adding more controls, and why the time pattern of partisan severity was also reversed over time. It is possible that public policies and individual behaviors regarding the spread of COVID-19 in Republican-leaning areas became more lax relative to Democratic-leaning areas, so that after controlling for major determinants of disease severity like racial composition and effective density, areas that voted for Donald Trump actually started to fare worse.

6 Conclusion

In this paper, we study heterogeneity in the severity of the COVID-19 pandemic across counties of the United States. We explore a wide range of correlates of severity jointly, in a unified estimation framework that allows for the inclusion of state fixed effects, controls for the differential timing of disease onset in various locations, and accommodates variation on both the intensive and extensive margins of cases and deaths. We document a strong and persistent role for population density, captured using a variety of metrics, as a correlate of cases and deaths. We argue that it is important to measure density correctly, using indicators of urbanicity, prevalent modes of transportation, household size and housing arrangements, and local effective density. We also show that the age structure and the proportion of people living in nursing homes are powerful and persistent predictors of disease severity, particularly the number of deaths. We explore correlations with a wide range of additional variables, finding for instance that minorities are more severely affected by the pandemic. Controlling for the timing of disease onset, more days spent under stay-at-home orders negatively predicts the number of deaths across counties. Finally, we find that areas with a large share of Trump voters are less severely affected by COVID-19, but that this effect is reversed when controlling for variables that are correlated with both Trump support and disease severity, in particular the shares of different minority groups. Once controlling for these, Trump-oriented counties are actually more severely affected by the disease. Many of these effects rise between March 15 and June 29, and remain statistically significant as of the end of our sample period. Time will tell whether this persistence persists.

Overall, our results suggest that policymakers should be sensitive to the specificities of different locations when designing policy responses to the spread of COVID-19, and their unwinding.

References

[1] Allcott, H., L. Boxell, J. C. Conway, M. Gentzkow, M. Thaler, D. Y. Yang (2020), "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic," *NBER Working Paper* #26946.

[2] Banerjee, A., L. Pasea, S. Harris, A. Gonzalez-Izquierdo, A. Torralbo, L. Shallcross, M. Noursadeghi, D. Pillay, N. Sebire, C. Holmes, C. Pagel, W. Keong Wong, C. Langenberg, B. Williams, S. Denaxas, and H. Hemingway (2020), "Estimating Excess 1-year Mortality Associated with the COVID-19 Pandemic According to Underlying Conditions and Age: A Population-Based Cohort Study," *The Lancet*, May 12.

[3] Barnett, M. L. and D. C. Grabowski (2020), "Nursing Homes Are Ground Zero for COVID-19 Pandemic," *JAMA Health Forum*, 1(3): e200369-e200369.

[4] Bisin, A. and A. Moro (2020), "Learning Epidemiology by Doing: The Empirical Implications of a Spatial SIR Model with Behavioral Responses", *Working Paper*, New York University.

[5] Bursztyn, L., A. Rao, C. Roth, D. Yanagizawa-Drott (2020), "Misinformation During a Pandemic", *NBER Working Paper* #27417.

[6] Dave, D., A. I. Friedson, K. Matsuzawa, J. J. Sabia (2020), "When Do Shelter-in-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across States and Adoption Time", *NBER Working Paper* #27091.

[7] Desmet, K. and R. Wacziarg (2020), "Understanding Spatial Variation in COVID-19 Across the United States", *NBER Working Paper* #27329, released on June 8.

[8] Ding, W., R. Levine, C. Lin, W. Xie (2020), "Social Distancing and Social Capital: Why U.S. Counties Respond Differently to COVID-19", *NBER Working Paper* #27393.

[9] Fernández-Villaverde, J. and C. I. Jones (2020), "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities," *NBER Working Paper* #27128.

[10] Ginsburgh, V., Magerman, G., Natali, I. (2020), "COVID-19 and the Role of Economic Conditions in French Regional Departments", *Working Paper*, ECARES, June.

[11] Harris, J. E. (2020), "The Subways Seeded the Massive Coronavirus Epidemic in New York City," *NBER Working Paper* #27021.

[12] Hay, J. A., D. J. Haw, W. P. Hanage, C. Jessica, E. Metcalf and M. J. Mina (2020), "Implications of the Age Profile of the Novel Coronavirus," Web Link.

[13] Israel, A., I. Feldhamer, A. Lahad, D. Levin-Zamir, G. Lavie (2020), "Smoking and the Risk of COVID-19 in a Large Observational Population Study", MedRxiv, doi: https://doi.org/10.1101/2020.06.01.20118877.

[14] Jinjarak, Y., R. Ahmed, S. Nair-Desai, W. Xin, and J. Aizenman (2020), "Accounting for Global COVID-19 Diffusion Patterns," *NBER Working Paper* #27185.

[15] Kapoor, R., H. Rho, K. Sangha, B. Sharma, A. Shenoy, G. Xu (2020), "God is in the Rain: The Impact of Rainfall-Induced Early Social Distancing on COVID-19 Outbreaks", *COVID Economics, Vetted and Real-Time Papers*, 24(1).

[16] Knittel, C. R., B. Ozaltun (2020), "What Does and Does Not Correlate with COVID-19 Death Rates", *NBER Working Paper* #27391, released on June 22.

[17] Kolko, J. (2020a), "The Changing Geography of COVID19", *Blog Post* (Web Link), June 21.

[18] Kolko, J. (2020b), "Where COVID19 Death Rates are Highest", *Blog Post* (Web Link), April 15, updated on May 13.

[19] Leamer, E. (2020), "What Explains the Large Differences in Rates of COVID-19 Infections Among California Counties?" *Working Paper*, UCLA.

[20] Lin, Z. and C. M. Meissner (2020), "Health vs. Wealth? Public Health Policies and the Economy During Covid-19", *NBER Working Paper* #27099.

[21] McLaren, J. (2020), "Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census Data", *NBER Working Paper* #27407, released on June 22.

[22] New York City Department of Health and Mental Hygiene COVID-19 Response Team (2020), "Preliminary Estimate of Excess Mortality During the COVID-19 Outbreak - New York City, March 11-May 2, 2020," *MMWR Morb Mortal Wkly Rep*, 69: 603-605.

[23] Pew Research Center (2020), "Republicans, Democrats Move Even Further Apart in Coronavirus Concerns", Web Link, June 25.

[24] Qi, H., S. Xiao, R. Shi, M. P. Ward, Y. Chen, W. Tu, Q. Su, W. Wang, X. Wang and Z. Zhang (2020), "COVID-19 Transmission in Mainland China is Associated with Temperature and Humidity: A Time-Series Analysis," *Science of The Total Environment*, 728(1).

[25] Rupasingha, A., S. J. Goetz and D. Freshwater (2006, with updates), "The Production of Social Capital in US Counties," *Journal of Socio-Economics*, 35, 83-101.

[26] Sá, F. (2020), "Socioeconomic Determinants of COVID-19 Infections and Mortality: Evidence from England and Wales", *CEPR Discussion Paper* #14781, May 19.

[27] Verwimp, P. (2020), "The Spread of COVID-19 in Belgium: a Municipality-Level Analysis", *ECARES Working Paper* #2020-25, released in July.

[28] Weniger, B. G. and C.-Y. Ou (2020), "Straight Talk from ex-CDC for the Long Slog Ahead", *Blog Post* (Web Link), May 3.

[29] Wells, C. R., P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer and A. P. Galvani (2020), "Impact of International Travel and Border Control Measures on the Global Spread of the Novel 2019 Coronavirus Outbreak," *Proceedings of the National Academy of Sciences of the USA*, 117(13): 7504-7509. Web Link.

[30] Xie, J. and X. Zhu (2020), "Association Between Ambient Temperature and COVID-19 Infection in 122 Cities from China," *Science of The Total Environment,* 724(1).

Table 1 - OLS Regressions for log 1+Cases and log 1+Deaths, June 29, 2020
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log 1+Cases	**Log 1+Cases, State FE**	**Log 1+Deaths**	**Log 1+Deaths, State FE**
Log population	1.005 (0.026)*** [0.704]	1.033 (0.029)*** [0.723]	0.694 (0.026)*** [0.627]	0.725 (0.031)*** [0.656]
Log population density	0.106 (0.022)*** [0.088]	0.065 (0.028)** [0.054]	0.005 (0.022) [0.006]	-0.044 (0.029) [-0.048]
Large central metro county or large fringe metro county	0.048 (0.072) [0.008]	0.011 (0.065) [0.002]	0.671 (0.075)*** [0.141]	0.603 (0.067)*** [0.127]
Medium metro county or small metro county	0.056 (0.051) [0.011]	0.006 (0.045) [0.001]	0.180 (0.052)*** [0.046]	0.166 (0.047)*** [0.042]
% people who commute by public transportation	0.029 (0.006)*** [0.042]	0.024 (0.006)*** [0.035]	0.083 (0.007)*** [0.155]	0.070 (0.007)*** [0.130]
Share of people aged 75 & above	-12.061 (0.974)*** [-0.132]	-10.896 (0.955)*** [-0.119]	-0.199 (1.003) [-0.003]	1.098 (0.991) [0.016]
% nursing home residents in pop.	0.208 (0.045)*** [0.044]	0.077 (0.043)* [0.016]	0.186 (0.046)*** [0.050]	0.068 (0.044) [0.018]
Log km to closest airport w/ flights from top 5 COVID countries	-0.059 (0.019)*** [-0.031]	-0.059 (0.019)*** [-0.032]	-0.062 (0.020)*** [-0.043]	-0.093 (0.019)*** [-0.065]
Average temperature, Feb., Mar. & Apr.	0.025 (0.002)*** [0.123]	0.026 (0.005)*** [0.126]	0.012 (0.002)*** [0.075]	0.016 (0.005)*** [0.104]
Log household median income	-0.008 (0.099) [-0.001]	-0.044 (0.100) [-0.005]	-0.082 (0.102) [-0.012]	-0.159 (0.103) [-0.023]
Social Capital Index, 2014	0.045 (0.019)** [0.026]	-0.010 (0.019) [-0.006]	0.054 (0.020)*** [0.041]	0.055 (0.020)*** [0.042]
Constant	-6.284 (1.105)***	-5.361 (1.146)***	-5.358 (1.138)***	-4.534 (1.188)***
R^2	0.79	0.84	0.62	0.71
N	3,137	3,137	3,137	3,137

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets.

Table 2 - OLS Regressions for log Cases and log Deaths, Synchronized Days from Onset at 70 days from Onset (for log cases) and 60 days from Onset (for deaths)

	(1)	(2)	(3)	(4)
	Log Cases	**Log Cases, State FE**	**Log Deaths**	**Log Deaths, State FE**
Log population	0.861 (0.030)*** [0.605]	0.957 (0.036)*** [0.673]	0.787 (0.044)*** [0.605]	0.882 (0.053)*** [0.678]
Log population density	0.175 (0.026)*** [0.144]	0.059 (0.034)* [0.049]	0.071 (0.040)* [0.061]	-0.014 (0.050) [-0.012]
Large central metro county or large fringe metro county	0.166 (0.082)** [0.031]	0.144 (0.074)* [0.027]	0.484 (0.112)*** [0.120]	0.447 (0.101)*** [0.111]
Medium metro county or small metro county	0.040 (0.058) [0.009]	0.012 (0.052) [0.003]	0.042 (0.083) [0.011]	0.020 (0.075) [0.006]
% people who commute by public transportation	0.042 (0.007)*** [0.071]	0.035 (0.007)*** [0.059]	0.063 (0.008)*** [0.158]	0.048 (0.008)*** [0.121]
Share of people aged 75 & above	-10.662 (1.204)*** [-0.118]	-9.726 (1.195)*** [-0.107]	-0.112 (1.675) [-0.001]	0.372 (1.745) [0.004]
% nursing home residents in pop.	0.292 (0.062)*** [0.059]	0.099 (0.061) [0.020]	0.877 (0.114)*** [0.154]	0.553 (0.114)*** [0.097]
Log km to closest airport w/ flights from top 5 COVID countries	-0.072 (0.021)*** [-0.044]	-0.065 (0.021)*** [-0.040]	-0.048 (0.024)** [-0.041]	-0.058 (0.024)** [-0.048]
Average temperature, Feb., Mar. & Apr.	0.013 (0.002)*** [0.070]	0.018 (0.006)*** [0.098]	0.001 (0.003) [0.009]	0.031 (0.009)*** [0.180]
Log household median income	-0.109 (0.118) [-0.014]	-0.238 (0.119)** [-0.030]	-0.223 (0.164) [-0.034]	-0.346 (0.166)** [-0.052]
Social Capital Index, 2014	0.034 (0.024) [0.020]	-0.007 (0.025) [-0.004]	-0.051 (0.035) [-0.028]	-0.014 (0.035) [-0.008]
Constant	-4.039 (1.313)***	-2.908 (1.373)**	-5.032 (1.839)***	-5.631 (1.932)***
R^2	0.69	0.76	0.58	0.69
N	2,755	2,755	1,446	1,446

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets. Onset day is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).

Table 3 - An Investigation of Donald Trump Effects – Log(1+Cases), June 29

	(1)	(2)	(3)	(4)
	Short Spec.	Baseline Controls	Adding Race Shares	Comprehensive Spec.
Panel A: Dependent Variable: Log (1+Cases), June 29				
Trump vote share, 2016 general election	-1.489 (0.138)*** [-0.110]	-1.576 (0.140)*** [-0.116]	1.129 (0.182)*** [0.083]	0.700 (0.203)*** [0.053]
R^2	0.76	0.79	0.83	0.83
N	3,111	3,109	3,109	3,007
Panel B: Dependent Variable: Log (1+Deaths), June 29				
Trump vote share,	-1.589	-1.333	1.318	1.473
2016 general election	(0.140)***	(0.145)***	(0.192)***	(0.217)***
	[-0.151]	[-0.126]	[0.125]	[0.138]
R^2	0.59	0.63	0.68	0.70
N	3,111	3,109	3,109	3,007

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets.
All columns contain an intercept.
Column 1 (the short specification) includes only a control for log population.
Column 2 adds controls for the baseline set of variables in Tables 1 and 2.
Column 3 adds variables measuring the % Black or African American, % Hispanic or Latino, % American Indian and Alaska Native and % Asian.
Column 4 adds controls for % high school graduate or higher (among persons age 25+), % with bachelor's degree or higher (among persons age 25+), 30-day mortality for heart attacks, 30-day mortality for heart failure, 30-day mortality for pneumonia, Gini index within bottom 99%, poverty rate, top 1% income share, % housing units in multi-unit structures, persons per household and log effective local density (i.e. the variables explored in Tables A5-A9, except share obese and share smoking).

Figure 1 - Effects on Log(1+Cases), by Date

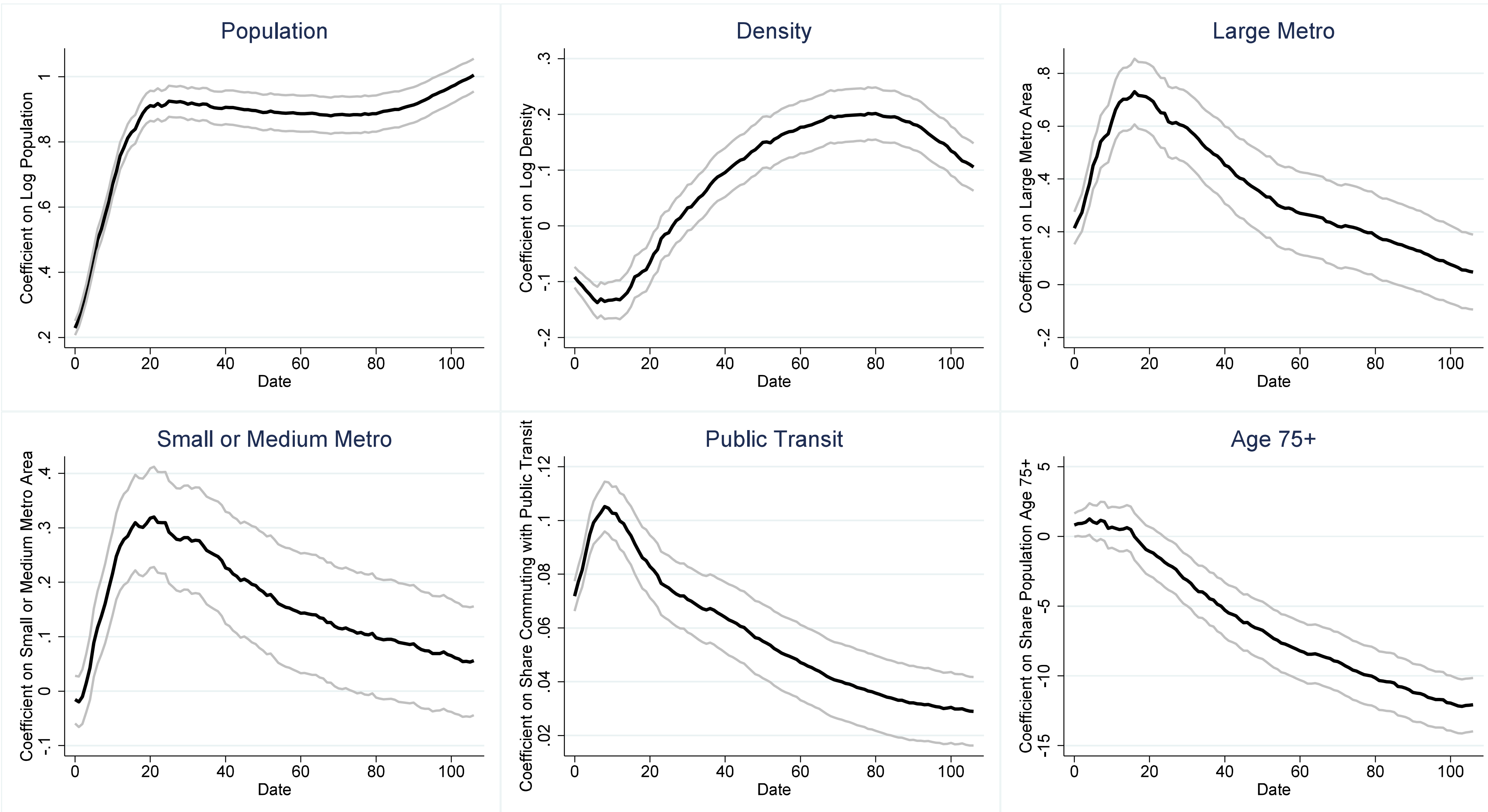


Figure 1A - Effects on Log(1+Cases), by Date (contd.)

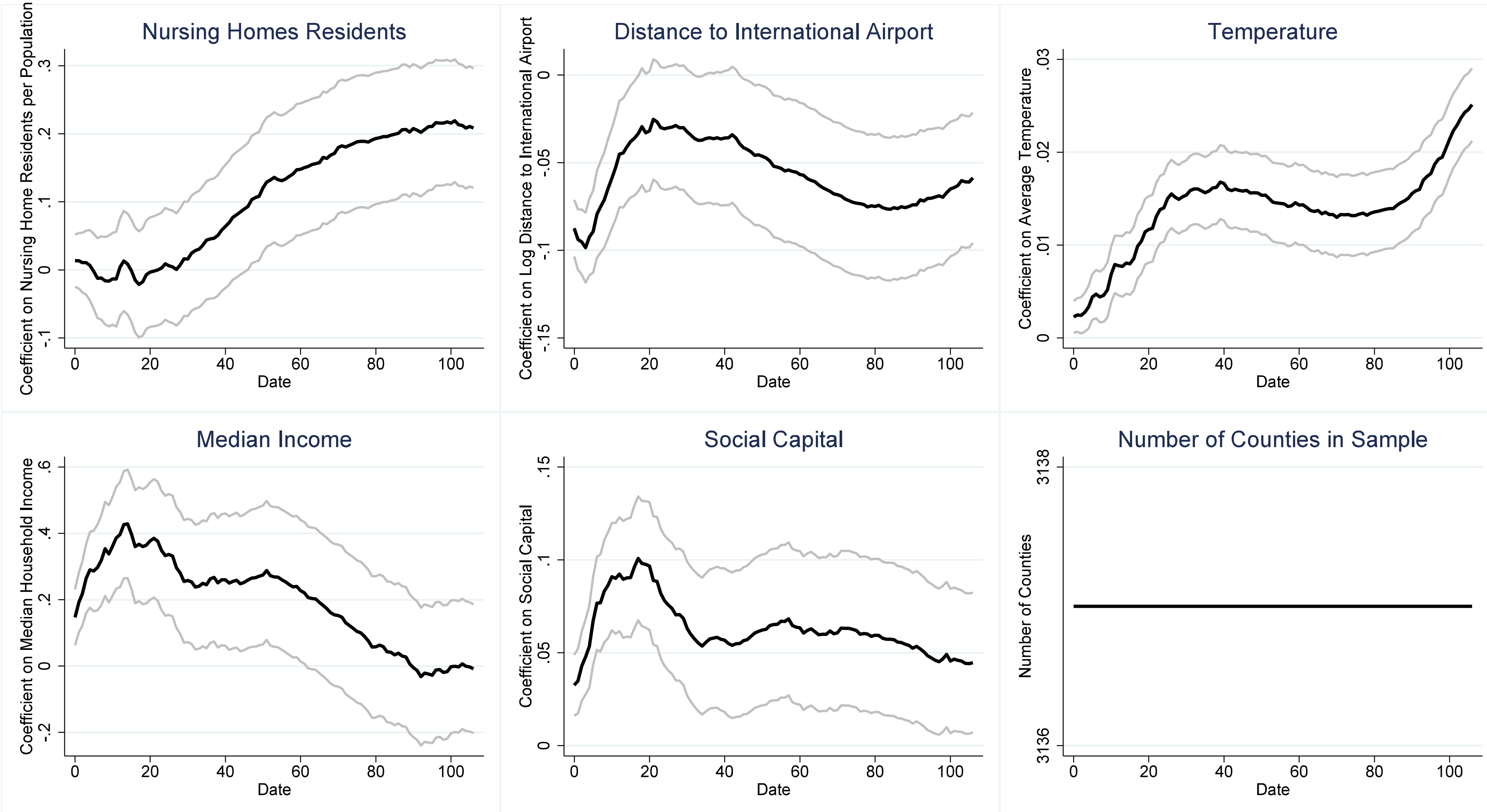


Figure 1B - Effects on Log(1+Deaths), by Date

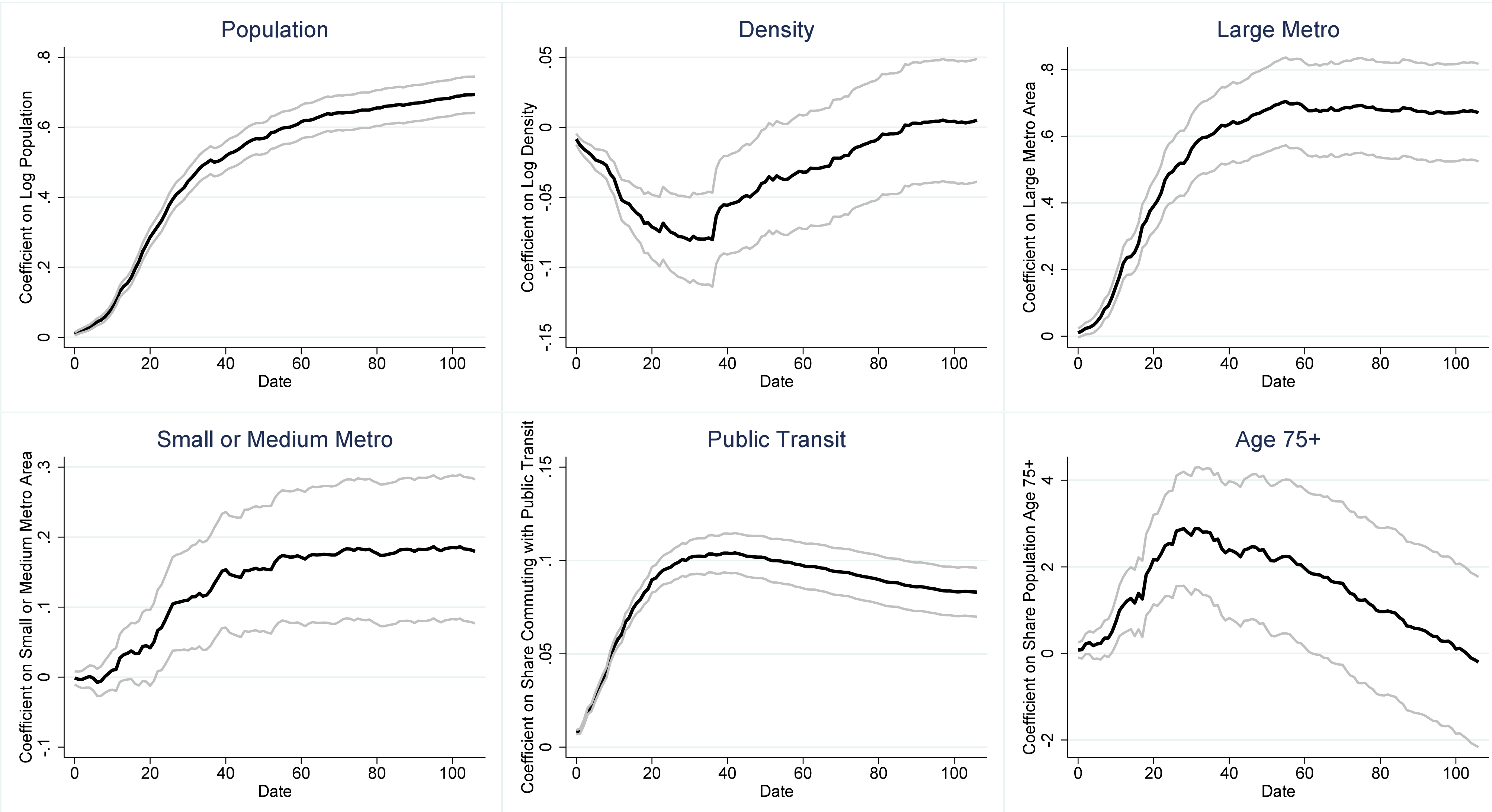


Figure 1B - Effects on Log(1+Deaths), by Date

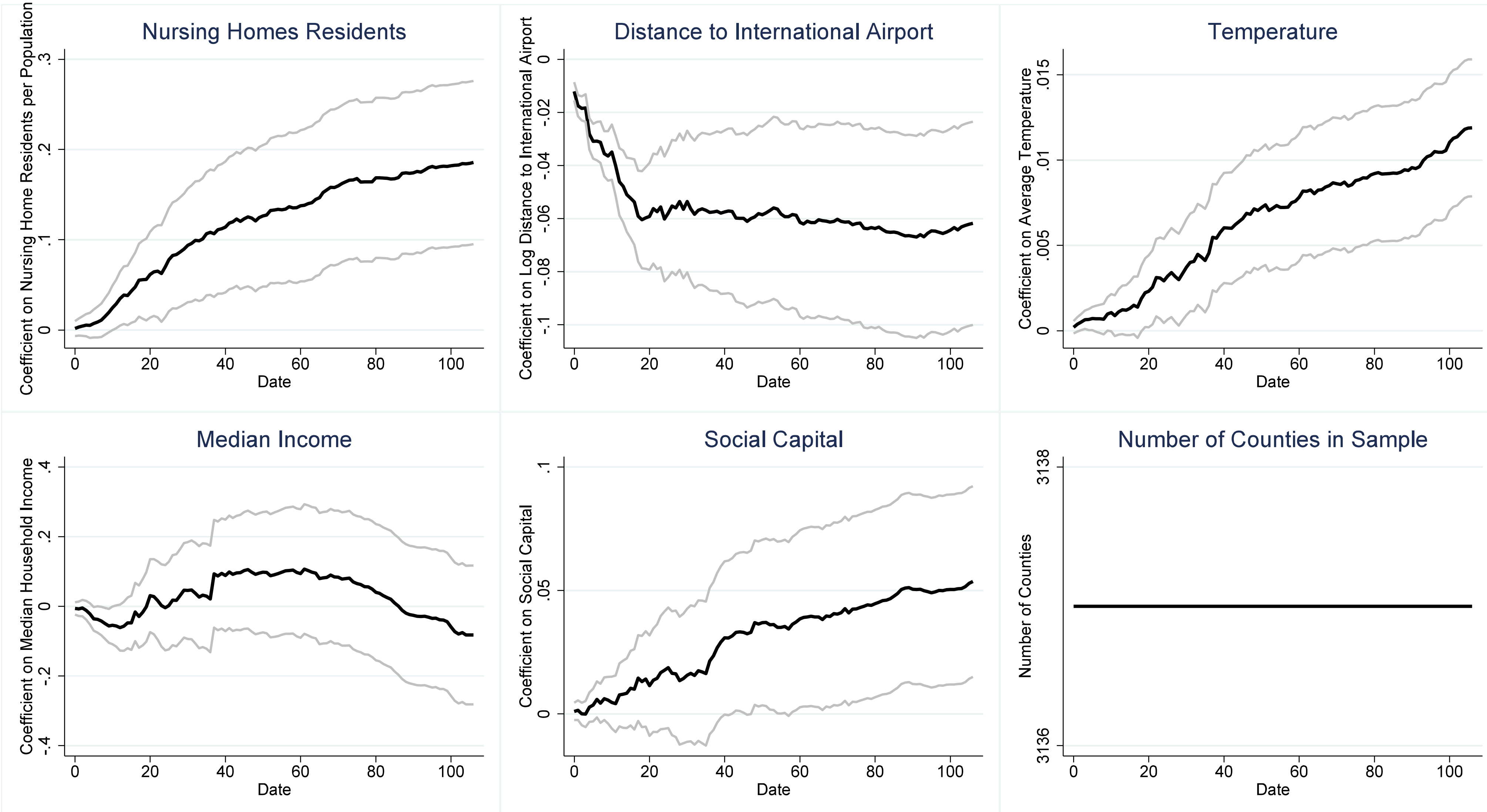
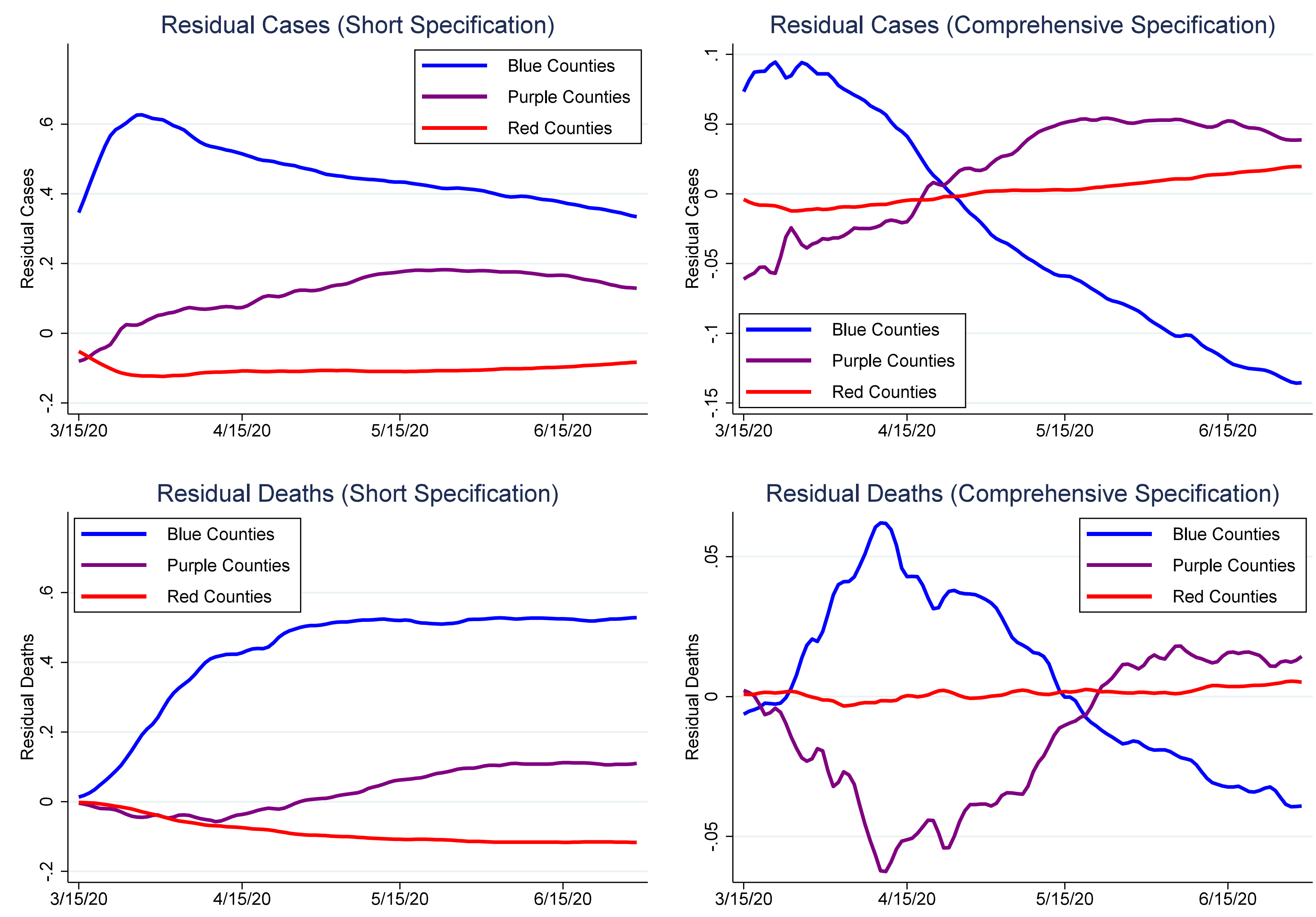


Figure 2 - The Political Divide in COVID-19 Severity



Appendix to "Understanding Spatial Variation in COVID-19 across the United States"

Klaus Desmet — SMU and NBER

Romain Wacziarg — UCLA and NBER

July 2020

Abstract

This Appendix contains: A) Description and sources of the data used in the analysis. B) Additional tables and figures mentioned in the main text.

A. Data Sources

A1. Dependent Variables

COVID-19 cases and deaths. Daily county-level data on COVID-19 cases and deaths. Source: *New York Times*, `https://github.com/nytimes/covid-19-data`. We adjusted the data in the following ways:

1. The source reports data cumulated for New York City overall (all 5 boroughs/counties together). We apportioned cases and deaths to each of the 5 boroughs/counties by county population shares.

2. The source reports data for all of Kansas City, which is made up of parts of several counties, each independent entries with their own cases and deaths (exclusive of Kansas City). Most of Kansas City is in Jackson County MO, so we added all Kansas City cases and deaths to that county's tally.

3. We did not make any modifications regarding any of the additional geographic specificities as described in the source data: "Counts for Alameda County (CA) include cases and deaths from Berkeley and the Grand Princess cruise ship; counts for Douglas County (NE) include cases brought to the state from the Diamond Princess cruise ship; all cases and deaths for Chicago are reported as part of Cook County (IL); counts for Guam include cases reported from the USS Theodore Roosevelt."

4. The source reports non-monotonic evolutions of cumulative cases and deaths for a very small set of counties, at the very beginning of the pandemic, when there were very few cases and deaths. The reason is unknown. We recoded cases and deaths that subsequently became lower to the level of the later lower number to ensure monotonic cumulative series for all counties.

A2. Independent Variables

Population and age. Age structure of population by county. Source: U.S. Census Bureau. *2018 American Community Survey 5-Year Estimates.* `https://data.census.gov/cedsci/`.

Population density. Population divided by land in square miles. Source: U.S. Census Bureau.

Metro county. Classification as large central metro county, large fringe metro country, medium metro county or small metro county. Source: National Center for Health Statistics (NCHS). *Urban-Rural Classification Scheme for Counties 2013.* `https://www.cdc.gov/nchs/data_access/urban_rural.htm#Data_Files_and_Documentation`

Public transportation. Share of population that goes to work by public transportation. Source: U.S. Census Bureau. *2018 American Community Survey 5-Year Estimates.* `https://data.census.gov/cedsci/`.

Nursing home residents. Percentage of population who are residents in nursing homes. Source: Centers for Medicare & Medicaid Services. *Nursing Home Compare Datasets: Provider Info.* `https://data.medicare.gov/data/nursing-home-compare`.

Temperature. Average temperature in February, March and April, 2009 to 2019. Source: National Oceanic and Atmospheric Administration. *NOAA's Gridded Climate Divisional Dataset (CLIMDIV).* `ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/`.

Distance to airport. Data on all international flights to the U.S. in 2019 come from Table T-100 from the Bureau of Transportation Statistics. For each U.S. airport, we take the average number of monthly passengers on direct flights from the top-5 countries in terms of COVID-19 cases on March 15, 2020 (China, Italy, Iran, South Korea and Spain). For each county in the U.S., we then compute the geodesic distance to the closest airport that received at least 250 passengers per month on direct flights from one of these 5 countries. `https://www.transtats.bts.gov/`

Household income. Log of median household income, 2009-2013. Source: U.S. Census Bureau.

Social capital. Social capital index created using principal component analysis using number of associations and organizations (including non-profits), voter turnout and census response rate in 2014 (variable sk14). Source: Rupasingha, A., S. J. Goetz and D. Freshwater (2006, with updates). `https://aese.psu.edu/nercrd/community/social-capital-resources`

Race. Black or African American alone, Hispanic or Latino, American Indian and Alaska Native alone, percentage 2014. Source: U.S. Census Bureau.

Education. High school graduate or higher, percentage of persons age 25+, 2009-2013, and bachelor's degree or higher, percentage of persons age 25+, 2009-2013. Source: U.S. Census Bureau.

Housing arrangements. Percent of housing units in multi-unit structures, 2009-2013, and persons per household, 2009-2013. Source: U.S. Census Bureau.

Smokers and obese. Percentage of the population that smokes and percentage of population that is obese. Source: Bergeron et al. (2016). `https://opportunityinsights.org/data/`.

Risk-adjusted mortality. 30-day risk adjusted mortality for heart attacks, heart failure and pneumonia. Source: Bergeron et al. (2016). `https://opportunityinsights.org/data/`.

Effective local density. Expected density in a one square kilometer around a randomly drawn individual from each county. If all county inhabitants are uniformly distributed across space, this measure is identical to standard population density. If the population is concentrated in a small subset of the county territory, this measure will be larger than standard population density. Own calculations based on 2020 population data from GPW. Source: Center for International Earth Science Information Network, *Gridded Population of the World*, Version 4: Population Count, Revision 11, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) (2018). `https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11`.

Trump vote share in the 2016 general election. Source: Dave Leip's Atlas of U.S. Presidential Elections. `https://uselectionatlas.org/`.

Stay-at-home orders. Days since first stay-at-home order. `https://commons.wikimedia.org/wiki/Data:Stay-at-home_orders_in_the_United_States.map#/map/0`.

References

[1] Bergeron, A., R. Chetty, D. Cutler, B. Scuderi, M. Stepner, N. Turner (2016), "The Association Between Income and Life Expectancy in the United States, 2001-2014," *Journal of the American Medical Association*, 315(16): 1750-1766.

[2] Rupasingha, A., S. J. Goetz and D. Freshwater (2006, with updates), "The Production of Social Capital in US Counties," *Journal of Socio-Economics*, 35, 83–101.

B. Additional Tables and Figures

Table A1 – Summary Statistics

Panel A - Summary Statistics for Various Indicators of Disease Severity (June 29, 2020)

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Total cases	3,142	822.580	4010.340	0	100,772
Cases per capita	3,142	501.226	762.467	0	15,481
Indicator for any case	3,142	0.969	0.173	0	1
Log 1 + Cases	3,142	4.380	2.132	0	11.521
Log Cases	3,045	4.461	2.097	0	11.521
Total Deaths	3,142	39.964	260.886	0	6,723
Deaths per capita	3,142	16.996	34.080	0	375
Indicator for any death	3,142	0.616	0.487	0	1
Log 1 + deaths	3,142	1.408	1.648	0	8.813
Log Deaths	1,934	2.025	1.758	0	8.813

Panel B - Summary Statistics for the Baseline Set of 11 Regressors

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Log Population	3,142	10.275	1.494	4.317	16.129
Log Density	3,140	3.786	1.784	-3.291	11.175
Large central or fringe metro county	3,142	0.139	0.346	0	1
Medium or small metro county	3,142	0.232	0.422	0	1
% people who commute by public transportation	3,141	0.902	3.066	0	60.700
Share of people aged 75 or older	3,142	0.079	0.023	0.013	0.241
% nursing home residents in pop.	3,142	0.603	0.448	0	5.047
Log km to closest airport w/ flights from top 5 COVID countries	3,142	5.562	1.144	-4.605	8.264
Average temperature, Feb., Mar. & Apr.	3,141	45.126	10.453	-0.317	73.067
Log household median Income	3,140	10.705	0.242	9.903	11.714
Social Capital Index, 2014	3,139	0.001	1.260	-3.183	21.809

Table A2 - OLS Regressions for log Cases and log Deaths, June 29, 2020
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log Cases	**Log Cases, State FE**	**Log Deaths**	**Log Deaths, State FE**
Log population	1.030 (0.027)*** [0.701]	1.056 (0.031)*** [0.719]	0.806 (0.039)*** [0.605]	0.881 (0.047)*** [0.661]
Log population density	0.110 (0.023)*** [0.089]	0.076 (0.030)** [0.062]	0.105 (0.035)*** [0.089]	0.027 (0.045) [0.023]
Large central metro county or large fringe metro county	0.041 (0.075) [0.007]	-0.007 (0.067) [-0.001]	0.398 (0.100)*** [0.091]	0.381 (0.091)*** [0.088]
Medium metro county or small metro county	0.056 (0.053) [0.011]	-0.003 (0.047) [-0.001]	0.016 (0.072) [0.004]	0.022 (0.066) [0.006]
% people who commute by public transportation	0.028 (0.007)*** [0.041]	0.021 (0.007)*** [0.032]	0.066 (0.008)*** [0.145]	0.052 (0.008)*** [0.112]
Share of people aged 75 & above	-13.250 (1.043)*** [-0.142]	-11.843 (1.027)*** [-0.127]	-1.229 (1.514) [-0.014]	-0.576 (1.545) [-0.007]
% nursing home residents in pop.	0.252 (0.050)*** [0.051]	0.100 (0.048)** [0.020]	0.731 (0.093)*** [0.136]	0.503 (0.094)*** [0.094]
Log km to closest airport w/ flights from top 5 COVID countries	-0.059 (0.020)*** [-0.032]	-0.056 (0.019)*** [-0.031]	-0.041 (0.023)* [-0.031]	-0.055 (0.023)** [-0.042]
Average temperature, Feb., Mar. & Apr.	0.025 (0.002)*** [0.125]	0.028 (0.005)*** [0.138]	0.007 (0.003)** [0.043]	0.021 (0.008)** [0.118]
Log household median income	-0.082 (0.104) [-0.010]	-0.115 (0.105) [-0.013]	-0.268 (0.148)* [-0.039]	-0.383 (0.151)** [-0.056]
Social Capital Index, 2014	0.055 (0.021)*** [0.032]	-0.003 (0.021) [-0.002]	-0.050 (0.032) [-0.027]	-0.017 (0.033) [-0.009]
Constant	-5.725 (1.162)***	-4.974 (1.205)***	-5.024 (1.645)***	-4.704 (1.749)***
R^2	0.77	0.83	0.58	0.67
N	3,042	3,042	1,933	1,933

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets.

Table A3 - OLS Regressions for log Cases and log Deaths, June 29, 2020, CFR<0.1
(Dependent variable listed in second row)

	(1)	(2)	(3)	(4)
	Log Cases	**Log Cases, State FE**	**Log Deaths**	**Log Deaths, State FE**
Log population	1.031 (0.028)*** [0.702]	1.058 (0.032)*** [0.720]	0.858 (0.040)*** [0.635]	0.959 (0.048)*** [0.709]
Log population density	0.108 (0.024)*** [0.087]	0.074 (0.031)** [0.060]	0.100 (0.036)*** [0.083]	0.016 (0.046) [0.014]
Large central metro county or large fringe metro county	0.023 (0.078) [0.004]	-0.005 (0.070) [-0.001]	0.360 (0.101)*** [0.084]	0.359 (0.091)*** [0.084]
Medium metro county or small metro county	0.062 (0.055) [0.013]	0.002 (0.049) [0.000]	0.014 (0.073) [0.004]	-0.003 (0.066) [-0.001]
% people who commute by public transportation	0.028 (0.007)*** [0.042]	0.023 (0.007)*** [0.035]	0.066 (0.008)*** [0.148]	0.050 (0.008)*** [0.114]
Share of people aged 75 & above	-13.405 (1.084)*** [-0.144]	-12.117 (1.062)*** [-0.130]	-2.627 (1.544)* [-0.030]	-2.146 (1.562) [-0.024]
% nursing home residents in pop.	0.219 (0.052)*** [0.044]	0.070 (0.050) [0.014]	0.613 (0.100)*** [0.110]	0.414 (0.100)*** [0.074]
Log km to closest airport w/ flights from top 5 COVID countries	-0.055 (0.020)*** [-0.030]	-0.054 (0.020)*** [-0.030]	-0.027 (0.023) [-0.021]	-0.044 (0.023)* [-0.034]
Average temperature, Feb., Mar. & Apr.	0.024 (0.002)*** [0.121]	0.026 (0.005)*** [0.129]	0.009 (0.003)*** [0.052]	0.014 (0.008)* [0.083]
Log household median Income	-0.111 (0.109) [-0.013]	-0.120 (0.109) [-0.014]	-0.403 (0.151)*** [-0.059]	-0.467 (0.153)*** [-0.069]
Social Capital Index, 2014	0.059 (0.021)*** [0.034]	-0.002 (0.021) [-0.001]	-0.044 (0.033) [-0.024]	-0.025 (0.032) [-0.014]
Constant	-5.394 (1.212)***	-4.823 (1.245)***	-4.220 (1.687)**	-4.182 (1.760)**
R^2	0.77	0.83	0.60	0.70
N	2,847	2,847	1,738	1,738

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets.

Table A4 - OLS Regressions for Log Cases and Log Deaths, Synchronized Days from Onset at 70 days from Onset (for Cases) and 60 days from Onset (for Deaths), Sample with CFR<0.1

	(1) Log Cases	(2) Log Cases, State FE	(3) Log Deaths	(4) Log Deaths, State FE
Log population	0.864	0.968	0.833	0.950
	(0.032)***	(0.037)***	(0.046)***	(0.055)***
	[0.612]	[0.685]	[0.635]	[0.724]
Log population density	0.165	0.043	0.053	-0.030
	(0.027)***	(0.036)	(0.042)	(0.052)
	[0.136]	[0.035]	[0.045]	[-0.025]
Large central metro county or large fringe metro county	0.155	0.132	0.491	0.445
	(0.087)*	(0.078)*	(0.118)***	(0.105)***
	[0.029]	[0.025]	[0.125]	[0.113]
Medium metro county or small metro county	0.042	0.005	0.052	-0.006
	(0.061)	(0.055)	(0.089)	(0.078)
	[0.010]	[0.001]	[0.014]	[-0.002]
% people who commute by public transportation	0.059	0.051	0.085	0.054
	(0.010)***	(0.011)***	(0.011)***	(0.011)***
	[0.075]	[0.064]	[0.164]	[0.105]
Share of people aged 75 & above	-10.497	-10.137	-1.368	-2.180
	(1.280)***	(1.266)***	(1.809)	(1.844)
	[-0.116]	[-0.112]	[-0.015]	[-0.025]
% nursing home residents in pop.	0.261	0.069	0.774	0.421
	(0.065)***	(0.064)	(0.127)***	(0.123)***
	[0.053]	[0.014]	[0.131]	[0.071]
Log km to closest airport w/ flights from top 5 COVID countries	-0.058	-0.051	-0.026	-0.038
	(0.022)***	(0.022)**	(0.025)	(0.024)
	[-0.035]	[-0.031]	[-0.022]	[-0.033]
Average temperature, Feb., Mar. & Apr.	0.013	0.017	0.003	0.029
	(0.003)***	(0.006)***	(0.004)	(0.010)***
	[0.067]	[0.090]	[0.019]	[0.168]
Log household median income	-0.113	-0.249	-0.318	-0.477
	(0.124)	(0.124)**	(0.173)*	(0.168)***
	[-0.014]	[-0.032]	[-0.050]	[-0.075]
Social Capital Index, 2014	0.030	-0.013	-0.057	-0.030
	(0.025)	(0.025)	(0.036)	(0.034)
	[0.018]	[-0.007]	[-0.032]	[-0.017]
Constant	-4.061	-2.814	-4.653	-4.694
	(1.376)***	(1.424)**	(1.924)**	(1.951)**
R^2	0.68	0.76	0.59	0.71
N	2,498	2,498	1,233	1,233

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets. Onset day is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths).

Table A5 - An Investigation of Effective Density

	(1)	(2)	(3)	(4)	(5)	(6)
	Log 1+Cases, June 29	**Log 1+Cases, June 29**	**Log 1+Cases, June 29**	**Log 1+Deaths, June 29**	**Log 1+Deaths, June 29**	**Log 1+Deaths, June 29**
Log population density	0.146	0.146	0.117	0.022	0.019	0.043
	(0.022)***	(0.022)***	(0.022)***	(0.023)	(0.023)	(0.023)*
	[0.122]	[0.122]	[0.098]	[0.024]	[0.021]	[0.046]
Large central metro county or large fringe metro county	0.041	0.041	0.080	0.682	0.667	0.722
	(0.071)	(0.072)	(0.073)	(0.074)***	(0.074)***	(0.076)***
	[0.007]	[0.007]	[0.013]	[0.143]	[0.140]	[0.152]
Medium metro county or small metro county	0.050	0.050	0.045	0.170	0.164	0.137
	(0.050)	(0.050)	(0.051)	(0.052)***	(0.052)***	(0.054)**
	[0.010]	[0.010]	[0.009]	[0.044]	[0.042]	[0.035]
Housing units in multi-unit structures, percent	0.015	0.014		0.018	0.023	
	(0.003)***	(0.003)***		(0.003)***	(0.004)***	
	[0.063]	[0.063]		[0.102]	[0.132]	
Persons per household	1.023	1.022		0.634	0.686	
	(0.100)***	(0.102)***		(0.104)***	(0.105)***	
	[0.118]	[0.117]		[0.094]	[0.102]	
% people who commute by public transportation	0.011	0.011		0.061	0.060	
	(0.007)	(0.007)		(0.008)***	(0.008)***	
	[0.016]	[0.016]		[0.113]	[0.111]	
Log effective local density		0.002	0.100		-0.108	0.075
		(0.031)	(0.027)***		(0.032)***	(0.029)***
		[0.001]	[0.047]		[-0.065]	[0.045]
Constant	-7.525	-7.510	-5.611	-6.195	-7.006	-4.827
	(1.092)***	(1.117)***	(1.121)***	(1.134)***	(1.157)***	(1.180)***
R^2	0.80	0.80	0.79	0.63	0.63	0.61
N	3,137	3,137	3,137	3,137	3,137	3,137

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets. All columns contain an intercept and controls for the remaining baseline set of regressors in Tables 1 and 2.

Table A6 - An Investigation of Race Effects

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
% Black or African American	0.037 (0.001)*** [0.253]	0.041 (0.002)*** [0.317]	0.035 (0.002)*** [0.303]	0.034 (0.002)*** [0.328]
% Hispanic or Latino	0.020 (0.001)*** [0.124]	0.020 (0.002)*** [0.132]	0.013 (0.002)*** [0.102]	0.012 (0.003)*** [0.082]
% American Indian and Alaska Native	0.011 (0.002)*** [0.039]	0.008 (0.003)** [0.027]	0.014 (0.003)*** [0.062]	0.034 (0.006)*** [0.093]
% Asian	-0.034 (0.007)*** [-0.043]	-0.025 (0.009)*** [-0.035]	0.006 (0.008) [0.010]	-0.018 (0.012) [-0.035]
R^2	0.83	0.74	0.68	0.64
N	3,137	2,755	3,137	1,446

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets. Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths). All specifications contain an intercept and controls for the baseline set of variables in Tables 1 and 2.

Table A7 - An Investigation of Education Effects

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
High school graduate or higher, percent of persons age 25+	-0.053 (0.004)*** [-0.172]	-0.065 (0.005)*** [-0.228]	-0.038 (0.005)*** [-0.160]	-0.056 (0.008)*** [-0.211]
Bachelor's degree or higher, percent of persons age 25+	0.007 (0.003)** [0.030]	0.007 (0.004)* [0.031]	0.018 (0.003)*** [0.096]	0.016 (0.005)*** [0.097]
R^2	0.80	0.70	0.63	0.59
N	3,137	2,755	3,137	1,446

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets. Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths). All columns contain an intercept and controls for the baseline set of variables in Tables 1 and 2.

Table A8 - An Investigation of Health Effects

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
Percentage of the population that smokes	-1.686 (0.343)*** [-0.054]	-1.110 (0.395)*** [-0.036]	-0.442 (0.383) [-0.015]	0.022 (0.655) [0.001]
Percentage of the population that is obese	0.404 (0.289) [0.016]	0.577 (0.337)* [0.024]	0.669 (0.322)** [0.029]	0.200 (0.547) [0.008]
30-day Mortality for Heart Attacks	-1.991 (0.750)*** [-0.031]	-2.133 (0.908)** [-0.032]	-1.395 (0.838)* [-0.023]	-0.460 (1.462) [-0.006]
30-day Mortality for Heart Failure	-0.154 (1.251) [-0.001]	-1.342 (1.481) [-0.013]	-3.630 (1.397)*** [-0.037]	-8.035 (2.318)*** [-0.073]
30-day Mortality for Pneumonia	1.089 (1.128) [0.012]	3.075 (1.324)** [0.033]	2.558 (1.260)** [0.029]	4.170 (2.009)** [0.044]
R^2	0.75	0.66	0.63	0.59
N	2,334	2,250	2,334	1,334

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets. Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths). All columns contain an intercept and controls for the baseline set of variables in Tables 1 and 2. Note the smaller number of observations due to lack of availability of data on obesity and smoking. 30-day mortality measures are risk-adjusted so are likely to capture mostly the quality of the health infrastructure / health care system in the county.

Table A9 - An Investigation of Inequality and Poverty Effects

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
Gini Index Within Bottom 99%	0.874 (0.442)** [0.037]	0.536 (0.520) [0.024]	0.615 (0.445) [0.032]	0.984 (0.721) [0.050]
Poverty Rate	1.151 (0.558)** [0.036]	1.735 (0.666)*** [0.058]	4.166 (0.562)*** [0.162]	7.509 (0.982)*** [0.282]
Top 1% Income Share	-0.102 (0.593) [-0.003]	-0.021 (0.686) [-0.001]	0.235 (0.597) [0.007]	-0.872 (0.922) [-0.027]
Log household median income	0.279 (0.144)* [0.033]	0.273 (0.169) [0.035]	0.673 (0.145)*** [0.099]	1.299 (0.234)*** [0.197]
R^2	0.77	0.68	0.65	0.60
N	3,026	2,728	3,026	1,441

* p<0.1; ** p<0.05; *** p<0.01. Standard errors in parentheses and standardized betas in brackets. Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths). All columns contain an intercept and controls for the remaining baseline set of variables in Tables 1 and 2. There is collinearity between poverty rate and median income (ρ = - 0.75). The coefficient on median income is robust but the coefficient on the poverty rate is sensitive to the inclusion of median income (it becomes zero without median income included).

Table A10 - Comprehensive Specification

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
Log population	1.035 (0.028)*** [0.704]	0.897 (0.033)*** [0.624]	0.773 (0.030)*** [0.650]	0.842 (0.051)*** [0.645]
Log population density	0.066 (0.024)*** [0.053]	0.140 (0.028)*** [0.114]	0.031 (0.025) [0.031]	0.091 (0.047)* [0.078]
Large central metro county or large fringe metro county	-0.142 (0.065)** [-0.024]	-0.043 (0.074) [-0.008]	0.405 (0.070)*** [0.086]	0.245 (0.106)** [0.061]
Medium metro county or small metro county	0.023 (0.045) [0.005]	0.013 (0.051) [0.003]	0.093 (0.049)* [0.024]	-0.008 (0.078) [-0.002]
% people who commute by public transportation	0.003 (0.007) [0.004]	0.017 (0.008)** [0.030]	0.044 (0.007)*** [0.083]	0.034 (0.010)*** [0.087]
Share of people aged 75 & above	-0.280 (1.134) [-0.003]	2.700 (1.355)** [0.030]	10.342 (1.216)*** [0.139]	11.514 (1.951)*** [0.132]
% nursing home residents in pop.	0.277 (0.046)*** [0.056]	0.192 (0.059)*** [0.038]	0.336 (0.050)*** [0.084]	0.719 (0.112)*** [0.126]
Log km to closest airport w/ flights fr. top 5 COVID countries	-0.023 (0.018) [-0.013]	-0.026 (0.020) [-0.016]	-0.016 (0.019) [-0.011]	-0.036 (0.025) [-0.030]
Average temperature, Feb., Mar. & Apr.	-0.002 (0.003) [-0.008]	-0.019 (0.003)*** [-0.100]	-0.013 (0.003)*** [-0.078]	-0.030 (0.004)*** [-0.178]
Log household median income	0.556 (0.160)*** [0.066]	0.704 (0.191)*** [0.090]	0.843 (0.171)*** [0.123]	1.750 (0.301)*** [0.265]
Social Capital Index, 2014	-0.024 (0.021) [-0.014]	-0.047 (0.025)* [-0.026]	-0.025 (0.022) [-0.018]	-0.080 (0.036)** [-0.043]
% Black or African American	0.039 (0.002)*** [0.276]	0.044 (0.002)*** [0.344]	0.033 (0.002)*** [0.294]	0.029 (0.003)*** [0.281]
% Hispanic or Latino	0.007 (0.002)*** [0.048]	0.009 (0.002)*** [0.061]	0.005 (0.002)** [0.040]	0.001 (0.004) [0.004]
% American Indian and Alaska Native	0.015 (0.003)*** [0.050]	0.013 (0.004)*** [0.042]	0.015 (0.003)*** [0.060]	0.022 (0.007)*** [0.061]
% Asian	-0.044 (0.009)*** [-0.056]	-0.029 (0.010)*** [-0.041]	-0.010 (0.009) [-0.016]	-0.024 (0.012)** [-0.048]

	(1)	(2)	(3)	(4)
	Log 1+Cases, June 29	**Log Cases, 70 days since onset**	**Log 1+Deaths, June 29**	**Log Deaths, 60 days since onset**
High school graduate or higher, percent of persons age 25+	-0.060 (0.005)*** [-0.201]	-0.065 (0.006)*** [-0.229]	-0.035 (0.006)*** [-0.145]	-0.044 (0.010)*** [-0.166]
Bachelor's degree or higher, percent of persons age 25+	0.010 (0.004)*** [0.044]	0.007 (0.004) [0.032]	0.008 (0.004)** [0.043]	0.004 (0.006) [0.023]
30-day Mortality for Heart Attacks	-0.742 (0.489) [-0.012]	-0.976 (0.626) [-0.016]	0.091 (0.524) [0.002]	0.734 (1.173) [0.011]
30-day Mortality for Heart Failure	2.500 (0.858)*** [0.024]	2.332 (1.074)** [0.023]	-0.815 (0.920) [-0.010]	-2.317 (1.997) [-0.022]
30-day Mortality for Pneumonia	-1.500 (0.798)* [-0.016]	-0.943 (0.986) [-0.010]	-0.859 (0.856) [-0.011]	0.016 (1.687) [0.000]
Gini Index Within Bottom 99%	-1.444 (0.414)*** [-0.061]	-1.659 (0.487)*** [-0.075]	-1.307 (0.443)*** [-0.068]	-0.010 (0.734) [-0.000]
Poverty Rate	-4.333 (0.572)*** [-0.137]	-4.348 (0.681)*** [-0.146]	-0.011 (0.613) [-0.000]	2.593 (1.114)** [0.097]
Top 1% Income Share	1.826 (0.536)*** [0.045]	1.723 (0.622)*** [0.046]	1.872 (0.575)*** [0.057]	0.037 (0.908) [0.001]
Housing units in multi-unit structures, percent	0.015 (0.004)*** [0.067]	0.010 (0.004)** [0.050]	0.008 (0.004)* [0.043]	0.001 (0.006) [0.005]
Persons per household	0.627 (0.116)*** [0.073]	0.502 (0.142)*** [0.060]	0.309 (0.125)** [0.044]	0.364 (0.231) [0.046]
Log effective local density	0.054 (0.033)* [0.025]	0.046 (0.039) [0.023]	-0.008 (0.035) [-0.005]	-0.020 (0.063) [-0.011]
Days since lockdown began (0 if no or before lockdown)		-0.004 (0.001)*** [-0.051]		-0.005 (0.001)*** [-0.074]
Constant	-8.977 (1.748)***	-8.268 (2.076)***	-14.125 (1.874)***	-23.648 (3.238)***
R^2	0.83	0.76	0.70	0.65
N	3,020	2,726	3,020	1,440

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. * *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets. The comprehensive specification includes all control variables in Tables 1 and 2 and A5-A9 except the percentage of obese persons and the percentage of smokers (from Table A8), due to the loss of observations that would result from their inclusion.

Table A11 - An Investigation of the Effects of Lockdowns

	(1)	(2)	(3)	(4)
	Log Cases, 70 days since onset	**Log Cases, 70 days since onset, State FEs**	**Log Deaths, 60 days since onset**	**Log Deaths, 60 days since onset, State FEs**
Days since lockdown began (0 if no or before lockdown)	-0.002 (0.001)* [-0.021]	-0.002 (0.002) [-0.027]	-0.004 (0.001)*** [-0.052]	-0.013 (0.002)*** [-0.179]
R^2	0.69	0.76	0.58	0.69
N	2,755	2,755	1,446	1,446

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets. Onset is defined as the day at which the number of cases reaches 1 per 100,000 (for cases) and 0.5 per 100,000 (for deaths). All columns contain an intercept and controls for the baseline set of variables in Tables 1 and 2.

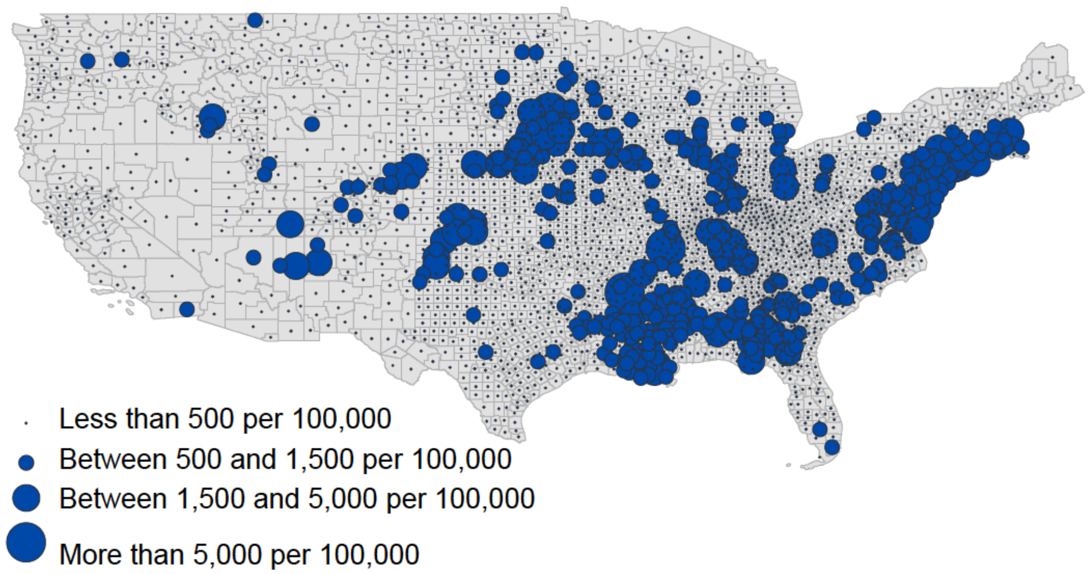
Table A12 - An Investigation of Donald Trump Effects (days-since-onset specification)

	(1)	(2)	(3)	(4)
	Short Spec.	**Baseline Controls**	**Adding Race Shares**	**Comprehensive Spec.**
Panel A – Dependent Variable: Log Cases, 70 Days from Onset				
Trump vote share, 2016 general election	-1.774 (0.156)*** [-0.143]	-1.644 (0.164)*** [-0.133]	1.449 (0.214)*** [0.117]	0.899 (0.239)*** [0.073]
R^2	0.66	0.70	0.75	0.76
N	2,744	2,742	2,742	2,716
Panel B – Dependent Variable: Log Deaths, 60 Days from Onset				
Trump vote share, 2016 general election	-2.067 (0.218)*** [-0.192]	-1.743 (0.230)*** [-0.162]	0.866 (0.326)*** [0.080]	0.735 (0.380)* [0.068]
R2	0.54	0.60	0.64	0.65
N	1,443	1,443	1,443	1,437

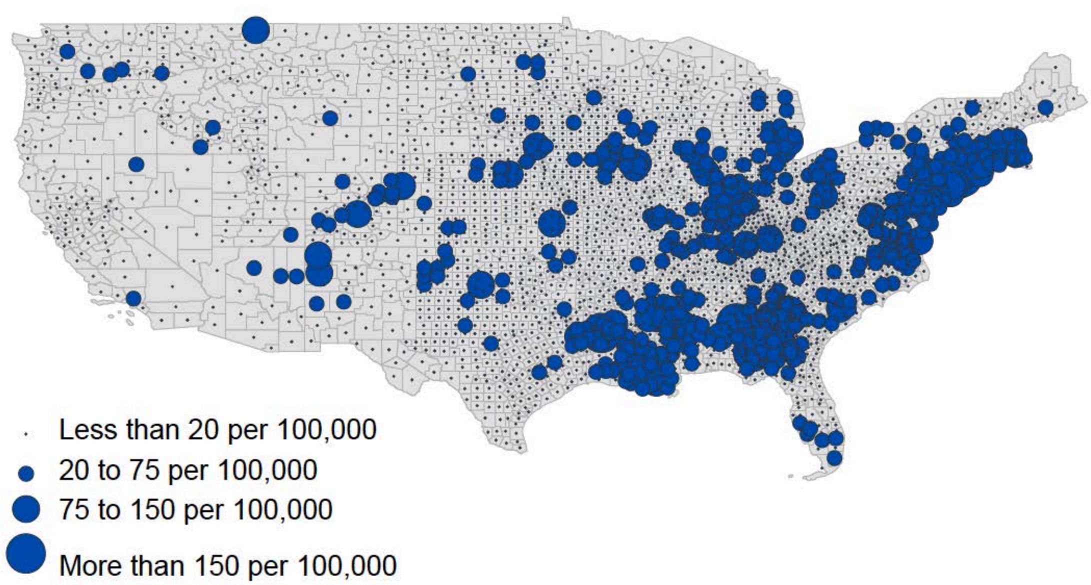
* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Standard errors in parentheses and standardized betas in brackets.
All specifications contain an intercept.
Column 1 (the short specification) includes only a control for log population.
Column 2 adds controls for the baseline set of variables in Tables 1 and 2.
Column 3 adds variables measuring the % Black or African American, % Hispanic or Latino, % American Indian and Alaska Native and % Asian.
Column 4 adds controls for % high school graduate or higher (among persons age 25+), % with bachelor's degree or higher (among persons age 25+), 30-day mortality for heart attacks, 30-day mortality for heart failure, 30-day mortality for pneumonia, Gini index within bottom 99%, poverty rate, top 1% income share, % housing units in multi-unit structures, persons per household and log effective local density (i.e. the variables explored in Tables A5-A9, except share obese and share smoking), plus number of days since the first stay-at-home order.

Figure A1 – Maps of the Variables Used in the Analysis

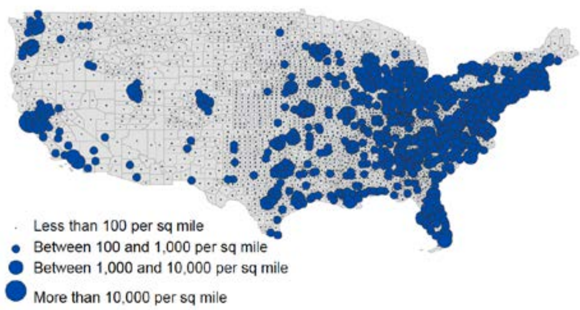
Cumulative Cases per 100,000 on May 26



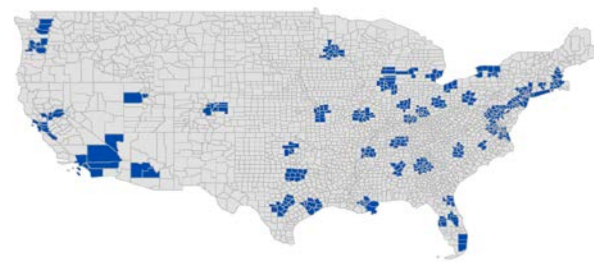
Cumulative Deaths per 100,000 on May 26



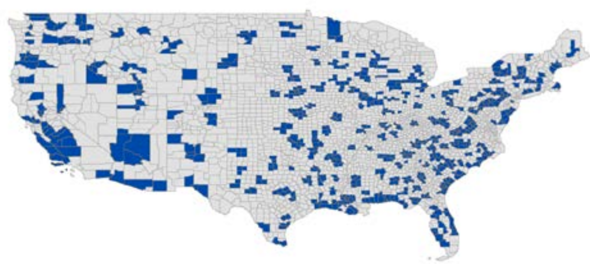
Population Density



Large Metro



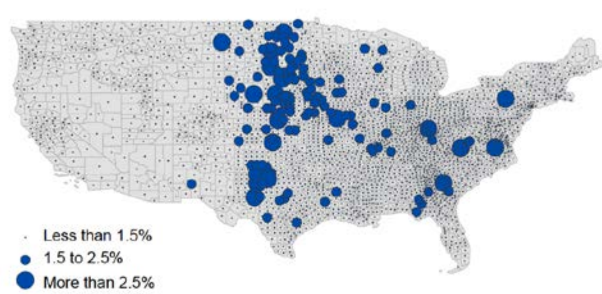
Medium or Small Metro



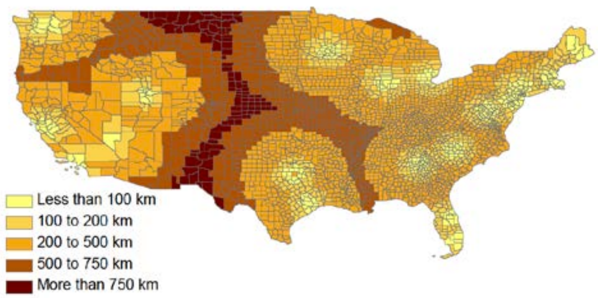
Commute by Public Transportation

Population Aged 75+

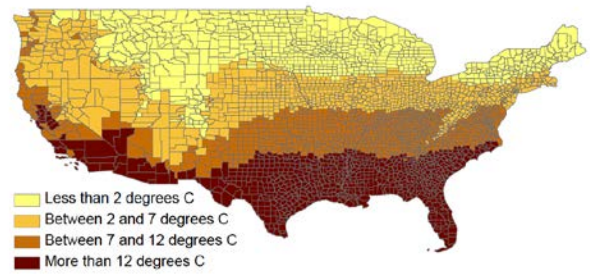
Nursing Homes Residents



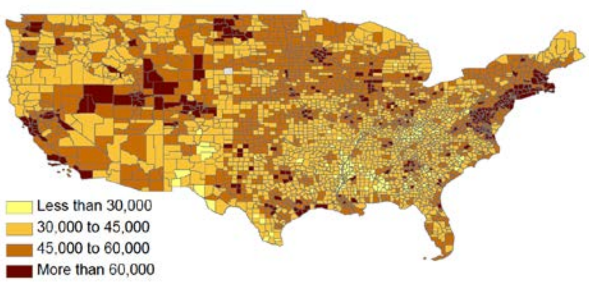
Distance to Airport with Flights to High-COVID Countries



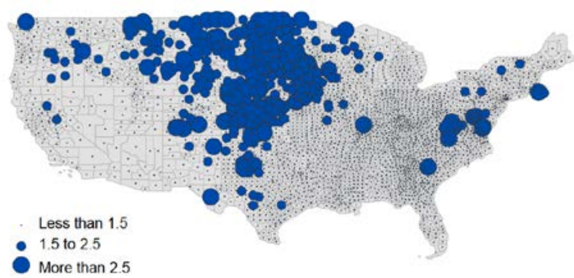
Average Temperature February to April



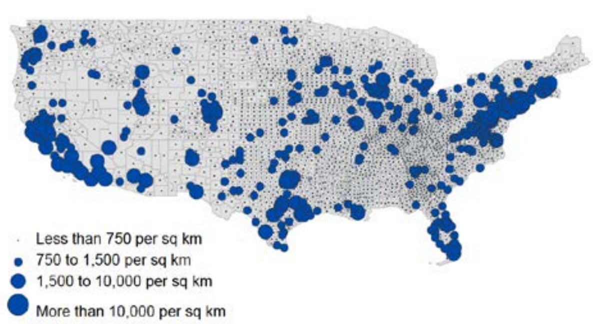
Median Household Income



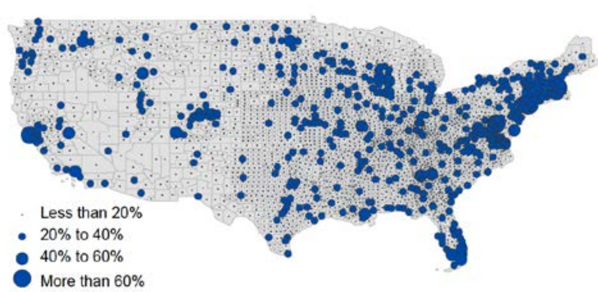
Social Capital



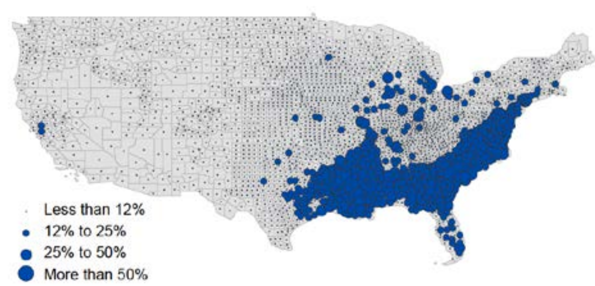
Effective Local Population Density



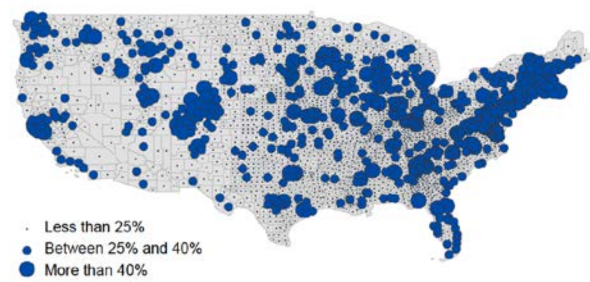
Multi-Unit Housing



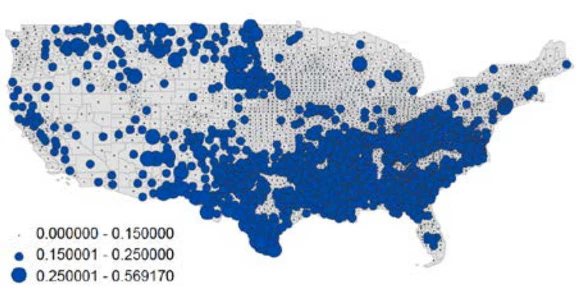
Share of African American Population



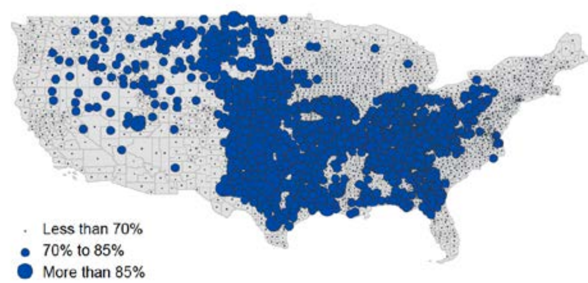
Share Bachelor's Degree or More



Poverty Rate



Trump 2016 Vote Share



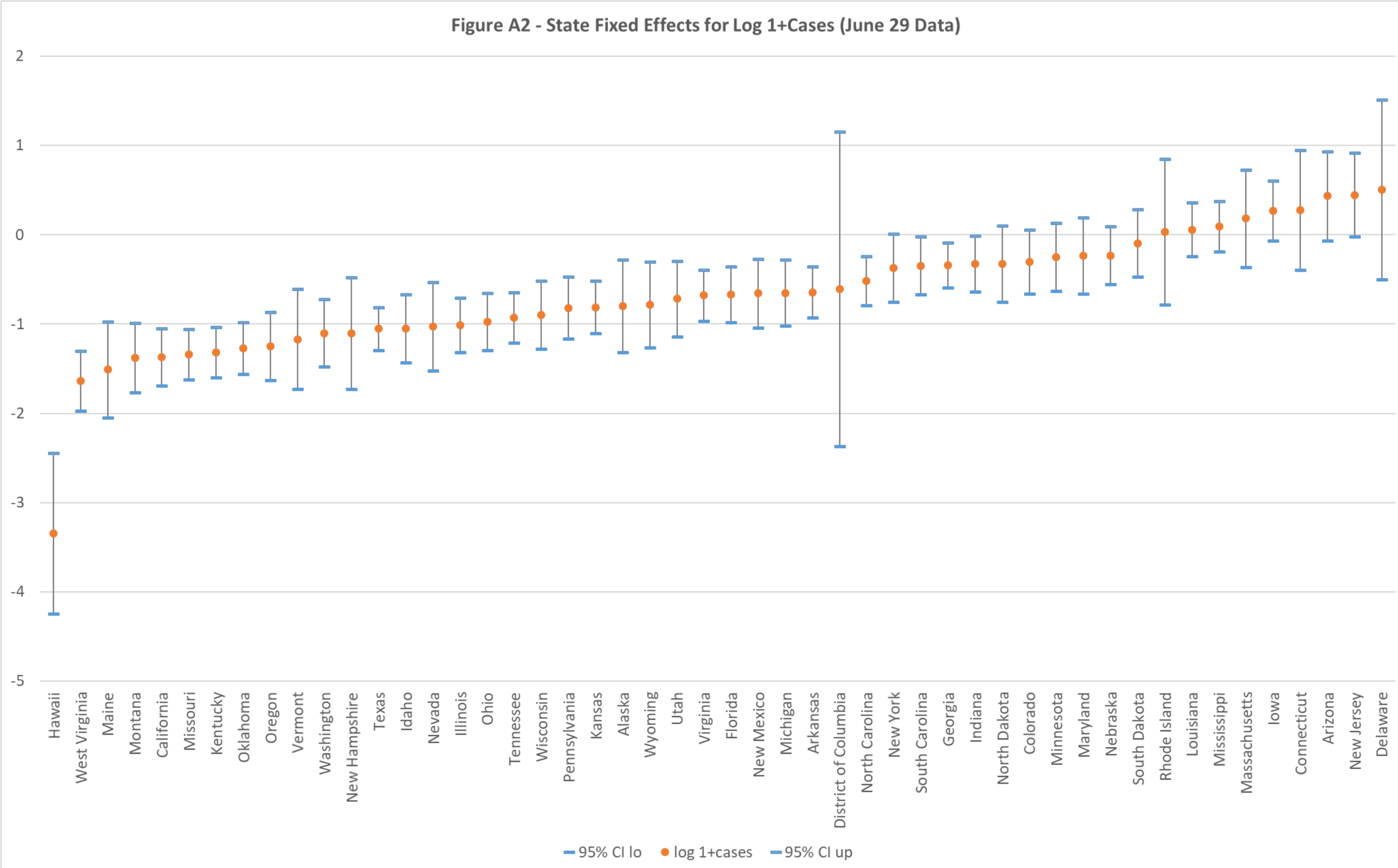


Figure A2 - State Fixed Effects for Log 1+Cases (June 29 Data)
2
1
0
-1
-2
-3
-4
-5
Hawaii
West Virginia
Maine
Montana
California
Missouri
Kentucky
Oklahoma
Oregon
Vermont
Washington
New Hampshire
Texas
Idaho
Nevada
Illinois
Ohio
Tennessee
Wisconsin
Pennsylvania
Kansas
Alaska
Wyoming
Utah
Virginia
Florida
New Mexico
Michigan
Arkansas
District of Columbia
North Carolina
New York
South Carolina
Georgia
Indiana
North Dakota
Colorado
Minnesota
Maryland
Nebraska
South Dakota
Rhode Island
Louisiana
Mississippi
Massachusetts
Iowa
Connecticut
Arizona
New Jersey
Delaware
95% CI lo
log 1+cases
95% CI up

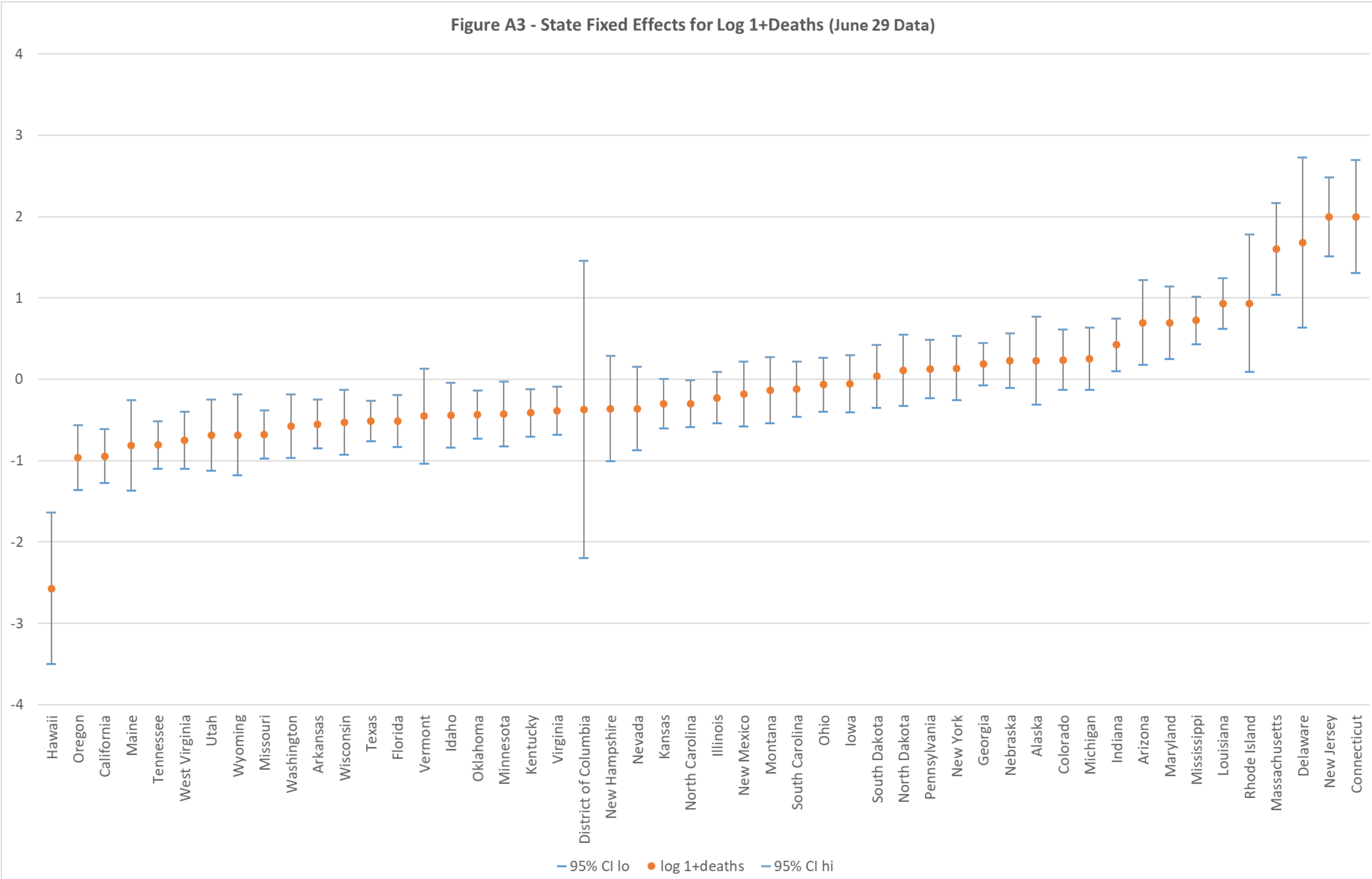


Figure A3 - State Fixed Effects for Log 1+Deaths (June 29 Data)
4
3
2
1
0
-1
-2
-3
-4
Hawaii
Oregon
California
Maine
Tennessee
West Virginia
Utah
Wyoming
Missouri
Washington
Arkansas
Wisconsin
Texas
Florida
Vermont
Idaho
Oklahoma
Minnesota
Kentucky
Virginia
District of Columbia
New Hampshire
Nevada
Kansas
North Carolina
Illinois
New Mexico
Montana
South Carolina
Ohio
Iowa
South Dakota
North Dakota
Pennsylvania
New York
Georgia
Nebraska
Alaska
Colorado
Michigan
Indiana
Arizona
Maryland
Mississippi
Louisiana
Rhode Island
Massachusetts
Delaware
New Jersey
Connecticut
95% CI lo
log 1+deaths
95% CI hi

Figure A4A - Effects on Log Cases, by Days Since Onset

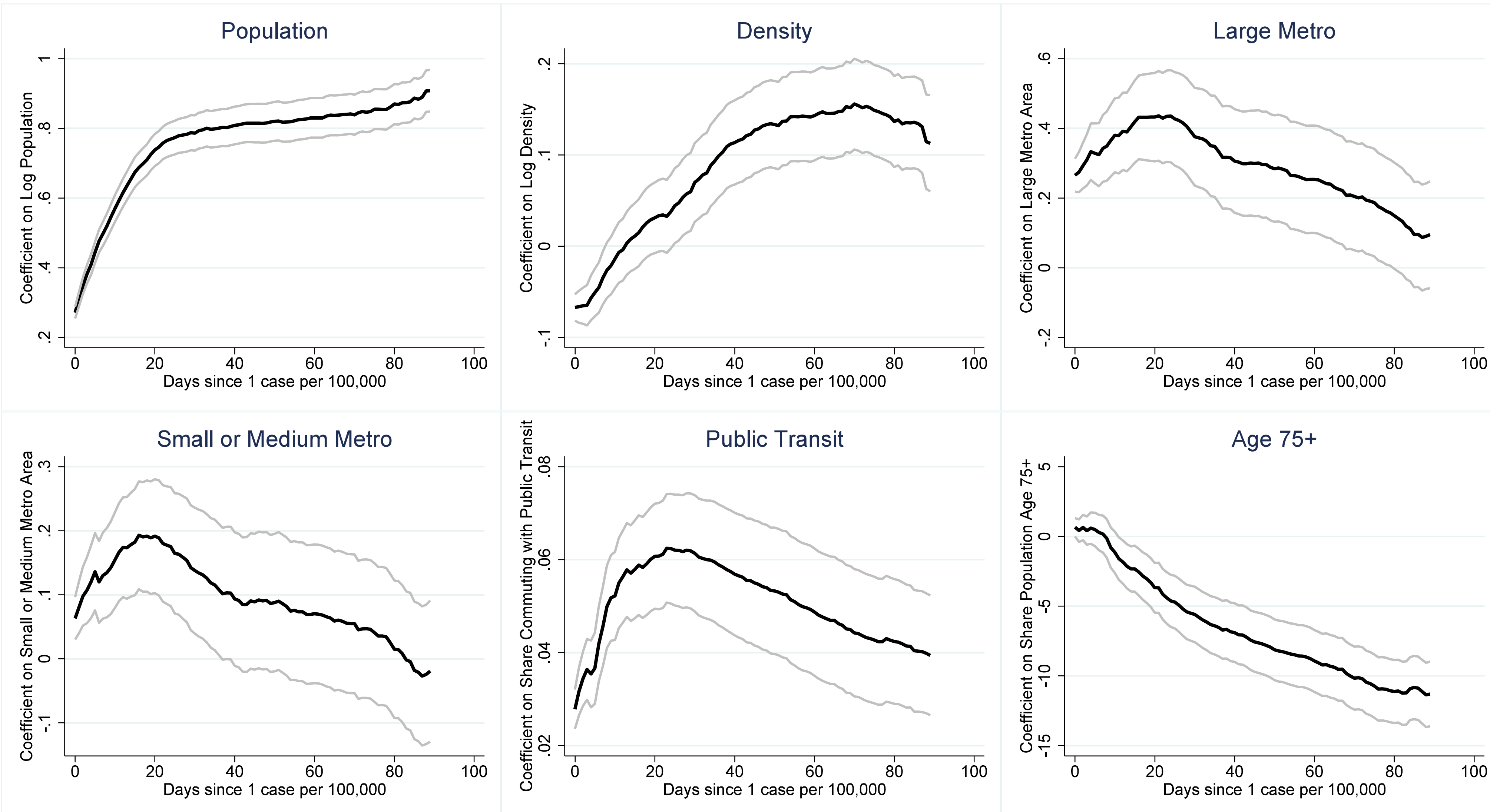


Figure A4A - Effects on Log Cases, by Days Since Onset (contd.)

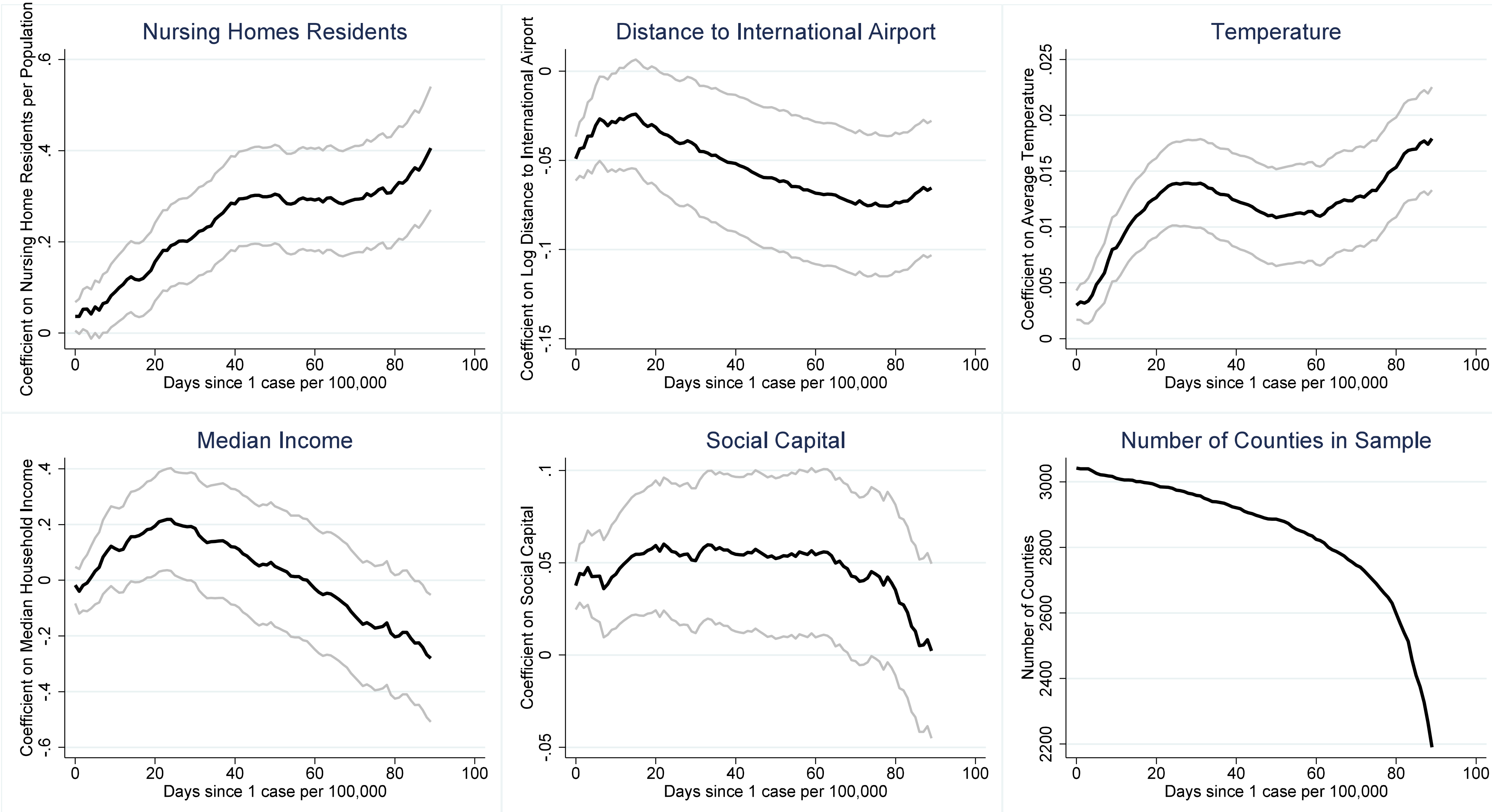


Figure A4B - Effects on Log Deaths, by Days Since Onset

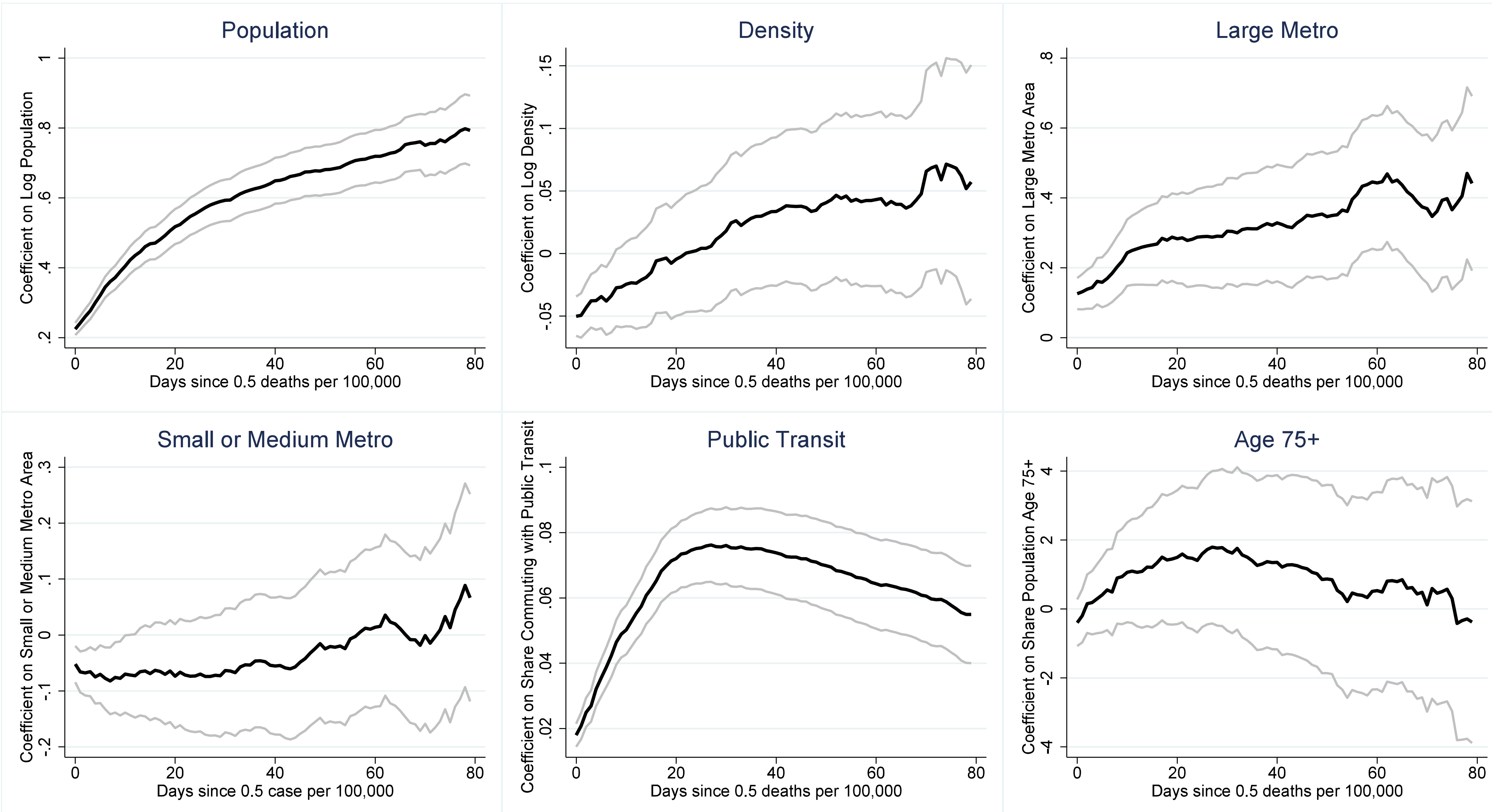


Figure A4B - Effects on Log Deaths, by Days Since Onset (contd.)

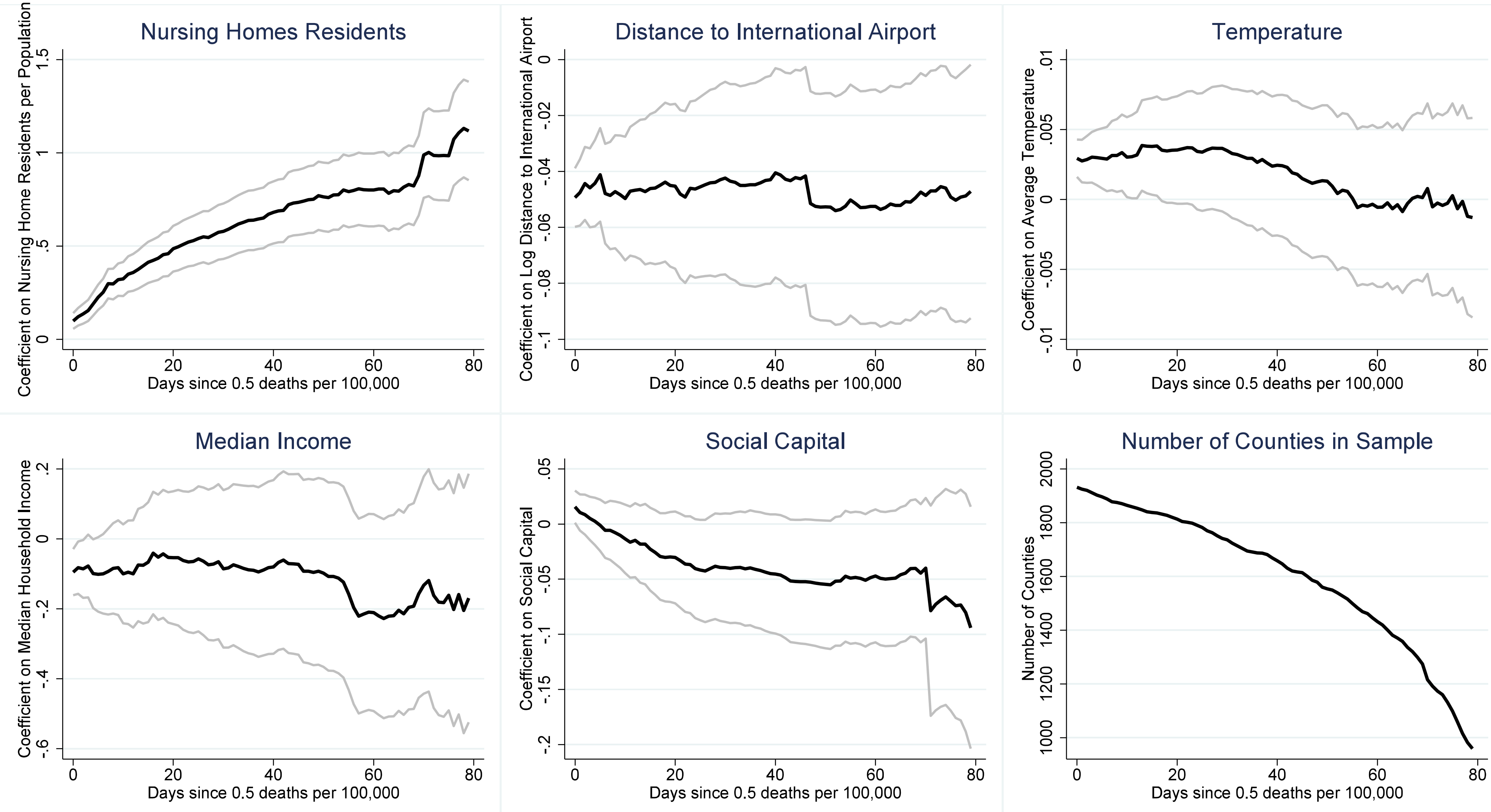


Figure A5A - Log Cases - Effect of Stay-at-Home Order

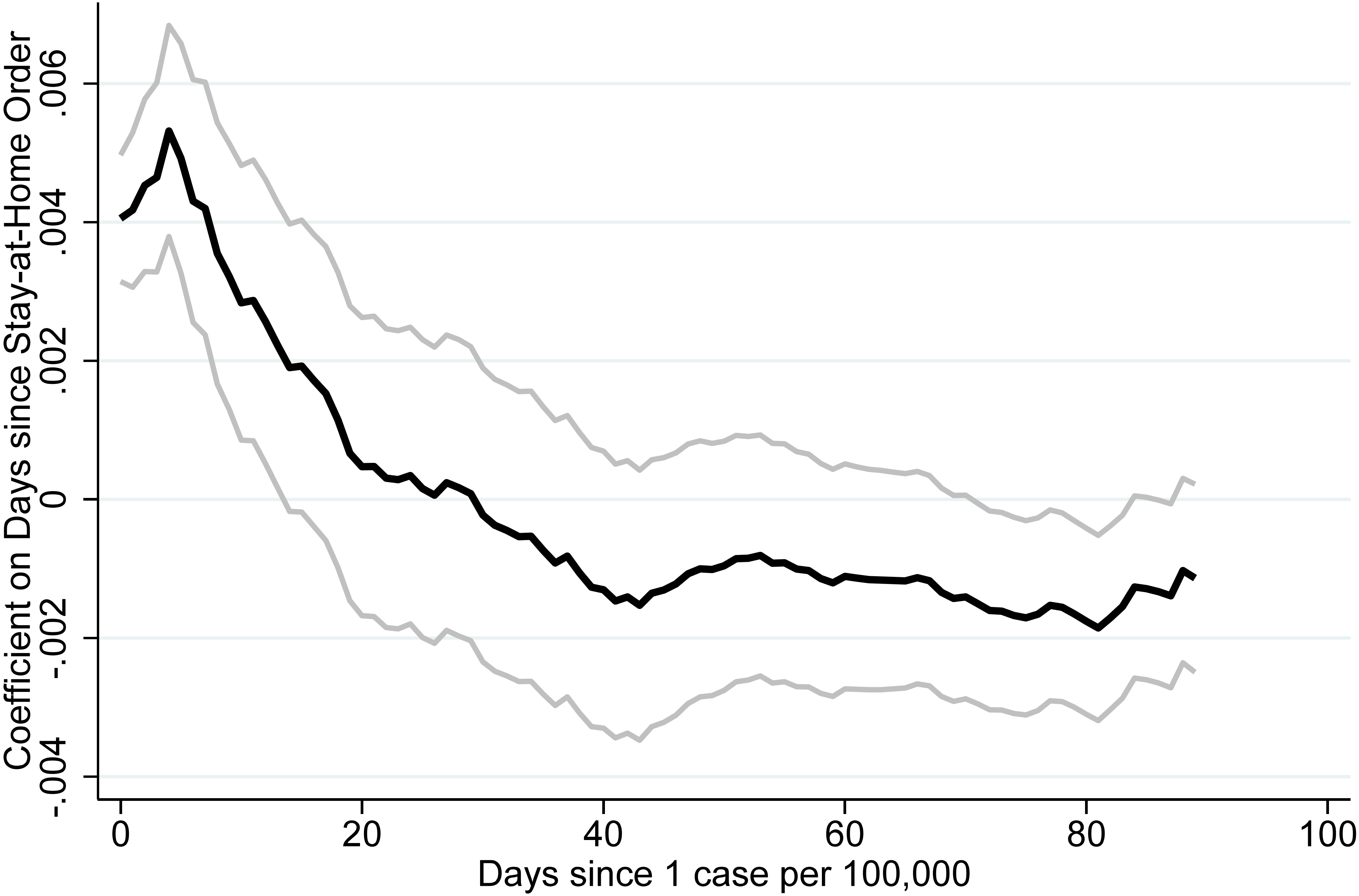


Figure A5B - Log Deaths - Effect of Stay-at-Home Order