I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13440

Urban Density and COVID-19

Felipe Carozzi
Sandro Provenzano
Sefi Roth

JULY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13440

Urban Density and COVID-19

Felipe Carozzi
London School of Economics

Sandro Provenzano
London School of Economics

Sefi Roth
London School of Economics and IZA

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Urban Density and COVID-19*

This paper estimates the link between population density and COVID-19 spread and severity in the contiguous United States. To overcome confounding factors, we use two Instrumental Variable (IV) strategies that exploit geological features and historical populations to induce exogenous variation in population density without affecting COVID-19 related deaths directly. We find that density has affected the timing of the outbreak in each county, with denser locations more likely to have an early outbreak. However, we find no evidence that population density is linked with COVID-19 cases and deaths. Using data from Google, Facebook and the US Census, we also investigate several possible mechanisms for our findings. We show that population density can affect the timing of outbreaks through higher connectedness of denser location. Furthermore, we find that population density is positively associated with proxies of social distancing and negatively associated with the age of the population, highlighting the importance of these mediating factors.

JEL Classification: I12, R12

Keywords: COVID-19, density, congestion forces

Corresponding author:
Sefi Roth
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom
E-mail: s.j.roth@lse.ac.uk

* We would like to thank Gabriel Ahlfeldt for useful comments and suggestions. The work by Provenzano was supported by the Economic and Social Research Council [grant number: ES/P000622/1].

1. Introduction

The ongoing COVID-19 pandemic has led to approximately 6.3 million confirmed cases and 375 thousand deaths globally as of 31 May 2020. Whilst the virus has affected most countries around the world to some extent, there is wide variation between and within countries in the spread and severity of cases. Given the significant health and economic consequences of the pandemic, it is vital to understand the key drivers of this variation to establish an adequate policy response. Historically, cities have been associated with the propagation of infectious diseases.[1] Has density - the defining feature of cities - promoted the spread of COVID-19? Have city dwellers been especially affected by the health consequences of the pandemic?

Estimating how population density shaped the severity of the COVID-19 outbreak is challenging for several reasons. First, population densities are not randomly assigned and they might be correlated with unobserved confounding factors. For example, population densities can be affected by locational productive advantages, whether natural or man-made (e.g. soil quality or transportation infrastructure), that may simultaneously affect local economic conditions and local densities. Insofar as COVID-19 incidence is affected by economic conditions, unobservable locational advantages can confound the effect of density on the spread of the disease. Second, differences in the timing of the onset of the disease can generate cross-sectional differences in the severity of the outbreak at one point in time. Finally, data on COVID-19 cases might be reported with error due to variation in local testing strategy and capacity.

In this paper, we estimate the causal relationship between population density and the impact of COVID-19 in urban counties of the contiguous United States. We overcome the empirical challenges mentioned above in several ways. We use two Instrumental Variable (IV) strategies borrowed from the agglomeration literature to induce plausibly exogenous variation in population density without affecting COVID-19 related deaths directly. More specifically, in our *geological* IV approach, we use the presence of aquifers, earthquake risk, and soil drainage capacity to build an instrument for density (as in Carozzi and Roth 2020). In our *historical* IV strategy, we use the traditional long-lag instrument, which measures urban population density in the 1880 US Census (as in Ciccone and Hall 1996). We study both, how density affected the timing of the outbreak in each county and the time adjusted number of deaths after that outbreak. We focus on the daily number of confirmed COVID-19 deaths rather than cases as our main outcome of interest as it is considered to be a more accurate indicator of local COVID-19 prevalence (Subbaraman 2020). Finally, we cross-validate our COVID-19 figures with data from different sources to ensure reported deaths are consistent with other measures of COVID-19 spread.

To the best of our knowledge, there are only two available studies that have examined the link between density and COVID-19 incidence in the United States.[2] Wheaton and Kin-

[1] Examples in economics include Duranton and Puga (2020) and Voigtländer and Voth (2013).

[2] The literature on the relationship between the 1918 Influenza pandemic (the Spanish Flu) and population density is naturally more developed and can shed light on the link between pandemics and density more broadly.

sella Thompson (2020) used data on 351 cities and towns in Massachusetts to provide a cross-section analysis of the per capita infection rate. They find that population density has an economically and statistically significant positive effect on the incidence of the disease. Almagro and Orane-Hutchinson (2020) also examine this link but use data on the number of tests and positives across NYC zip codes. They also find a significant positive relationship between population density and the share of positive tests, but this relationship seems to decline over time. Importantly, these studies provide descriptive evidence on the correlation between density and the spread of the pandemic, but do not attempt to identify a plausibly causal relationship, nor they discuss the timing of the outbreak.

We find convincing evidence that density has affected the time of the outbreak in each county, with dense locations more likely to have an early outbreak. However, we find no evidence that population density is linked with COVID-19 incidence once we adjust for the timing of the onset of the disease in each county. On first reflection, this second result seems surprising given that the virus spreads via human contact and denser areas provide more opportunities for human interactions. Nevertheless, several mediating factors might explain why the direction of this relationship is in fact ambiguous. For example, variation in density might affect the behavioural responses to the pandemic, which can itself affect the spread and severity of the disease.

We examine several potential mechanisms for our main results using data from Google, Facebook, and the US Census. We begin by exploring the effect of density on Americans' behavioural responses to the pandemic since the spread of the virus is not exclusively a biological phenomenon but also a social one (Papageorge et al. 2020). We show that density is negatively associated with work-related activity between January and April, suggesting that compliance with social distancing measures might be an important mediating factor. Relatedly, we examine whether population density is associated with differences in political preferences. This is motivated by documented partisan differences in Americans' responses to the pandemic. We find that density is negatively associated with the share of Republican voters, which have been shown to be less engaged in social distancing and other efforts to reduce transmission (Allcott et al. 2020). Finally, given that older age is considered to be a significant mortality risk factor of COVID-19 (Zhou et al. 2020), we examine the effect of density on the share of the older population. We find that population density is linked with a smaller share of residents above 60 years of age, highlighting the possibility that the lower share of older residents might also mediate the hypothesised positive effect of density on COVID-19 incidence. Collectively, these results provide suggestive evidence of mechanisms generating offsetting negative effects of density on the severity of the COVID-19 outbreak.

Interestingly, while it may seem intuitive that the influenza pandemic was positively associated with population density as the virus spread via human contact, a review of the literature produce mixed results. For example, Garrett (2007) finds a positive relationship between mortality rates and population density in the US. In contrast, Mills, Robins and Lipsitch (2004) find no statistical association between population density and the initial reproductive number (R) using data on 45 US cities. Chowell et al. (2008) also find no association between transmissibility, death rates and indicators of population density in England and Wales. Ferguson et al. (2006) studies the development of the 1918 pandemic and finds early onset in dense urban cores before a more smooth development of the disease across space.

Our study provides several contributions to the existing academic literature on urban density and has significant policy implications. First, to the best of our knowledge we are the first to credibly estimate the causal relationship between population density and COVID-19 related mortality. Second, our results show population density appears to affect the impact of COVID-19 only through the timing of outbreaks and not through the rate of subsequent spread. As such, our results highlight some weaknesses of the popular hypotheses of the supposedly detrimental effects of the COVID-19 pandemic on cities. Third, our study predicts that in case of a second wave of the pandemic, denser places might be affected earlier (due to their connectedness) but once affected, the spread and severity may not differ from less dense places. Importantly, our results also highlight the gravity of non-pharmaceutical interventions and measures (i.e. social distancing) in containing the spread of the virus.

2. Data

Our dataset combines information on COVID-19 cases and deaths, population density, demographics, social connectedness, behavioural adjustment, voting behaviour and geological features at the US county level. The period under investigation ranges from the the 22nd of January, when the first US case was confirmed in 'King County', up until the 1st of June 2020. We restrict our sample to urban counties[3] in the contiguous United States which leaves us with 1,759 counties representing $\sim$ 93% of the total US population. For certain parts of the analysis, we focus on the outbreak dynamic and therefore reduce the sample further to those counties that had at least one confirmed COVID-19 related death 45 days before the end of our sample period. Our final sample consists of 1,197 counties representing $\sim$ 82 % of the total US population (see Figure A.1). In the following, we describe the dataset and provide further information about the sources and URLs for download in Appendix B and descriptive statistics in Table A.1.

COVID-19 Cases and Deaths

We obtain a panel of daily confirmed COVID-19 fatalities and cases for US counties from usafacts.org. The most intuitive indicator to monitor the COVID-19 outbreak is the daily number of confirmed cases. However, this figure is likely to be distorted by varying local testing strategy and capacity. Furthermore, the ability of the virus to spread across asymptomatic people makes the task of recording the number of infections in the community extremely difficult (Subbaraman (2020)). Therefore, we mainly use the daily number of confirmed COVID-19 deaths as this is a more accurate indicator of the local COVID-19 prevalence. Furthermore, COVID-19 is significantly more lethal than recent epidemics, making the focus on death toll very relevant to understand the pandemic more broadly. In order to ensure that our COVID-19 estimates are consistent, we cross-validate our COVID-19 figures with official data from the Centers for Disease Control and Prevention (CDC). In the left panel of

[3]Urban counties are those that are classified as either 'metropolitan' or 'micropolitan' core-based statistical areas in the 2010 census.

Figure A.2, we compare our total COVID-19 fatality counts by county to the latest figures on officially confirmed deaths due to COVID-19. In the right panel, we compare total fatalities to CDC excess death estimates. Both graphs exhibit strong linear relationships and support the validity of our COVID-19 data.[4]

Population Density

Based on the US census for 2010, we compute two indicators of population density. The first indicator is simply the total population of a county over its total area i.e. this indicator assumes a uniform spatial distribution of populations within a county. The second indicator computes the population density for all census-blocks within a county and then takes the total population weighted mean as our indicator of 'weighted population density'. Population-weighted density is meant to measure average "experienced" density and was popularized in economics after work in Rappaport (2008) and Glaeser and Kahn (2004). It can be obtained using spatially disaggregated data on the spatial distribution of population and weighting each small unit population density by its relative population in the county.

Instrumental Variables:

For our geological instrumental variable estimates we use three different instruments. More specifically, we use variables measuring earthquake risks and presence of aquifers from the United States Geological Survey (USGS) (also used in Duranton and Turner (2018)), and data on soil drainage quality from NRCS State Soil Geographic Data Base. We match our grid cells to the geological data using grid cell centroids to spatially impute data on aquifers, earthquake risks and soil drainage quality. For our historical instrument, we use population density obtained from the 1880 United States census. We impute this data on the county level using spatial matching based on the assumption of uniform population distribution within 1880 counties. [5]

Behavioral Adjustment/Social Distancing:

To measure how much people in different counties adjusted their behaviour as a response to the COVID-19 outbreak we use the 'COVID-19 Community Mobility Reports' by Google (Google CMR). This database aggregates extensive anonymised mobile device GPS user data and estimates the percentage change in activity in certain place categories by county and day. The five week period from January 3rd to February 6th before the start of the COVID-19 outbreak in the US serves as the corresponding baseline period. We focus on two such place categories: (1) 'places of work' and (2) 'retail & recreation' (including restaurants, cafes,

[4] In contrast, the correlation between county level COVID-19 fatalities and USAFacts is -0.001 and insignificant indicating that COVID-19 mortality is not simply an amplification of fatalities occurring under normal circumstances but rather follows distinct patterns that are consistently capture by our database.

[5] Note that, while the assumption of uniform distribution is clearly a simplification which could lead to measurement error, this should not have a substantial impact on our main estimates. This is because measurement error in the instruments could affect the relevance of the instruments but should not generate bias in the coefficients of interest unless the measurement error itself is correlated with COVID-19 incidence.

shopping centers, theme parks, museums, libraries, and movie theatres).

Other Variables:
We obtain data on county-level demographic characteristic estimates for 2018 from the US census. Indicators include the share of male, Black, Asian and over 60 years old population in each county. Social connectedness is measured using Facebook's Social Connectedness Index (Facebook SCI), which measures the intensity of the link between locations using the number of friend links in this social network (See Bailey et al. (2018) for further details on the SCI).

3. Empirical Analysis

The top left panel of Figure 1 illustrates the positive cross-sectional correlation between a county's population density - calculated as the total population over the surface area - and the number of COVID-19 related deaths per capita. This is the basic fact that had been noticed in Wheaton and Kinsella Thompson (2020), Dubner (2020) ,and, as early as April 2020. Similar graphs, again displaying positive relationships using population-weighted densities and number of cases are reported in Appendix Figure A.3.

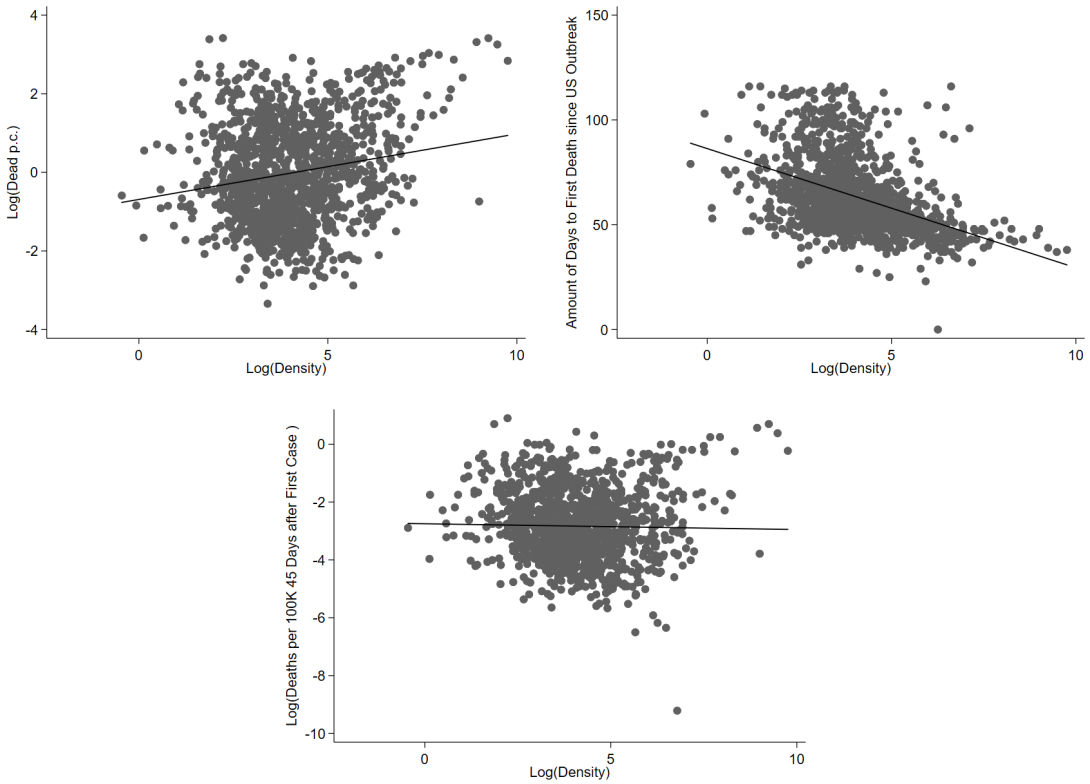
Naturally, these cross-sectional patterns do not constitute conclusive evidence that urban density results in faster or more deadly COVID-19 spread. There are at least two problems that could arise in this context. First, the positive correlation in the top left panel of Figure 1 can be the result of differences in the timing of the onset of the disease across locations. Second, certain location characteristics which are correlated with both density and COVID-19 spread could induce a correlation in the absence of any actual causal link. We discuss this second issue in detail in the next section.

The top right panel of Figure 1 illustrates the point on differences in the timing of the onset of the disease across locations by showing the relationship between population density and the number of days between the 22nd of January and the first fatality in each county. The figure exhibits a clear negative relationship, indicating that dense locations experienced COVID-19 fatalities earlier than more sparsely populated locations.

We can adjust for the differences in the timing of the onset of the disease by computing the number of deaths after a fixed number of days from that onset. This is what is typically shown in cross-country comparisons of the evolution of the pandemic. In our case, we can compute the number of COVID-19 deaths at a specified time after the outbreak started in a county. We define the start of the outbreak as the first day with a reported case and compute the number of deaths 45 days after this date for all counties. The link between these time-adjusted variables and density is illustrated in the bottom panel of Figure 1.

The relationship is almost flat after time-adjusting, suggesting that density does not simply translate into a higher COVID-19 incidence as there might be several mediating factors. We turn to this point later when we discuss mechanisms in section 3.3 but first we illustrate some descriptive results. Figure 2 shows the change in mobility relative to January

Figure 1
COVID-19 and Population Density



Notes: The horizontal axis represents the logarithm of the county's population density. In the top left panel, the vertical axis represents the logarithm of the number of fatalities per thousand inhabitants. In the top right panel, the vertical axis represents the number of days between the 22nd of January and the first fatality in each county. Black markers correspond to counties forming part of a CBSA. Black fit lines estimated via Ordinary Least Squares.

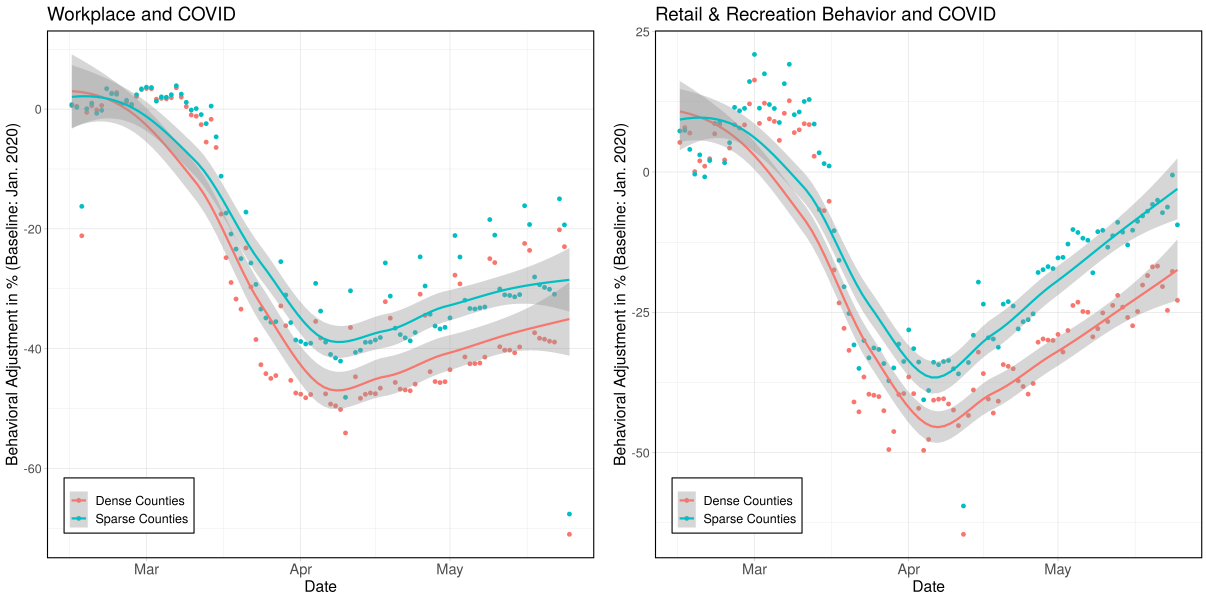
2020 for sparse and dense counties, with the split based on median county density.[6] The left panel corresponds to changes in workplace-related mobility and the right panel corresponds to changes in mobility for shopping and recreation. In both cases, we observe a sharp reduction in mobility relative to the January baseline. Moreover, in both cases, the reduction is more acute in denser counties. These behavioural changes - whether driven by policy or spontaneous - can provide one mechanism that contains disease spread in dense locations. We will return to this point below.

3.1. Estimation

Before we can obtain specific estimates for the relationship between time-adjusted COVID-19 related mortality and density, we also need to deal with potential confounders affecting both density and the prevalence of the disease. Climate conditions, for example, can simultaneously influence household location decisions (see Glaeser, Kolko and Saiz 2001) and

[6]The data is based on COVID-19 Community Mobility Reports released by Google and are based on data from portable device users in United States counties.

Figure 2
Changes in Mobility Relative to January Baseline



Notes: The figures plot the daily change and local regression curve (LOESS) over time in mobility relative to the January 2020 baseline for sparse counties and dense counties, with the split based on median weighted county density. The left panel refers to adjustment of workplace-related activity. The right panel refers to retail and recreational activity including restaurants, cafes, shopping centres , theme parks, museums, libraries, and movie theatres.

COVID-19 spread.[7] Local amenities such as waterfronts or low precipitation levels can themselves influence travel patterns - e.g. by increasing tourist arrivals - which could in turn affect COVID-19 rates. Insofar as some of these elements are observable, we can include them as controls in our regressions. Yet some confounders may be unobservable due to their inherent nature or lack of accurate data. For instance, locational productive advantages can simultaneously affect local economic conditions and increase local densities.[8] Examples range from natural factors such as fertile or irrigable lands to man-made infrastructures such as ports or highways. Insofar as COVID-19 incidence is affected by economic conditions, unobservable locational advantages can confound the effect of density on the spread of the disease.

To overcome the problem posed by potential unobservable confounders we borrow canonical instruments for density from the agglomeration literature (see Combes, Duranton and Gobillon 2011 for a description) and our previous work on the relationship between density and air pollution (see Carozzi and Roth 2020). Specifically, we will instrument population density with either geological factors which can affect the costs of compact urban development or a long-lags in population density.

[7] A number of recent papers document a negative effect of temperature on COVID-19 incidence, at least in temperate weathers. See for example Prata, Rodrigues and Bermejo 2020; Tobías and Molina 2020.

[8] Locational advantages increase local densities because higher land prices in these areas trigger a substitution of land for capital in the production of structures (i.e. an increase in building heights).

We use three geological instruments: the fraction of the urban footprint with aquifer presence, a measure of average earthquake risks and an estimate of soil drainage quality. The rationale for the aquifer instrument is that new dwellings in the periphery of urban areas need either to connect with the municipal network or to directly connect with an underwater source. As the cost of connecting to the municipal network is increasing in the distance to other connected dwellings and the option of the underwater source is only available if there is an aquifer where the dwelling is located, cities with more land over aquifers can sprawl out further, contain more sparse development and lower densities. This instrument is motivated by the work in Burchfield et al. (2006) which reports that aquifers in the urban fringe are associated with urban sprawl. The rationale for our earthquake risk instrument is the expectation that the risk of an earthquake might influence building regulations, construction practices and the space between buildings, thus also affecting urban density. We also expect this instrument to satisfy the exogeneity condition, once we condition for distance to sea, latitude and longitude, and state fixed effects. Finally, the soil drainage quality variable is expected to affect land suitability for building at different densities. In fully urbanized land, a significant fraction of rainfall is drained through drainage networks and sewage systems Konrad (2003). However, at lower densities, soil drainage capacity is important to avoid stagnant water and, possibly, floods. In addition, high drainage soil is not ideal for laying down heavy infrastructure, making the task of building high density development more expensive.

We use a separate instrument for density based on historical population as recorded in the 1880 US census. Settlements in this period were in place before much of the technological revolutions in transportation that have affected location patterns in the last decades and also precede current patterns of industrial location. The use of historical population instruments for density was popularized by Ciccone and Hall (1996) and have often been used in the literature on agglomeration economies since (see Combes and Gobillon (2015) for a review).

Our main estimating equation will regress measures of COVID-19 presence on the logarithm of population density:

$$Y_i = \alpha_s + \beta Ln(Pop.Density)_{it} + \gamma' X_i + \varepsilon_i \tag{1}$$

where i indexes individual counties, α_s is a set of state effects and X_i is a set of controls. In all specifications, we control for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. Our outcomes include different measures of COVID-19 presence. In most of our analysis these are either variables capturing the time it took for the disease to arrive at a county or a time-adjusted measure of COVID-19 presence - the logarithm of the number of COVID-19 fatalities in the county 45 days after the first case was confirmed. Finally, we will consider two alternative measures of density: total population divided by surface area of the county, and population-weighted density.

3.2. Results

We first report baseline cross-sectional correlations between population density and COVID-19 cases and deaths on the 1st of June. In Table A.2, we estimate equation 1 via Ordinary Least Squares (OLS) using the logarithm of the number of cases per 100,000 inhabitants and the logarithm of the number of deaths per 100,000 inhabitants as outcome variables. We find positive and statistically significant effects of population density on COVID-19 incidence, in line with the descriptive evidence reported in Figure 1. Specifically, when using the conventional measure of population density we find elasticities of 22% and 8% for cases and deaths, respectively. This suggests that a 1% increase in population density increases cases and deaths per 100,000 people by 0.22% and 0.08%. When using our population-weighted measure of density, we also find positive elasticities, though these are of slightly smaller magnitude and statistically insignificant in the case of deaths per 100,000 inhabitants. The findings for COVID-19 cases are consistent with the evidence presented by Wheaton and Kinsella Thompson (2020) and Almagro and Orane-Hutchinson (2020). Yet this should not be taken as conclusive evidence that density has a causal effect on the spread of COVID-19. As argued above, potential differences in the timing of the onset of the disease across locations or the presence of potential unobservable confounders can induce substantial bias in these coefficients.

Estimates reported in Table 1 deal with these empirical issues by looking explicitly at differences in the onset of the COVID-19 epidemic across locations and incorporating our instrumental variable strategy. In panels A and B we report estimates for the effect of density on the number of days to the first case and the number of days to the first death. These numbers are measured relative to the date of the first reported case in the United States, so that small numbers correspond to an earlier onset of the disease. In column 1, we report OLS estimates obtained after controlling for state effects and covariates. In columns 2 and 3, we show IV estimates obtained using our Geological and Historical instruments respectively to overcome potential confounders. We find that doubling density is associated with approximately 3 days earlier onset of the disease. Estimates are fairly consistent across panels A and B, as well as across estimation methods. We find that denser areas have indeed experienced earlier onsets of the disease whether we use days to the first case or days to the first death. These estimates are large, demonstrating the importance of adjusting for differences in the timing of the onsets across locations when estimating the relationship between population density and COVID-19 incidence.

In Panel C of Table 1, we examine our main outcome of interest; the effect of population density on COVID-19 related mortality. As mentioned previously, we focus on confirmed COVID-19 related deaths rather than cases as our main outcome of interest as it is considered to be a more accurate indicator of local COVID-19 prevalence. Given our results from Panels A and B, we adjust for differences in the timing of the onset of the disease by constructing our outcome variable as the number of deaths per 100,000, 45 days after the first case. In column 1, we find that the cross-sectional correlation observed in Table A.2 becomes *negative* and statistically insignificant, suggesting that the positive link between population

TABLE 1
ONSET OF THE DISEASE AND DEATHS AFTER 45 DAYS

	OLS	IV	
A. Days to First Case			
Log(Population Density)	-4.578***	-4.093***	-4.617***
	(0.231)	(0.656)	(0.576)
IV F-stat		24.5	122.8
Obs.	1745	1745	1719
B. Days to First Fatality			
Log(Population Density)	-5.493***	-4.627***	-4.097***
	(0.407)	(1.194)	(1.010)
IV F-stat		26.0	84.0
Obs.	1324	1324	1302
C. Log(Deaths per 100,000 after 45 Days)			
Log(Population Density)	-0.105	-0.105	0.010
	(0.070)	(0.146)	(0.086)
F-stat		23.5	78.7
Obs.	1197	1197	1175
Instrument		Geological	Historical
State Effects	No	Yes	Yes

Notes: The main explanatory variable in all models is the natural logarithm of population density. Panels A and B report the estimates for the number of days to the first case and death respectively. Panel C reports the result for the log of the number of deaths per 100,000 residents in a county, 45 days after the first case. Column (1) corresponds to OLS estimates, column (2) and (3) presents 2SLS estimates using the Geological and Historical instruments respectively. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (3) add state effects. SStandard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

density and COVID-19 incidence might have been confounded by differences in the timing of the onsets. In columns 2 and 3, we use our instrumental variable approach to test this hypothesis more convincingly, using an arguably exogenous variation in density which does not affect COVID-19 related mortality directly. Importantly, our first stage estimates yield F-stats of 20 and 69, indicating that our instruments are not weak. Our second stage results reveal a statistically insignificant relationship between population density and COVID-19 related deaths in both columns, portraying a similar picture as the OLS estimate presented in column 1. Our 2SLS results are unsurprisingly less precise, but the overall picture is clear. We find no evidence that population density is linked with COVID-19 related deaths.

We further investigate the link between density and COVID-19 incidence in Table A.3, using population-weighted density as our main regressor of interest. Unfortunately, since our geological instruments do not provide a strong first stage in this setting, our IV analysis relies solely on our long lag instrument. Reassuringly, we find that the overall results are similar to those obtained in Table 1. Panels A and B show denser counties had earlier onsets of the disease compared to sparse counties. In panel C, we find a *negative* association between

weighted density and COVID-19 related deaths when using OLS. However, our IV estimates again show a statistically insignificant elasticity. We therefore conclude that variation in density did not result in faster spread of COVID-19 in the United States beyond the effect on early onset of the disease despite prior descriptive evidence suggesting otherwise.

On first reflection, our results seem surprising given that the virus spreads via human contact and denser areas can provide more opportunities for human interactions. Nevertheless, there are several mediating factors that might offset this intuitive mechanism. For example, density itself might attract younger residents who are less likely to develop symptoms. In addition, both behavioural and/or policy induced changes in behaviour may be different in dense counties. In fact, studies on previous pandemics (e.g. the 1918 influenza pandemic) also show that population density is not necessarily linked with the spread and severity of a disease (Mills, Robins and Lipsitch 2004). Our analysis thus far has explored how changes in density, ceteris paribus, affected COVID-19 incidence. In the next section, we explore potential mechanisms underlying our reduced-form results.

3.3. Mechanisms

Variation in density might lead to changes in several local conditions, which can themselves affect the spread and severity of the disease. These types of changes may provide mechanisms that reinforce or offset the hypothesised positive effects that have been suggested in the literature, both in terms of timing of the local onset of the pandemic and subsequent spread. We turn to study some of these mechanisms by estimating the effect of density on other determinants of COVID-19 spread. To do so, we re-estimate equation 1 using these hypothetical mediators as outcomes. The resulting estimates do not provide definite proof regarding the mechanisms explaining the effect of density on COVID-19 incidence, but should be interpreted as suggestive evidence in this regard.

We begin by looking at possible factors explaining the early onset of the disease in denser cities by studying whether social connectedness with other counties in the US is affected by density. Our proxy for this variable relies on Facebook's Social Connectedness Index.[9] This index is based on the relative frequency of friendship links between users of the social-network, with higher index values corresponding to a larger number of friendship links. To proxy for social connectedness with other counties we aggregate the SCI of each US county with all other counties and normalizes it by the own-county SCI. The resulting variable takes larger values in counties that have a larger number of connections with other counties. Coefficients resulting from estimating equation 1 using the logarithm of this proxy as an outcome variable are provided in Panel A of Table 2. As above, we report both OLS estimates (column 1) and 2SLS estimates using our geological and historical instruments (columns 2 and 3). We observe consistently positive elasticities of roughly 0.5 across columns, indicating denser counties are more intensely related to other counties in the US. These results can explain our findings of early onsets of COVID-19 cases and deaths in denser counties illustrated in

[9] Kuchler, Russel and Stroebel (2020) study how social networks provided a channel for the spread of the disease based on the SCI.

Figure 1 and Table 1

TABLE 2
SUGGESTED MECHANISMS

	OLS	IV	
A. Social Connectedness			
Log(Population Density)	0.552***	0.429***	0.395***
	(0.019)	(0.045)	(0.030)
IV F-stat		24.5	122.8
Obs.	1758	1758	1732
B. Δ Workplace Related Activity			
Log(Population Density)	-4.033***	-5.095***	-3.661***
	(0.172)	(0.520)	(0.286)
IV F-stat		17.3	70.9
Obs.	1355	1355	1336
C. Δ Retail Related Activity			
Log(Population Density)	-3.024***	-2.840***	-3.406***
	(0.473)	(1.101)	(0.622)
IV F-stat		16.3	60.7
Obs.	1289	1289	1270
D. Republican Vote Share 2016			
Log(Population Density)	-0.052***	-0.013	-0.080***
	(0.003)	(0.012)	(0.007)
IV F-stat		24.5	122.8
Obs.	1759	1759	1733
E. Share of Pop. Above 60 Years			
Log(Population Density)	-0.016***	0.001	-0.014***
	(0.001)	(0.005)	(0.003)
F-stat		24.5	122.8
Obs.	1759	1759	1733
Instrument		Geological	Historical
State Effects	No	Yes	Yes

Notes: The main explanatory variable in all models is the natural logarithm of population density. In Panel A, we present the results for the social connectedness of a county based on Facebook's Social Connectedness Index. Panels B and C report the results on behavioural adjustment of workplace and retail-and-recreation related activities relative to the January baseline respectively. Panel D features the results on votes for the Republican party in the 2016 presidential election. Panel E reports the estimates for the share of population above 60 years of age. Column (1) corresponds to OLS estimates, column (2) and (3) presents 2SLS estimates using the Geological and Historical instruments respectively. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (3) add state effects. Standard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

Next, we study how density affects behavioural responses to the pandemic (e.g. compliance with social distancing measures). We use data from Google COVID-19 Community Mo-

bility Reports (Google CMR) to measure how mobility patterns in each county have changed relative to baseline levels measured in January 2020. In Panels B and C of Table 4, we show the relationship between county density and the change in mobility to workplaces and retail activities, respectively. We find population density is associated with a larger decline in mobility for both of these purposes between January and April. Doubling density reduces workplace-related mobility by approximately 2.5-3%, and retail or entertainment related mobility by 2-2.5%. Given the significant variation in density across US counties, these estimates are large. Insofar as social distancing reduces the spread of the disease, these differences in behaviour might explain why we find limited differences in spread by location after accounting for the timing of onset of the disease.

Several factors could explain this difference in behaviour across dense and sparse counties. One candidate that could account for both policy responses and individual differences in behaviour relates to ideological or political views. There have been significant partisan differences in Americans' response to the pandemic. For example, Allcott et al. (2020) show that Republican county vote share has a positive and significant association with the number of weekly visits to points of interest during the peak of the social distancing measures in April. Anecdotal evidence also reveals substantial differences in the tone of the Democratic and Republican parties when discussing the pandemic and its consequences. If density is associated with reduced support for the Republican party, residents of denser areas may be more likely to comply with the social distancing advise. In Panel D of Table 2, we estimate this link using voting data from the 2016 presidential election as a proxy for Republican support. We find that population density has a negative association with the share of republican voters, an observation that should come as no surprise for observers of US politics.[10] This difference in attitudes or political preferences across locations could explain, at least in part, the observed differences in the behavioural response to the pandemic illustrated in Figure 2.

Finally, in Panel E of Table 4 we examine the effect of density on the share of population above 60 years of age. This is of particular importance given that older age considered to be a significant risk factor (Zhou et al., 2020) and that population density is likely to affect the age structure of local areas via its impact on employment opportunities (see Glaeser 1999). Indeed, we find that population density is linked with a smaller share of residents above 60 years of age. This result highlights the possibility that the lower share of older residents might mediate the hypothesised positive effect of density on COVID-19 incidence.

We can arrive to three conclusions from the results reported in Table 2. First, dense counties are more connected with other locations and this may account for earlier onset of the COVID-19 epidemic in these areas. Second, the behavioural response to the disease was larger in denser counties, with less mobility for work and leisure purposes in these locations. We hypothesize that this may be linked to differences in attitudes towards the

[10]Please note that this relationship remains highly robust upon controlling for the share of Afro Americans as well as the population above 60 years of age. In fact, when adding these additional controls, the relationship remains between -0.04 and -0.05 and significant at the 99% confidence level for all three estimation approaches.

pandemic across the density distribution. Finally, dense counties are "younger" than sparse counties and this could reduce the number of deaths in these areas. Our points relating to behavioural responses and demographics provide probable explanations for the surprisingly flat relationship between density and COVID-19 spread reported in panel C of Tables 1 and A.3.

3.4. Robustness Checks

In this section, we provide several robustness checks for our main findings. We first look at our results for the number of COVID-19 deaths 45 days after the onset of the disease in each county. In Panel A of Appendix Table A.4, we test whether the null effect of density is affected by flexibly controlling by week of onset in each state. This goes beyond simply time-adjusting the outcome variable of interest as it also incorporates differences in knowledge regarding the disease or country-wide behavioural adjustments. We find that our qualitative results remain unchanged, with coefficients being insignificantly different from 0 across specifications. Note that the point estimates obtained under OLS and 2SLS with geological instruments are both *negative*. In panel B, we test whether our results are affected by excluding the New York metropolitan area.[11]. In this case, we find a negative and statistically significant relationship between density and time-adjusted COVID-19 deaths in the first two columns. We interpret these results with caution, as we are imposing sample selection that simultaneously exclude the MSA with the largest initial outbreak and the highest density. Results in Table A.4 further emphasize that the time adjusted number of deaths does not appear to be affected positively by density.

We also check the robustness of our results regarding suggested mechanisms for the link between density and COVID-19 deaths to our definition of density. We reproduce Table 2 using the population-weighted densities as the main regressor of interest. Recall that in this case we can only use our long lag instrument as geological instruments are weak predictors of population-weighted densities. Results are presented in Appendix Table A.5 and are qualitatively analogous to those presented for the conventional measure of density. Hence, we conclude that evidence in support for our suggested mechanisms does not depend on the chosen measure for density.

Finally, we test whether density affects the time-adjusted number of reported cases of COVID-19. As argued above, the number of cases is more likely to be affected by variation in testing resources and asymptomatic cases. This motivates our focus on number of deaths in much of the main analysis. Yet, data on reported cases can be used instead. In Panel A of Table A.5, we report estimates of the relationship between density and the number of cases per capita 45 days after the first reported case in the county. IV estimates are not completely conclusive, with a negative and marginally significant elasticity reported using the geological IVs and a positive but insignificant elasticity when using the historical instrument. We replicate our estimates using the log of the number of cases per 100,000 people after 30 days as the dependent variable. Results are reported in Panel B of Table A.5 and show

[11]We use the census 2010 definition corresponding to the New York-Northern New Jersey-Long Island CBSA.

insignificant or even negative effects of density on the time-adjusted number of cases in US counties. We conclude that the data does not yield evidence indicating a clear effect of density on the spread of the disease.

4. Conclusions

Urban areas are often places of intense social interactions, crowded living and close contact. Whether Justinian's Constantinople, fourteenth century Florence or 1918 Philadelphia - cities have historically been associated with the propagation of infectious disease. In the first three months of the COVID-19 global pandemic, large, dense urban areas around the world such as New York, Madrid and London were identified as disease hotspots. Increased awareness of the risks of present and future epidemics has understandably prompted a debate about the future of cities. Does density - the defining feature of cities - promote the spread of the disease? Will this affect the long-run outlook of urban areas?

Our analysis of the onset of the COVID-19 pandemic in the United States raises a series of important points regarding these questions. First, density is associated with an early arrival of COVID-19, so that urban cores and superstar cities get a head start on the spread of the disease. Second, the subsequent spread - once COVID-19 has arrived - is not faster than in smaller towns or sparsely populated peripheries. Cities get hit first, but do not necessarily get hit harder. Third, several mechanisms may explain these findings. Large cities are intensely inter-connected with other locations, which can explain early onset. Yet, in the case of within-city spread, many different offsetting forces may be at play. Crowding may promote the spread of the disease but differences in demographics or precautionary measures may contain it. As a result, it is important to distinguish differences in spread between and within locations.

This paper is based on reported patterns for the spread of the disease in the US over a relatively short period of five months. Therefore, drawing definitive conclusions of long-term spread across urban systems is hardly warranted. Yet, our results may be useful for understanding and predicting the dynamics of future waves of viral disease outbreak across urban areas. As such, our findings may help policy makers to put in place sensitive measures to contain outbreaks. Lastly, by showing that the time-adjusted number of COVID-19 related deaths appears not to be affected by density, we also cast doubts on hasty predictions on the consequences of dense urban living.

References

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Y Yang. 2020. "Polarization and public health: Partisan differences in social distancing during the Coronavirus pandemic." *NBER Working Paper*, , (w26946).

Almagro, Milena, and Angelo Orane-Hutchinson. 2020. "The Determinants of the Differential Exposure to COVID-19 in New York City and Their Evolution Over Time." *Covid Economics: Vetted and Real-Time Papers*, , (13).

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social connectedness: Measurement, determinants, and effects." *Journal of Economic Perspectives*, 32(3): 259–80.

Burchfield, Marcy, Henry G Overman, Diego Puga, and Matthew A Turner. 2006. "Causes of sprawl: A portrait from space." *The Quarterly Journal of Economics*, 121(2).

Carozzi, Felipe, and Sefi Roth. 2020. "Dirty density: Air quality and the density of American cities." *IZA Discussion Paper*.

Chowell, Gerardo, Luis MA Bettencourt, Niall Johnson, Wladimir J Alonso, and Cecile Viboud. 2008. "The 1918–1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact." *Proceedings of the Royal Society B: Biological Sciences*, 275(1634): 501–509.

Ciccone, Antonio, and Robert E Hall. 1996. "Productivity and the Density of Economic Activity." *The American Economic Review*, 54–70.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The empirics of agglomeration economies." In *Handbook of regional and urban economics*. Vol. 5, 247–348. Elsevier.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011. "The identification of agglomeration economies." *Journal of economic geography*, 11(2): 253–266.

Dubner, Stephen J. 2020. "What Does Covid-19 Mean for Cities (and Marriages)?" *Freakonomics Podcast Ep.401*.

Duranton, Gilles, and Diego Puga. 2020. "The Economics of Urban Density." *Journal of Economic Perspectives*.

Duranton, Gilles, and Matthew A Turner. 2018. "Urban form and driving: Evidence from US cities." *Journal of Urban Economics*, 108: 170–191.

Ferguson, Neil M, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. 2006. "Strategies for mitigating an influenza pandemic." *Nature*, 442(7101): 448–452.

Garrett, Thomas A. 2007. "Economic effects of the 1918 influenza pandemic." *Federal Reserve Bank of St. Louis*.

Glaeser, Edward L. 1999. "Learning in cities." *Journal of urban Economics*, 46(2): 254–277.

Glaeser, Edward L, and Matthew E Kahn. 2004. "Sprawl and urban growth." In *Handbook of regional and urban economics*. Vol. 4, 2481–2527. Elsevier.

Glaeser, Edward L, Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of economic geography*, 1(1): 27–50.

Konrad, Christopher Peter. 2003. "Effects of urban development on floods."

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel. 2020. "The geographic spread of COVID-19 correlates with structure of social networks as measured by Facebook." National Bureau of Economic Research.

Mills, Christina E, James M Robins, and Marc Lipsitch. 2004. "Transmissibility of 1918 pandemic influenza." *Nature*, 432(7019): 904–906.

Papageorge, Nicholas W, Matthew V Zahn, Michèle Belot, Eline van den Broek-Altenburg, Syngjoo Choi, Julian C Jamison, Egon Tripodi, et al. 2020. "Socio-Demographic Factors Associated with Self-Protecting Behavior during the COVID-19 Pandemic." Institute of Labor Economics (IZA).

Prata, David N, Waldecy Rodrigues, and Paulo H Bermejo. 2020. "Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil." *Science of the Total Environment*, 138862.

Rappaport, Jordan. 2008. "A productivity model of city crowdedness." *Journal of Urban Economics*, 63(2): 715–722.

Subbaraman, N. 2020. "Why daily death tolls have become unusually important in understanding the coronavirus pandemic." *Nature*.

Tobías, Aurelio, and Tomás Molina. 2020. "Is temperature reducing the transmission of COVID-19?" *Environmental Research*, 186: 109553.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "The three horsemen of riches: Plague, war, and urbanization in early modern Europe." *Review of Economic Studies*, 80(2): 774–811.

Wheaton, William C, and Anne Kinsella Thompson. 2020. "The Geography of COVID-19 growth in the US: Counties and Metropolitan Areas." *Available at SSRN 3570540*.

Zhou, Fei, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. 2020. "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study." *The lancet*.

Appendix

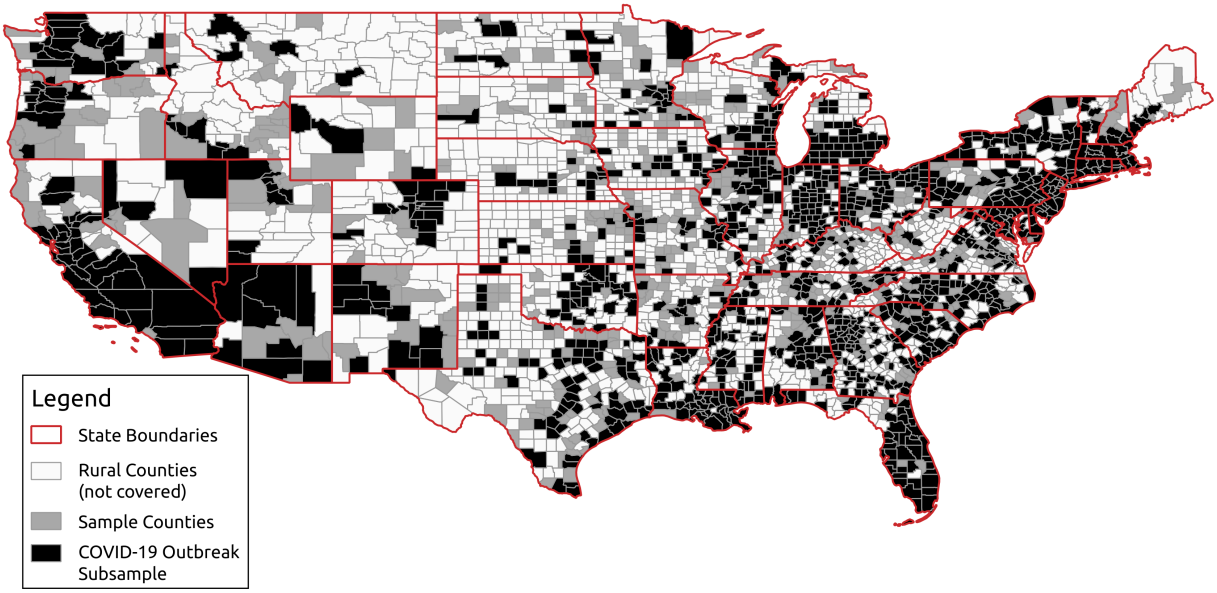
A. Additional Tables and Figures

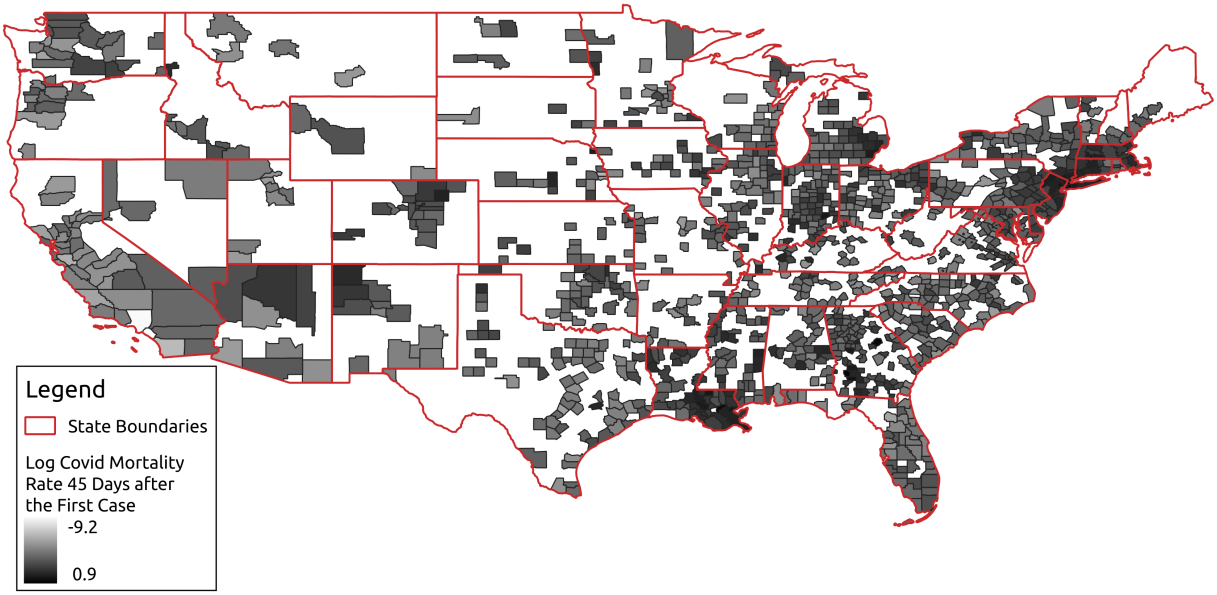
TABLE A.1
DESCRIPTIVE STATISTICS

	Mean	Standard Deviation
A. Whole Sample		
Population Density	147	696
Weighted Population Density	522	1,117
Population	173,406	432,333
COVID-19 Deaths 45 Days after first Case	25.7	171.5
COVID-19 Cases 45 Days after first Case	488.1	2,280.4
Δ Workplace Related Activity	-40.6	7.8
Δ Retail Related Activity	-35.6	12.0
Number of Counties:	1,759	
Share of US population:	93%	
	Mean	Standard Deviation
B. COVID-19 Outbreak Subsample		
Population Density	195	822
Weighted Population Density	644	1,308
Population	225,227	467,881
COVID-19 Deaths 45 Days after first Case	36.8	204.4
COVID-19 Cases 45 Days after first Case	686.6	2,706.5
Δ Workplace Related Activity	-41.8	7.9
Δ Retail Related Activity	-36.3	11.1
Number of Counties:	1,197	
Share of US population:	82%	

Notes: Descriptive statistics presenting the mean and standard deviation for a set of key variables of interest. Panel A corresponds to the whole sample of urban counties. Panel B corresponds to the COVID-19 subsample consisting of counties that had at least one confirmed COVID-19 case 45 days before the end of our sample period on the 1st of June 2020 (Panel B).

Figure A.1
Sample Counties, COVID-19 and Population Density





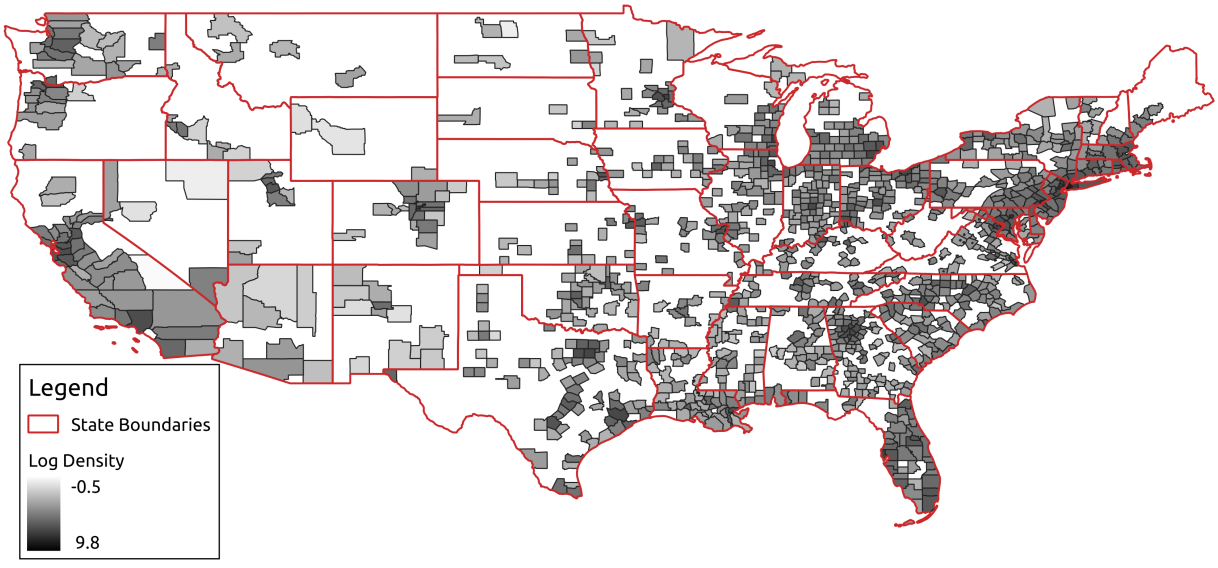
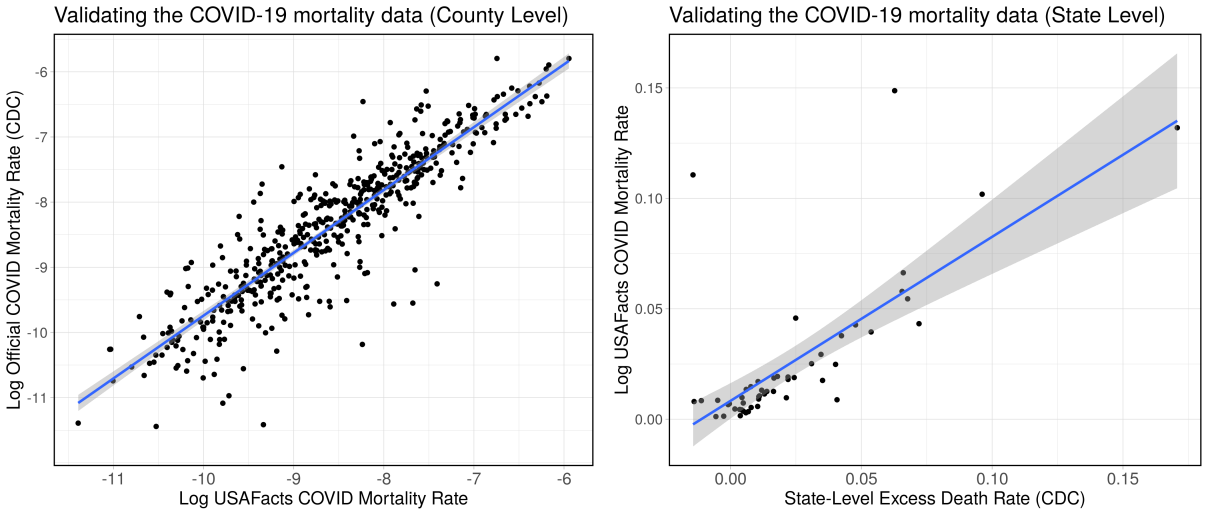
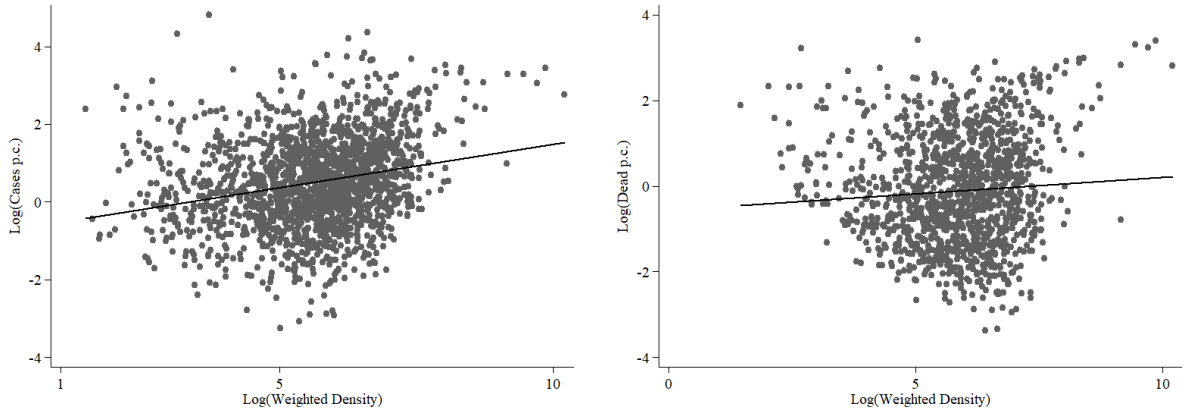


Figure A.2
Validating Covid-19 Figures



Notes: In the left panel, the vertical axis represents the log of the officially confirmed COVID-19 mortality rate per county by the CDC and the horizontal axis the COVID-19 mortality rate by USAFacts. The right panel plots the USAFacts state-level mortality rate (vertical axis) over the excess death estimates by the CDC (horizontal axis). Blue fit lines estimated via Ordinary Least Squares including the 95% confidence interval in grey.

Figure A.3
Cases and Deaths per 100,000 vs. Weighted Density



Notes: The horizontal axis represents the logarithm of the county's population-weighted density. In the left panel, the vertical axis represents the logarithm of the number of cases per 100,000 inhabitants. In the right panel, the vertical axis represents the logarithm of the number of fatalities per thousand inhabitants. Black markers correspond to counties forming part of a CBSA. Black fit lines estimated via Ordinary Least Squares.

TABLE A.2
CASES AND DEATHS: BASELINE OLS ESTIMATES

	Log(Cases per 100,000)		Log(Deaths per 100,000)	
Log(Population Density)	0.234***	0.219***	0.074	0.081**
	(0.033)	(0.027)	(0.057)	(0.040)
Obs.	1745	1745	1319	1319
	Log(Cases per 100,000)		Log(Deaths per 100,000)	
Log(Weight. Density)	0.237***	0.206***	0.083	0.057
	(0.035)	(0.026)	(0.066)	(0.042)
State Effects	No	Yes	No	Yes
Obs.	1745	1745	1319	1319

Notes: Baseline OLS estimates. Columns (1) and (2) use the log of cases per 100,000, columns (3) and (4) the log of deaths per 100,000 inhabitants on the 1st of July as dependent variables. In the top, the log of population density constitutes the explanatory variable, in the bottom it is the log of population weighted density. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (4) add state effects. Standard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

Table A.3

Weighted Densities: Onset of the Disease and Deaths after 45 Days

	OLS	IV
A. Days to First Case		
Log(Weight. Density)	-4.212***	-9.290***
	(0.262)	(1.506)
IV F-stat		31.5
Obs.	1745	1719
B. Days to First Fatality		
Log(Weight. Density)	-5.241***	-9.418***
	(0.482)	(2.667)
IV F-stat		19.8
Obs.	1324	1302
C. Log(Deaths per 100,000 after 45 Days)		
Log(Weight. Density)	-0.101**	0.022
	(0.049)	(0.188)
F-stat		21.0
Obs.	1197	1175
Instrument		Historical
State Effects	No	Yes

Notes: The main explanatory variable in all models is the natural logarithm of weighted density. Panels A and B report the estimates for the number of days to the first case and death respectively. Panel C reports the result for the log of the number of deaths per 100,000 inhabitants in a county, 45 days after the first case. Column (1) corresponds to OLS estimates and column (2) presents 2SLS estimates using the Historical instrument. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (3) add state effects. Standard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

TABLE A.4
ROBUSTNESS: DENSITY AND DEATHS 45 DAYS AFTER FIRST CASE

	OLS	IV	
A. Controlling for Week of Onset Effects			
Log(Population Density)	-0.091	-0.075	0.086
	(0.056)	(0.191)	(0.102)
Instrument		Geological	Historical
F-stat		19.2	69.2
State Effects	Yes	Yes	Yes
Obs.	Yes		No
N	1197	1197	1175
B. Excluding New York State			
Log(Population Density)	-0.105	-0.105	0.010
	(0.070)	(0.146)	(0.086)
F-stat		23.5	78.7
Obs.	1197	1197	1175
Instrument		Geological	Historical
State Effects	No	Yes	Yes

Notes: Robustness tests corresponding to Table 1 Panel C, additionally controlling for the the week of the onset (Panel A) and excluding New York State (Panel B). The main explanatory variable in all models is the natural logarithm of population density. The dependent variable is the log of the number of deaths per 100,000 inhabitants in a county 45 days after the first case. Column (1) corresponds to OLS estimates, column (2) and (3) refer to 2SLS estimates using the Geological and Historical instruments respectively. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (3) add state effects. Standard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

Table A.5
Robustness: Cases 30 and 45 Days after First Case

	OLS	IV	
A. Log(Cases after 45 Days)			
Log(Population Density)	0.094**	-0.254*	0.112
	(0.043)	(0.138)	(0.071)
IV F-stat		25.2	117.4
Obs.	1716	1716	1691
B. Log(Cases after 30 Days)			
Log(Population Density)	0.027	-0.250*	0.022
	(0.045)	(0.130)	(0.072)
F-stat		23.5	78.7
Obs.	1734	1734	1708
Instrument		Geological	Historical
State Effects	No	Yes	Yes

Notes: The dependent variables are the log of the number of cases 45 days (Panel A) and 30 days (Panel B) after the first confirmed case. Column (1) corresponds to OLS estimates, column (2) and (3) refer to 2SLS estimates using the Geological and Historical instruments respectively. In all models, we include controls for average maximum and minimum temperatures, average yearly precipitation, latitude, longitude, distance between the county centroid and the closest sea front and distance to the closest waterfront. The specifications in columns (2) and (3) add state effects. Standard errors in parenthesis are clustered at the CBSA level. ***p<0.01, **p<0.05, *p<0.1.

TABLE A.6
ROBUSTNESS: SUGGESTED MECHANISMS AND WEIGHTED DENSITIES

	OLS	IV
A. Social Connectedness		
Log(Weight. Density)	0.482***	0.743***
	(0.023)	(0.085)
IV F-stat		34.5
Obs.	1758	1732
B. Δ Workplace Related Activity		
Log(Weight. Density)	-3.244***	-6.935***
	(0.227)	(1.011)
IV F-stat		20.4
Obs.	1355	1336
C. Δ Retail Related Activity		
Log(Weight. Density)	-2.844***	-6.884***
	(0.541)	(1.546)
IV F-stat		16.4
Obs.	1289	1270
D. Republican Vote Share 2016		
Log(Weight. Density)	-0.053***	-0.150***
	(0.004)	(0.019)
IV F-stat		34.7
Obs.	1759	1733
E. Share of Pop. Above 60 Years		
Log(Weight. Density)	-0.017***	-0.027***
	(0.001)	(0.005)
F-stat		34.7
Obs.	1759	1733
Instrument		Historical
State Effects	No	Yes

Notes: Corresponds to Table 2, using the log of weighted density as the main explanatory variable.

B. Data Sources

- **USAfacts.org COVID-19 Data**
  The USAFacts is a non-profit civic initiative that provides data on the US population and government and works in partnership with the Penn Wharton Budget Model and the Stanford Institute for Economic Policy Research (SIEPR). The data can be retrieved at: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/. [Last visited: June 2nd 2020]

- **CDC Official COVID-19 Mortality Rate** This database comprises confirmed or presumed COVID-19 fatalities and is limited to counties with at least 10 COVID-19 deaths. It should be noted, the dataset is incomplete because of the time lag between the death and the official certificate submitted to the National Center for Health Statistics (NCHS). For this reason, we this data corresponds only to 514 counties. Our version of the data dates to the 23rd of May. The latest figures can be downloaded at: https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-in-the-United-St/kn79-hsxy. [Last visited: June 1st 2020]

- **CDC Excess Mortality** Excess mortality corresponds to the deviation of total deaths to average expected deaths based on the experience in past years for each state and week from Feburary to May 2020. Our version of the CDC excess mortality estimate dates to the 27th of May 2020. The latest estimates can be downloaded at: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm. [Last visited: June 1st 2020]

- **US Census** contains information about demographics on the country level and can be accessed via: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html. [Last visited: May 14th 2020]

- **'COVID-19 Community Mobility Reports' by Google**
  This report contains information about the behavioral activity change and social distancing in response to the COVID outbreak by county and day. For more detail on this database please visit https://www.google.com/covid19/mobility/data_documentation.html?hl=en. [Last visited: June 3rd 2020]

- **Social Connectedness Data** Obtained after presenting a brief email application for the data based on this paper's outline to Mike Bailey and others at Facebook. April 6 2020 Release Version.