NBER WORKING PAPER SERIES

INEQUALITY AND THE CORONAVIRUS: SOCIOECONOMIC COVARIATES OF BEHAVIORAL RESPONSES AND VIRAL OUTCOMES ACROSS US COUNTIES

Caitlin S. Brown
Martin Ravallion

Working Paper 27549
http://www.nber.org/papers/w27549

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020

The idea for this paper grew out of discussions with Dominique van de Walle, who has kindly commented as the paper evolved. We thank Michael Lokshin, Mead Over and Vasia Panousi for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Inequality and the Coronavirus: Socioeconomic Covariates of Behavioral Responses and Viral Outcomes Across US Counties
Caitlin S. Brown and Martin Ravallion
NBER Working Paper No. 27549
July 2020
JEL No. I12,I14,I32

ABSTRACT

Not much is obvious about how socioeconomic inequalities impact the spread of infectious diseases once one considers behavioral responses, correlations among multiple covariates and the likely non-linearities and dynamics involved. Social distancing responses to the threat of catching COVID-19 and outcomes for infections and deaths are modelled across US counties, augmenting epidemiological and health covariates with within-county median incomes, poverty and income inequality, and age and racial composition. Systematic socioeconomic effects on social distancing and infections emerge, and most effects do not fade as the virus spreads. Deaths, once infected, are less responsive to socioeconomic covariates. Richer counties tend to see greater gains in social distancing and lower infection rates, controlling for more standard epidemiological factors. Income poverty and inequality tend to increase the infection rate, but these effects are largely accountable to their correlation with racial composition. A more elderly population increases deaths conditional on infections, but has an offsetting effect on the infection rate, consistent with the behavioral responses we find through social distancing.

Caitlin S. Brown
School of Public Policy
Central European University
brownc@spp.ceu.edu

Martin Ravallion
Department of Economics
Georgetown University
ICC 580
Washington, DC 20057
and NBER
mr1185@georgetown.edu

1. Introduction

Epidemiological models of the spread of infections have typically assumed homogeneous populations in a given context.[2] While these models have provided valuable starting points for context-specific predictions, a richer modelling framework would allow for heterogeneity in the characteristics within the specific human populations exposed to infectious diseases. Much of that heterogeneity reflects inequalities in various dimensions.

In the case of infections such as COVID-19, the reliance on non-pharmaceutical interventions points to the importance of understanding the role played by socio-economic inequalities, as sources of heterogeneity in initial conditions and behavior. A large body of social and economic thought then becomes relevant to control of the spread of infections. For example, it has been argued that there is an important complementarity between policies that aim to support consumption by poor people and health-care policies in an epidemic.[3] The socio-economic incidence of COVID-19 can also be viewed as an instance of the longstanding concerns about the socioeconomic inequalities of health more broadly, as relevant to social policy and assessing social progress. The context of an infectious disease raises a further question as to whether such inequalities persist, or fade, as the disease spreads over time.

What evidence is there to help inform our understanding of these issues? Micro data on COVID-19 cases and/or deaths that include socio-economic characteristics at the unit-record level are rare.[4] Instead, this paper explores the empirical relationships across the 3,000 counties of the US. We merge recorded counts of cases and deaths at county level with socio-economic characteristics—average incomes, race, income inequality and poverty—and data on other covariates as suggested by the epidemiological literature. We use these data to try to better understand the spread of this infectious disease and behavioral responses to it.[5]

[2] For example, the word "poverty" does not appear in the classic epidemiological texts by Anderson and May (1991) and Gordis (2013). Nor does the word appear in the fully revised version of Gordis test, by Celentano and Szklo (2019). The need to incorporate human behavior into epidemiological models is discussed further in Ferguson (2007) and Fenichel et al. (2011). There has been greater awareness of socio-economic factors in the spread of infectious diseases with the recent emergence of the sub-field of social epidemiology (Honjo, 2004). Ellison (2020) shows how the standard Susceptible-Infectious-Recovered (SIR) model can incorporate socioeconomic factors.
[3] See, for example, Ravallion (2020) and references therein.
[4] See the discussion in Chen and Krieger (2020) with regard to this point in the US.
[5] We do not address the reverse effect—how COVID-19 might impact inequality going forward—although some of our results have bearing on this issue.

This is not the first paper to study the relationship between COVID-19 outbreaks and socioeconomic characteristics at the county level in the US; the antecedents we know of are Chen and Krieger (2020), Chin et al. (2020), Wu et al. (2020), Knittel and Ozaltun (2020) and McLaren (2020). These studies have been valuable, but a number of issues remain. The paper takes three main points of departure.

The first difference with past work is that we ground our empirical analysis on a theoretical model of behavioral responses to an infectious disease through endogenous social distancing. The model shows that, once one considers the potential income-constrained behavioral responses, even the directions of the effects of key covariates become uncertain. With regard to incomes, we distinguish two, potentially opposing, effects. The first is a "protection effect," whereby poverty curtails the ability to avoid infection through social distancing; for example, many low-paying jobs cannot be done from home. The second is an "adjustment-cost effect," recognizing that people cannot quickly and fully adjust to a lower level of socio-economic interactions when an epidemic hits. Furthermore, the marginal cost of adjustment may well depend on socio-economic characteristics such as age, race and income. The adjustment-cost effect could work in the opposite direction to the protection effect, given that a more affluent area may have more social and economic interactions, including externally to that area. Rational but partial adjustment in response to the epidemic will attenuate the preferred level of social distancing in more affluent areas. Another potential ambiguity is in the (much discussed) effect of an elderly population, which enhances vulnerability to serious illness, once infected, but may well have the opposite effect on the probability of infection, via social distancing. Thus, the overall effect on fatalities is unclear.

Second, to have a realistic hope of identifying the effect of incomes and other population characteristics, it is important to consider the multiple (correlated) covariates jointly, so as to better disentangle their individual effects. Socio-economic characteristics, such as poverty and race, are known to be correlated with each other as well as with epidemiological and biomedical factors, such as population density and health pre-conditions, thus confounding the claims one hears about the importance of those factors. Similarly, average incomes are likely to be negatively correlated with the incidence of poverty across areas, clouding inferences about whether the poverty effect is about average incomes or income inequality, with distinct implications for policy. Among the aforementioned studies, while Knittel and Ozaltun (2020) and Wu et al. (2020) look at multiple covariates jointly, they only study death rates. We consider multiple socio-economic variables as covariates for social distancing, the spread of the infection and for the severity of illness, conditional on the number of infections, as indicated by fatalities.

The third main difference to past work is that we propose an identification strategy in modelling fatalities conditional on (endogenous) infection rates. The dynamics of adjustments in behavior in response to the virus, and the lags in reporting, entail that deaths in any time period are unlikely to be directly proportional to infections in that period. In other words, the case fatality rate is likely to vary with the number of cases, which is endogenous. Our identification strategy is motivated by the idea of the epidemiological curve (often called the "epi curve"), whereby new infections increase over time up to some point and then decline as immunity builds up such that the number of susceptible people falls.[6] Following epidemiological theory, the epi curve is treated as a key factor determining observed counts of COVID-19 cases at county level, but is taken to only matter to deaths via cases. Under this exclusion restriction, we can identify causal effects on the counts of deaths conditional on infections.[7] We acknowledge and address a potential threat to this identification strategy associated with capacity constraints in the local health-care systems.

The following section provides some relevant background from the literature, while Section 3 outlines our theoretical model, which points to ambiguities in how levels of incomes and relative inequality impact social distancing and the spread of infection. Sections 4 and 5 describe our data and econometric methods respectively. The empirical models for social distancing, infections and deaths rely on the variables typically used in epidemiological models but augmented to include socio-economic characteristics relevant to inequality in various dimensions. On implementing the model empirically, Section 6 shows that, controlling for standard epidemiological covariates, US counties with higher median income tend to see more improvement in social distancing in response to the epidemic and a lower infection rate ceteris paribus. However, the latter effect is due to the (negative) correlation of median income with the poverty rate. Controlling for the median, a higher poverty rate—reflecting more unequal distributions of income from the perspective of the poor—is associated with a higher infection

---

[6] See, for example, the expositions on the epi curve in Anderson and May (1991) and Gordis (2013).
[7] While Knittel and Ozaltun (2020) study deaths per capita of the population, they do not identify the death rate among those infected, which one can expect to be determined differently to the infection rate.

rate. Income inequality also matters to the infection rate. Racial composition—interpretable as the fractionalization aspect of inequality[8]—also matters, and this covariate seriously confounds inferences about the role of income inequality and poverty, as well as epidemiological inferences, notably about the role of population density. The socio-economic covariates also impact the observed fatalities conditional on infections, but their effects are noticeably weaker. Section 7 concludes.

2. Foundations in the literature

A disease such as COVID-19 spreads when there is an effective contact between an infected individual and an uninfected susceptible person. The greater the number of additional people infected by each infected person, the faster the disease spreads. The progression of an outbreak over time is often modelled by the epi curve, and flattening the epi curve for COVID-19—mainly by reducing the contact rate—has become an important policy goal in 2020 across the world. The policy instruments have included various social distancing measures (recommendations on inter-personal contacts, the use of face masks, restrictions on large events/meetings, school closures, and shelter-in-place orders), among other recommendations such as frequent hand-washing. However, while the decisions taken are influenced by public health communications and controls, compliance and behaviors regarding social distancing are also personal choices. This is especially so in the US where some political leaders have resisted the stricter policies found elsewhere, such as lockdowns; both policies and compliance have varied across the country (Brzezinski et al. 2020).

Epidemiological models have long emphasized the role of local population density as a relevant factor driving the spread of a disease, as discussed by (inter alia) Anderson and May (1991) and Tarwater and Martin (2001).[9] Population density raises the contact rate by increasing the interaction between infected and uninfected individuals. (The epidemiological research on the role of population density has been influential in social distancing policy responses, such as bans on mass gatherings.) It can be expected that the marginal effect of higher density will

[8] Fractionalization refers to the population distribution across ethnic groups. For further discussion see Alesina et al. (2003).
[9] This also holds within confined spaces. See, for example, Lu et al.'s (2020) explanation of the spread of COVID-19 through a restaurant's air conditioning in China. Park et al. (2020) document a similar outbreak in a Korean call center.

decline as density rises, and reach zero at very high density, when it becomes physically hard to move. In Section 3 we will make the case for a new measure of density more appropriate to infectious diseases.

The influence of specific health conditions has also been emphasized with regard to the severity of the disease once infected. With reference to COVID-19, a number of pre-existing health conditions appear to exacerbate its effects, including cardiovascular disease, respiratory disease, and hypertension (see the review in Ssentongo 2020). Age also seems to be strongly correlated: once infected, older people have been found to be more likely to have severe symptoms leading to hospitalization, and in many cases, death (CDC COVID-19 Response Team, 2020; Verity et al. 2020; Ioannidis et al. 2020). At the time of writing, the role of age in determining the effects of COVID-19 is not fully understood, and it may well be that age reflects the higher incidence of the aforementioned comorbidities among older people.[10]

Stepping back, it is less clear how a high incidence of elderly people would impact infection rates, as this also depends on behavioral factors in specific contexts. While age may be considered a health-related factor with regard to the severity of illness once infected, it is also relevant as a behavioral covariate of the spread of infection. With higher retirement rates, the elderly will tend to face less economic pressure to be active outside the home, thus reducing their contact rates. Time-use surveys for the US indicate that elderly people have substantially lower contact rates in normal times (Cornwell 2011). We can think of this as a lower marginal cost of extra social distancing for the elderly during an epidemic. Against this, elderly people concentrated in residential care homes become more vulnerable, as seen in the US, as well as in other countries, during the new coronavirus pandemic. The key point is that, on a priori grounds, it is unclear whether elderly people will have a higher fatality rate from COVID-19, once one allows for the behavioral response through social distancing and (hence) infection rates.

A number of researchers have also pointed to poverty and race as covariates of COVID-19 incidence in the US (Chen and Krieger 2020; Chin et al., 2020; McLaren, 2020).[11] Chen and Krieger (2020) estimate COVID-19 death rates (per capita) that are almost twice as high for poverty rates over 20% as for those under 5%. The gradient in death rates is even steeper (a

[10] Early evidence has suggested that a weaker immune system among elderly adults may also contribute to higher mortality rates (Du et al. 2020, Zheng et al. 2020).
[11] On COVID-19 and racial inequalities in America also see the discussion in Yancy (2020).

factor of almost six) between the category with the highest percentage of the non-white population versus the lowest. More generally, large disparities in health outcomes along racial lines are well-documented; Black Americans have substantially lower life expectancy and higher infant mortality than other racial groups (National Center for Health Statistics 2016). Kirby and Kaneda (2010) note large racial and ethnic disparities in health insurance coverage that persist across the lifecycle. As is well known, poverty and race are correlated in America; for example, the official poverty rate in 2018 was 21% for Black Americans versus 12% overall (Semega et al. 2019).

While poverty has not been a prominent causative factor in traditional epidemiological models, the literature in the social sciences has pointed to many ways in which poverty might be expected to result in greater vulnerability to the new coronavirus. It is well documented that many of the risk factors associated with the severity of COVID-19 are correlated with income. For example, poverty has been found to increase the odds of having diabetes and heart disease (Gaskin et al. 2014; O'Connor and Wellenius 2012). Health outcomes in the US are strongly correlated with income and education, with poorer people generally experiencing worse health (Braverman et al. 2010).[12] In addition, poor (often underinsured) people may delay seeking medical help when it is costly or difficult to obtain (Jacob et al. 2015).

Poorer people are also likely to have a harder time isolating as a means of protecting themselves from infection. We can think of this as an effect of low income on the marginal cost of extra social distancing, which almost invariably comes at a cost (pecuniary or otherwise). Poor families may have little or no buffer of food stocks or savings to fall back on, and depend on short-term, often casual, labor, such that lockdown is a costly proposition. Whether people are able to shelter-in-place is likely to depend on their employment type, job security and savings (to offset loss of income from not working). Papageorge et al. (2020) find that income is strongly associated with self-protective behavioral responses, with poorer individuals much less able to practice social distancing, and much less able to tele-work.

[12] Health outcomes more generally seem to vary with income, with higher incomes being associated with better mental and physical health (Ettner 1996; Marmot 2002). Chetty et al. (2016) found that the average gap in life expectancy between the richest 1% and poorest 1% was 15 years for men and 10 years for women.

Nor do the assets and home environments of poorer people typically permit them to protect themselves well from the new coronavirus.[13] For the US, the Census Bureau's American Housing Survey (AHS) shows that poorer households have more persons per room, lower square footage per person, and are more likely to have inadequate plumbing and heating.[14] Poorer people are also more reliant on public transport, which increases their vulnerability to infection. Poorer areas also tend to face tighter fiscal and administrative constraints on policy effectiveness. Poverty may also affect death rates, including through its relationship with pre-existing health conditions or access to health services.[15]

These arguments suggest that we might expect to find a negative income effect on how quickly the virus spreads across communities. We can call this the "protection effect" since it mainly operates through the greater challenges facing poorer people in protecting themselves and their communities from the virus. This protection effect could matter to the number of cases, through a lower ability to social distance, as well as to the number of deaths, such as through a higher incidence of pre-existing health conditions and health-care access. The protection effect points to the possibility that poorer people face a sharper trade-off between their current economic welfare and their exposure to the virus, with implications for social protection policy.

This is clearly not the whole story. The potential for an offsetting income effect arises when places with higher average incomes have a higher customary density of personal interactions both in production and consumption, including links to external sources of infection through travel (for work and leisure) and by attracting visitors. Similarly, poverty is often associated with greater social isolation.[16] Richer people may well be better connected in both work and leisure activities. Interpersonal interactions help maintain and expand the networks that facilitate the creation and spending of wealth. For example, the 2018 Consumer Expenditure Survey shows that individuals in the top income quintile spend, on average, almost five times as much as people in the bottom quintile on food outside the home, and more than three times as much on entertainment.[17] Social and economic interactions in various forms (including with

[13] Here we focus on the US. There is evidence for other countries that poorer people have home environments that are less conducive to implementing prevailing recommendations for protection, including social distancing (Brown et al., 2020). Bargain and Aminjonov (2020) find that people in poorer regions responded less through their mobility for work and other activities.
[14] For data and further details, see the AHS Table Creator.
[15] Evidence for the US can be found in Ettner (1996). For a review of the evidence see Ravallion (2016, Chapter 7).
[16] Survey evidence on this point (for Canada) can be found in Stewart et al. (2009).
[17] For details, see the US Bureau of Labor Statistics site.

other locations where the virus may be present) could well be more pervasive in richer places. Against this view, one can also point to the fact that many relatively low-skilled jobs are done under working conditions that would create high contact rates (as in the case of the meat-packing industry). Local policies will also matter, such as the extent of enforced closures to restaurants, retail shops and events. Incomes and other socioeconomic covariates may matter via the local political economy.

These income effects are all present pre-COVID. They remain relevant to the extent that adjusting to a new, lower, level of such interactions in an epidemic is costly and hence partial. The adjustment may take time, creating interactions between socioeconomic covariates and the stage of the epidemic.

In the next section, we formalize these ideas in the form of a simple model of behavioral responses to an epidemic, which highlights the potential ambiguities, before we turn to the data.

3. A theoretical model of social distancing

While governments and public agencies can provide guidance and rules that alter behavior during an epidemic, the degree of compliance is in no small measure a matter of constrained personal choice. In the absence of medical treatment, three main things can be expected to matter to one's personal exposure to an infectious disease: (i) one's degree of social distancing (s, taken to be bounded above at s^{max}); (ii) one's income (y) as it determines one's expenditure on protective assets and goods, and (iii) one's personal characteristics, including location (x).[18] Define vulnerability to the virus as the probability of infection given s, y and x:[19]

$$p = p(s, y, x) \tag{1}$$

It is reasonable to assume that social distancing reduces the probability of infection but that it does so at a declining rate, i.e., that $p(.)$ is decreasing and convex in s ($p_s < 0, p_{ss} > 0$, where subscripts denote partial derivatives). For concreteness, one can think of the personal characteristic, x, as being above a certain age, say 65.

We do not assume that $p(s, y, x)$ is minimized. Rather, people trade-off the perceived benefit of social distancing against the cost, which includes costs in adjusting to the pandemic given pre-pandemic activities. In modelling this choice, we postulate that each person has a

[18] It is convenient to treat s and x as scalars, but one can readily generalize the analysis to treat them as vectors.
[19] For some purposes of our analysis we will treat p and s as random variables, with expected values.

customary level of interactions with others, which entail an initial, pre-epidemic, level of social distancing $\phi(y,x)$. In the pandemic, the person re-optimizes by choosing a new level of social distancing, above the prior level ϕ. How much higher this new level of social distancing is relative to ϕ determines the personal adjustment cost incurred in responding to the epidemic. However, as discussed in Section 2, the direction of the effect of either higher incomes or, say, being elderly on ϕ is ambiguous on a priori grounds.

The cost of social distancing is $c[s,\phi(y,x),y,x]$. We interpret $c[.]$ as the sum of a current adjustment cost and the expected future cost of being infected:

$$c[s,\phi(y,x),y,x] = a[s-\phi(y,x),y,x] + p(s,y,x)l(y,x) \quad (2)$$

Here $a(.)$ is the current adjustment cost and $l(y,x)$ is the (pecuniary and non-pecuniary) loss from infection (with $l(y,x)=0$ if one does not get infected). Adjustment costs entail that $a(.)$ is continuously increasing in $s-\phi$ with a rising adjustment cost as social distancing increases ($a_{ss}>0$). The nature and extent of these adjustment costs, and the losses from infection, will depend on many features of the markets and institutions in a given society. For example, if credit and risk markets work reasonably well then that would make adjustment easier. While we recognize that these contingent factors exist, we do not spell them out in detail here.

Even if there is no benefit from social and economic interactions, this problem will have an interior solution for social distancing. This will entail weighing the current adjustment cost against the expected future loss from infection, i.e., choosing s such that $a_s(.)+p_s(.)l(.)=0$. However, benefits from interactions are highly plausible, so we include this feature by letting the perceived personal benefit from social and economic interactions be $b(s,y,x)$, which is assumed to be a strictly and continuously decreasing function of s (which naturally reduces interactions).

Starting from zero, a small amount of social distancing will not presumably come with much loss, as one gives up the least important interactions. However, as one gets closer to full lockdown, at s^{max}, the loss will be large. So it is reasonable to assume that the net marginal benefit of social distancing ($b_s - c_s$) is strictly decreasing in the amount of social distancing ($b_{ss} - c_{ss} < 0$).

We further assume that the net marginal benefit is increasing in income $b_{sy} - c_{sy} > 0$. We do not consider it plausible that a higher income lowers the marginal benefit of social distancing ($b_{sy} < 0$), but we can allow that possibility as long as it is not outweighed by the tendency for poorer people to face higher marginal costs of social distancing ($c_{sy} < 0$), as discussed in Section 2.

We can now characterize the personally-preferred level of social distancing in response to the epidemic, which equates marginal benefit with marginal cost:[20]

$$b_s(s, y, x) = c_s[s, \phi(y, x), y, x] \tag{3}$$

Let the solution be $s(y, x)$. On implicitly differentiating (3) with respect to y, we have:

$$\frac{\partial s}{\partial y} = \frac{c_{sy} - b_{sy}}{b_{ss} - c_{ss}} + \frac{c_{s\phi}\phi_y}{b_{ss} - c_{ss}} \tag{4}$$

The first term on the right-hand side is the protection effect stemming from how a lower income implies a higher net marginal cost of extra social distancing. The protection effect is positive under our assumptions. The second term on the RHS of (4) is what we term the adjustment-cost effect, the sign of which cannot be determined based on the assumptions so far. On noting that $c_{s\phi} = -a_{ss} < 0$ (twice differentiating Equation 2), we see that if richer people interact more—a lower customary level of social distancing ($\phi_y < 0$) —then the adjustment-cost effect will work in the opposite direction to the protection effect, and we cannot say a priori whether poorer people will be more, or less, vulnerable to the virus. On the other hand, if $\phi_y > 0$ then the two effects work in the same direction; higher incomes yield greater social distancing.

Recall that income also appears directly in Equation (1) because of assets and goods that help protect from exposure to the virus. When social distancing is optimized (satisfying Equation 3) we can write the probability of infection as:

$$p = p[s(y, x), y, x] \equiv v(y, x) \tag{5}$$

It would be reasonable to assume that the function p is decreasing in y at given s. Then we see that if $\phi_y > 0$, poorer people will be more exposed to the virus when one takes account of both the direct effect and the behavioral response through social distancing, i.e., $v_y < 0$. Again, it should be emphasized that this is only one possibility consistent with our assumptions; if poorer people have higher customary levels of social distancing and the adjustment-cost effect is strong enough then we may find that it is the relatively well-off economically who are more exposed. Furthermore, if the adjustment-cost effect is sufficiently strong then there may be no income

[20] When there are governmental rules for social distancing these will add to the adjustment cost, and the solution of the following equation for s gives the optimal level of compliance with those rules.

effect on infection rates, or even a positive income effect, even though higher incomes facilitate better protection by other means at a given level of social distancing.

A positive income effect on social distancing creates a link between the extent of absolute income poverty in a society and its overall infection rate. Since poorer people are not as able to finance social distancing (notably because of its high marginal cost to them), the greater the number of poor people in any distribution the higher the likely infection rate.

We can go a step further and ask how greater relative inequality impacts the infection rate. Here the curvature of the relationship between social distancing and income becomes relevant. Suppose, for example, that the poor cannot afford to shelter-in-place, as they would not then have any income coming in, but the rich can readily do so without much loss, and that the infection rate depends on the aggregate number who shelter-in-place. Then a reduction in income inequality through mean-preserving transfers from the rich to the poor will increase social distancing and reduce the infection rate; the extra income for poor people will allow them to stay home more while the loss of income to the rich will have little effect on their behavior.

To provide a more general formulation of this argument, take the expected value of the functions $s(y,x)$ and $v(y,x)$ across those in group i (county i in our empirical case), giving:

$$S_i = E_i[s(y,x)] + v_i \tag{6}$$

$$C_i/N_i = E_i[v(y,x)] + \varepsilon_i \tag{7}$$

Here S_i is the observed mean level of social distancing in group i, with population size N_i, C_i is the number of cases observed, and the expectations are taken over all persons in group i, with zero-mean error terms v_i and ε_i. On applying Jensen's inequality, $E_i[s(y,x)] < (>)\ s(E(y),x)$ if the function s is increasing concave (convex) in y. For example, if the marginal increments to social distancing from extra income are highest for the poorest and tend to fall as income rises then higher inequality—interpretable as a transfer of income from the poor to the rich—will reduce average social distancing. Similarly, $E_i[v(y,x)] < (>)\ v(E(y),x)$ if v is decreasing concave (convex) in y.

To give a simple example with a closed-form solution, set $b(.)$ to a constant,[21] and allow adjustment costs of the form: $a(.) = \frac{1}{2}(s-\theta y)^2$, $(\theta > 0)$. Also let $p = (1-s)f(y)$ for some strictly decreasing function $f(y)$. (Both s and $f(y)$ can be taken to be scaled in the $(0,1)$

[21] Recall that adjustment costs and the loss from infection entail that an equilibrium exists even without any benefit from interaction.

interval to assure that that is also true of p.) Here $f(y)$ captures how a higher income helps protect from the infection, such as by allowing a home environment (with lower density and better facilities) that promotes compliance with social distancing. The expected loss from infection is taken to be a constant and set to unity. Optimal social distancing is $s = \theta y + f(y)$, which is increasing in y if $f'(y) > -\theta$, which we assume. Two cases illustrate the range of possibilities even in this simple example:

Case 1: Suppose that the marginal reductions in the probability of infection at given social distancing rise as income rises ($f"(y) < 0$). Then mean social distancing falls with higher inequality (s is concave in y). The infection rate falls with higher income, but we cannot say how it is affected by inequality (allowing s to vary optimally); a sufficient condition for inequality to increase the mean infection rate is that $f(y) > 1 - s$.[22]

Case 2: Instead, imaging that the marginal reductions in the probability of infection at given s fall as income rises ($f"(y) > 0$). Now mean social distancing rises with higher inequality. The infection probability still falls with higher income, but the implication for the infection rate of higher inequality is ambiguous. If $f(y) < 1 - s$ then the mean infection rate rises with inequality.

Nor is this the only way that inequality can matter—including inequality in other dimensions besides incomes. For example, it may harder to achieve the cooperation required to put in place collectively-beneficial pandemic response policies in societies that are more unequal in terms of both power and income.[23] Here we have only focused on the implications of income inequality for behavior with regard to social distancing.

What about the effect of a higher x on social distancing? Again, on implicitly differentiating (3), we have:

$$\frac{\partial s}{\partial x} = \frac{c_{sx} - b_{sx}}{b_{ss} - c_{ss}} + \frac{c_{s\phi}\phi_x}{b_{ss} - c_{ss}} \qquad (8)$$

The first term on the right-hand side of (8) is positive if being elderly lowers the net marginal cost of social distancing ($c_{sx} - b_{sx} < 0$). If, in addition, the elderly have a higher customary level of social distancing ($\phi_x > 0$) then we will expect the elderly to be social distancing more

[22] On letting the optimal s vary with y, it is readily verified that $\partial p/\partial y = (1-s)f'(y) - [\theta + f'(y)]f(y)$ and $\partial^2 p/\partial y^2 = [1 - s - f(y)]f"(y) - 2[\theta + f'(y)]\, f'(y)$.

[23] Similar arguments have been made about inequality and the provision of public goods. For a review of these arguments see Bardhan et al. (2000).

than others in response to the epidemic. On the other hand, the net effect of being elderly on social distancing is ambiguous if $\phi_x < 0$, given the adjustment cost.

One can think of this model as characterizing an equilibrium that emerges from a dynamic process of adjustment reflecting serial correlation in infection rates. Consider Equation (7). In equilibrium, the count of cases (C_i) is directly proportional to population (N_i). We can embed this in a dynamic adjustment model as:

$$lnC_{it} - lnE_i[v(y,x)] = \beta(lnC_{it-1} - lnE_i[v(y,x)]) + \gamma lnN_{it} + \varepsilon_i \ (1 > \beta > 0, \beta + \gamma = 1) \quad (9)$$

Notice that the short-run elasticity of cases to population, γ, is less than unity, even though the equilibrium level of cases is directly proportional to population. Given the dynamics, we should not expect a homogeneous relationship between counts of current infections and population size.

Even without lags, it can be conjectured that the initial socioeconomic effects evolve over time. (The vector x can include the time passed since the first infection locally.) One possibility is that they fade with the spread of infections, along with the mixing of different groups. Alternatively, the learning process and the aforementioned economic constraints on social distancing may mean that those who can afford to do so will protect themselves over time, including by mixing less. By the former view, socioeconomic inequalities have a diminished effect as the infection spreads, while by the latter, initial socioeconomic inequalities persistent and may even have magnified effects over time. We test for the presence of this aspect of the dynamics by introducing interaction effects with the time since the first recorded case.

We have seen that, once one allows for behavioral responses through social distancing, adjustment costs and dynamic effects, it is theoretically ambiguous whether the poor or elderly will be more vulnerable to infection, or whether more unequal income distributions will yield higher infection rates. Our empirical analysis will try to throw light on the matter.

4. Data and descriptive statistics

In measuring social distancing, we use data from Unacast's Social Distancing Scoreboard, which assigns a grade to each county, comparing daily mobility during the COVID-19 outbreak with a pre-COVID-19 baseline (beginning of March 2020), using mobile phone GPS data.[24] This is a composite index of improvement in social distancing across three dimensions:

[24] Further information can be found at Unacast's website.

the change in average distance traveled, the change in non-essential travel, and the decrease in human encounters. We convert Unacast's alphabetic grade to ordinal numeric variables, where a higher value indicates more social distancing relative to baseline values.

We also draw on Google's COVID-19 Community Mobility Report, which tracks changes in visits and time spent at various activities relative to a baseline (the median value for the corresponding day of the week between January 3 and February 6, 2020). The data are generated using location history for Google user accounts, and is reported for groceries and pharmacy, retail and recreation, transit stations, workplaces and residential locations.[25]

For data on confirmed COVID-19 cases and deaths, we draw on the Centers for Disease Control and Protection (CDC).[26] We use the most recent numbers available at the time of writing (June 18, 2020). Using the counts of both cases and deaths attributed to COVID-19 across 3,143 US counties, we find large modal point masses around their lower bounds with a long right tail, and with variances much larger than their means. The mean count of cases per county is 688 while the standard deviation is 3547; the median is 56 and the maximum is over 86,000 (in Cook County, IL).[27] The overdispersion is also evident for deaths. The median death count is 1, but the mean is 37 (and the standard deviation is 256). Kings County in New York (Brooklyn) recorded 7,000 deaths; the second highest is Queens, NY with 6,400 deaths.[28]

The data on cases and deaths are reported, and there are undoubtedly reporting errors, with some cases unreported or miss-diagnosed. As a robustness check, we also use data on excess deaths (defined as the difference between observed number of deaths and expected deaths (based on historical trends) and COVID-19 testing. Testing data are at the state level from the COVID-19 Tracking Project; excess deaths are also at the state level and are estimated by the CDC. We use excess deaths from all causes excluding COVID-19, which is a potential indicator

[25] More information can be found at Google's mobility website. The data are expected to be representative of Google users, but only contain information for users who opt in to tracking their location history and who have regular connectivity to the Google network.

[26] The data are available through USA Facts. An alternative source is the New York Times data site for COVID-19 (obtainable from the their Github repository). However, the NYT site records cases and deaths according to the county in which they occurred, while CDC does so according to the person's place of residence. The latter is more in keeping with our interest in the socioeconomic covariates, which are characteristics of places of residence. We also perform robustness checks using data from the NYT and find our results to be similar, with the main difference being that the NYT data suggest a lower elasticity of cases to days since first infection than for the CDC data.

[27] Figure A1 in the Appendix provides the kernel density functions for cases. Los Angeles has 78,000 cases, followed by Queens County, NY, with 64,000. The five New York City boroughs have over 200,000 cases.

[28] As with cases, New York City in total recorded more than 22,000 deaths. Outside of New York, Cook County recorded 4,300 deaths and Los Angeles County recorded 3,000.

of the extent of misclassified COVID-19 deaths and deaths indirectly related to the virus (for example, deaths resulting from an overburdened health system).[29] Of course, excess deaths do not stem solely from under-reporting of COVID-19 deaths, but also reflect fatalities due to other causes among those who do not seek treatment because of concern about catching the virus.

The Appendix (Table A2) lists cases, excess death estimates, and testing statistics by state. As expected, New York has the highest number of cases both in total (174,523) and per 100,000 (893). Nearby states, New Jersey, Rhode Island, Massachusetts and Connecticut, have substantially fewer cases per 100,000 (between 1,200 and 1,900), though still much higher than states in the South, Mid-West, and West (Washington, which had the earliest outbreak, has 380 cases per 100,000). Deaths follow a similar pattern, with relatively high deaths per 100,000 for states in the North-East of the country. Excess death counts suggest that significant undercounting may have occurred in some states with lower recorded deaths and cases, such as Pennsylvania, Michigan, Illinois, and Texas. The number of tests done within states also varies widely, with New York and California completing more than 3 million tests as of June 18, though their testing rates (per 1,000 residents) are much more in line with many other states.

We do not attempt to isolate the causal effect of social distancing on the infection rate, and we are skeptical about the prospects of finding a valid exclusion restriction for identifying the causal effect. Nonetheless, it is of interest to see what the bivariate relationship looks like between social distancing and infection rates: Figure 1 shows the counts of COVID-19 cases (panel a) plotted against the social distancing indicator and the corresponding graph of the proportion of counties with high infection rates (panel b), for various definitions of "high."[30] We see a strong negative relationship, which is consistent with the prevailing view (and the assumption of our model) that social distancing reduces infection rates for COVID-19. Going from the least improvement in social distancing to the most is associated with a reduction in the mean proportion of counties with over 400 cases per 100,000 from about 40% to 10%.

[29] More information on the calculation of excess deaths can be found on the CDC's website.
[30] In all cases, the linear regression has a slope significantly different from zero. For panel (a) the OLS slope coefficient is -0.60 (with a standard error of 0.01). For panel (b), the coefficients are (in %) -6.18 (0.35), -5.83 (0.40), -4.73 (0.39), -3.53 (0.37) for cut-offs of 100, 200, 300 and 400 respectively.

Table 1 provides summary statistics. Data on county population, population density, demographics and poverty rates are from the US Census Bureau.[31] Median income and the poverty rate are estimated from survey data but complemented by small-area estimation methods.[32] Overall, the poverty rate is 15%, ranging from 2.6% (Douglas County, CO) to 54% (Oglala Lakota County, SD). Gini indices are also estimated by small-area methods rather than data, and varies widely across counties, from the lowest value of 0.25 in (aptly named?) Loving County, Texas, to a high of 0.66 in East Carroll Parish, Louisiana.[33] The share of Black Americans refers to the proportion of the population that identifies as Black only, while share "Hispanic" also includes those who identify as Hispanic in addition to one or more other races.[34]

Health indicators come from the Centers for Disease Control and Prevention (CDC). The prevalence of diabetes and obesity is the crude rate among adults aged 18 years and older for each respective condition. For the incidence of asthma, high blood pressure, and COPD, we use state-level data from the U.S. Chronic Disease Indicators; all variables are the incidence among those 18 years and older. (Note that these state-level variables will therefore drop out of any estimation with state fixed effects). High blood pressure is measured for those with diabetes; that is, the proportion of diabetics with high blood pressure. On average, the prevalence of diabetes is 10%, asthma is 9%, and COPD is 7%. For those with diabetes, 72% have high blood pressure.

The Appendix provides robustness checks for an extended set of controls. Since public responses to COVID-19 may well be heavily influenced by state-level policies, we include the political party of the Governor in our robustness tests.[35] We also consider the role of weather, and in particular, the role of temperature, in both influencing social distancing (cooler weather suggests people may stay inside more often) and cases.[36]

[31] Specifically, population and demographics are obtained from the US Census Bureau. County population density is from this (public) GitHub page. Poverty estimates are from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program using the official national poverty line.
[32] These are model-based extrapolations. Further details can be found at the US Census Bureau site on small-area estimation methods.
[33] Loving County is also the second least populous county, with a 2017 population of only 169. The second lowest is Skagway Municipality in Alaska. NYC has a Gini index of 0.60, fifth highest. This range is similar to countries of the world, for which the range is from 0.24 in Slovenia to 0.63 in South Africa (as reported in the World Bank).
[34] The ACS does not include an Hispanic only category. The share of non-white is equal to the proportion of the population who does not identify as white only.
[35] Recent literature has suggested that places that have tended to vote Republican are less likely to adopt social distancing recommendations (Allcott et al. 2020).
[36] There is some evidence to suggest that the virus does not transmit as well in warm weather; for example Xu et al. (2020) and Sajadi et al. (2020).

5. Econometric models

We first discuss some key aspects of the specification of our models, before we turn to the empirical implementation.

Model specifications: We postulate a two-equation model for infections and deaths. The first equation is for the cumulative counts of cases ($\mathrm{C}_i \geq 0$ for county $i = 1, .. n$):

$$\mathrm{C}_i = exp(\alpha_0 + \alpha_1 ln\mathrm{T}_i + \alpha_2\mathbf{X}_i + \alpha_3 ln\mathrm{N}_i + \varepsilon_i) \ \ (i = 1, .., n) \qquad (10)$$

Here T_i is the number of days since the first case, the vector $\mathbf{X}_i$ comprises the covariates of COVID-19 infection, N_i is the population of county i, and the α's are parameters to be estimated. The days since the first infection enter as a log transformation to reflect the typical nonlinearity in the spread of infectious diseases, with a turning point emerging in new cases at some point (giving the epi curve).[37] Notice that we allow the population effect on the counts of cases to be non-homogeneous (of degree zero), to allow for lags in adjustment (as discussed in Section 3).

As noted in Section 4, the cases data are highly non-normally distributed, also exhibiting substantial overdispersion, such that the variance is substantially higher than the mean. Given these properties of the cases data we estimate the parameters of (10) as a negative binomial (NB) model using maximum likelihood.

The equation for recorded counts of deaths conditional on cases is as follows:

$$\mathrm{D}_i = exp(\beta_0 + \beta_1\mathbf{H}_i + \beta_2\mathbf{X}_i + \beta_3 ln\mathrm{C}_i + \mu_i) \ \ (i = 1, .., n) \qquad (11)$$

Where β's are parameters to be estimated and μ_i is another (zero mean) error term. Here the vector $\mathbf{H}_i$ includes the health risk factors—the pre-existing conditions that have been identified as relevant to the severity of the disease once infected. Here too we allow for possible non-homogeneity, in this case meaning that the count of deaths need not be directly proportional to infections. The death count exhibits a similar degree of non-normality and overdispersion to the counts of cases (or their logs), so we also estimate (11) as a NB.

Of course, $ln\mathrm{C}_i$ is endogenous in equation (11); for example, the latent characteristics of counties that tend to increase infection rates may make it more likely that the cases are severe or

[37] Note that the count of cases in our model is cumulative, so the function flattens out over time, with the decline in new cases. We also did a nested test comparing the log function with a quadratic in days; the coefficient on log days remained significant while those on the quadratic in days were not.

that health care is worse. We deal with this problem by adding a control function to (11), using the residuals from (10), and bootstrapping the standard errors. The days passed since the first case are assumed to alter cases (in keeping with the epi curve) but only matter to deaths via cases. The main threat to this identifying assumption appears to be the possibility that the strain on health facilities increases with days since the onset of the infection, leading to higher death rates. To help address this concern we include controls for county health-care capacity in the model of deaths.

We estimate regressions for social distancing (S_i for county $i = 1, .. n$), analogous to our model for cases. Here we postulate a latent continuous variable as a function of the same covariates as for cases:

$$\mathrm{S}_i = \gamma_0 + \gamma_1 ln\mathrm{T}_i + \gamma_2\mathbf{X}_i + \gamma_3 ln\mathrm{N}_i + \upsilon_i \ (i = 1, .., n) \tag{12}$$

From the Unacast data, we do not observe S_i but rather an ordinal signal—the index grades, which range from A+ to F. Given the nature of the data, we estimate this equation as an ordered logit model.

Explanatory variables and functional forms: While population density has long been seen as a key predictor of contacts and (hence) the spread of infections, we suggest that the more relevant variable is population squared per unit area.[38] To see why, note that in a county of population size N_i, the potential number of distinct contacts is $\mathrm{N}_i(\mathrm{N}_i - 1)/2 \cong \mathrm{N}_i^2/2$ for large N_i. (Note that this is a measure of potential interactions, not actual interactions, which is endogenous.) We call N_i^2 per unit land area the "potential interaction density" (PID) or "density" for short. Note that the log transformation entails that the regression coefficient on PID is the same as that on population density; what changes is the coefficient on (log) population. We postulate that the (per capita) contact rate at the county-level is an increasing function of density, but the slope of this function declines as density increases (as discussed in Section 2). We represent this by a log transformation of the county's PID (for which the elasticity is positive but less than unity). We also allow for the possibility that density matters to the death rate. As discussed in Sections 2 and 3, there is evidence to suggest that the share of the elderly can also be expected to influence infection rates, and in a potentially offsetting way to how this variable

[38] Note that here we are referring to the average density of a county. High local density (such as in one's residential building) is another matter.

impacts the severity of the illness once the virus takes hold; as such, the population share of elderly is also included in the vector $\mathbf{X}_i$.

Given that we are interested in estimating the protection and adjustment-costs effects on social distancing and infections, we additionally include the log of the median income of the county in $\mathbf{X}_i$. Since this is a key functional-form assumption for the interpretation of our results, we also did an encompassing (nested) test in which the median entered the NB model for cases in both log and linear form, the latter implying that cases are an exponential function of the median, i.e., that the infection rate is concave (convex) in the median whenever the infection rate is increasing (decreasing) in the median. With our full set of controls, the coefficient on the linear median had a very high standard error (z-score=0.63) while the coefficient on the log median had a z-score of 1.57 (significant at the 11% level). This suggests that the log transformation is to be preferred to a linear specification.[39]

Also motivated by our arguments in Sections 2 and 3, we allow the poverty rate of a county to matter to both the spread of the virus and its severity, after controlling for density and the share of the population 65 years and older. Note that the poverty rate is highly correlated with median household income (r=-0.89); the regression coefficient of the log poverty rate on log median is -1.45 (standard error=0.01). When we control for median income, the poverty rate behaves more like a measure of relative distribution; the partial correlation between the two variables (controlling for log median) is 0.43. So, the poverty rate and the Gini index can be treated as measures of aspects of inequality. They are not the same thing of course, and one cannot predict one of them very well from the other.[40] As a result, the Gini index may still have some extra explanatory power.

Given existing evidence on race and COVID-19 (Section 2), we include the population share of Black Americans in $\mathbf{X}_i$. Two possible function forms were considered, namely the log of the population share and the log of the fractionalization index.[41] In nested tests the former functional form clearly dominated. The population share of Black Americans may well be proxying for other unobserved factors; for example, Black Americans are more often in essential

[39] We found that an improvement in fit if one replaced the log median with a quadratic function of the (linear) median. The quadratic function was concave (positive coefficient on the median and negative on the median squared). The gain in fit was very small, however, so the more parsimonious log specification was preferred.
[40] Regressing the log of the Gini index on the log of the median and the log of the headcount index, the R^2 is 0.31.
[41] For two ethnic groups, the fractionalization index is $2s_1(1 - s_1)$ where s_1 is the population share of group 1.

services with greater exposure to the virus, such as health care, food preparation, and certain services.[42] Race is also highly correlated with our density variable (r=0.41), median income (-0.17), the share of the population 65 years and older (-0.30) and poverty (0.36). (Table A1 in the Appendix gives the correlation matrix). As a robustness check, in the Appendix we also consider the share of the Hispanic population in a county, along with the more general share of those who are non-white.

As noted in Section 2, the literature has also pointed to some comorbidities that influence vulnerability to the virus and so should be included in $\mathbf{H}_i$. We include measures of diabetes, asthma, hypertension and lung disease (specifically, the incidence of COPD). We also include a measure of health-care capacity; namely, the number of intensive care unit (ICU) beds per 1,000 residents and the number of hospitals per 100,000 residents. Here our expectation is that counties with greater health-care capacity will both detect more cases in their populations and attract cases for treatment from other counties.

6. Estimation results

In each case, we start with a specification of Equations (10)-(12) in which only some basic epidemiological and, in the case of deaths, health (capacity and comorbidity) variables appear as covariates. We then progressively add the socioeconomic characteristics up to the most complete model. To account for any unobservable factors at the state-level that may affect cases and deaths, such as health systems and policies, we also estimate a specification that includes state fixed effects.

Social distancing: We start first with the estimates from equation (12) for changes in social distancing in response to the epidemic in Table 2. Higher PID is associated with weaker improvements in social distancing, but larger populations generally do not matter independently of density. An exception is when we include state fixed effects, indicating that within states one tends to see less improvement in social distancing in the more populous counties. There is generally no effect of the time period since the first infection on the improvement in social

[42] Recent data from the Current Population Survey show that a majority of nursing and home health aides identify as Black/African American or Hispanic/Latino, as do a majority of those working in food preparation and most personal care and service occupations; see https://www.bls.gov/cps/cpsaat11.htm

distancing, although a weakly significant positive effect does emerge when we introduce state effects (Column 7).

We find a positive median-income effect on social distancing. This effect has an elasticity less than unity, implying that social distancing is an increasing and concave function of the median.[43] Thus, median-income inequality between counties reduces aggregate social distancing. However, the poverty rate has a positive effect, controlling for the median (and other covariates), suggesting that more unequal counties are seeing more social distancing at a given median income. We find no significant effect of the Gini index in any of our specifications.

We also see a strong positive effect of a higher share of elderly people on social distancing, consistent with the expectation discussed in Sections 2 and 3 that elderly people are better able to isolate at home—many of them may well be retired. We find a significant negative effect of a higher share of Black Americans on social distancing, although this is not robust to introducing state fixed effects.

Table A3 in the Appendix provides equivalent regressions using the Google COVID-19 Mobility data as outcome variables. We find similar results to those in Table 2; namely, that the socioeconomic variables related to income are all negatively associated with travelling to retail and recreation areas, transit station and workplaces, and positively associated with more time in residential spots. Interestingly, we find that while a higher share of the Black American population is correlated with less retail and recreation travel, it is positively associated with more time at workplaces, suggesting that the burden of essential work falls on Black Americans.

The Appendix (Table A4) also provides regressions with a fuller set of controls, including the share of Hispanic population, share of non-white population, temperature by month, COVID-19 tests per capita, and whether the state Governor is from the Republican Party. These do not change the main results.

Cases: The results for Equation (10) are found in Table 3.[44] We see strong positive effects of county density, population, and the days since first case. The infection rate is higher in

[43] Notice that the regression coefficient on the (log) median rises substantially when one adds the (log) poverty rate. However, when one calculates the total effect of the median in a regression with log median and log poverty rate allowing the latter to vary with the median consistently with the data one gets 0.54—very close to the coefficient on the median when the poverty rate is excluded. (The log poverty rate has an average elasticity with respect to the median of -1.45, with a standard error of 0.01.)

[44] The corresponding results using the cases data from the New York Times are reported in the Appendix, Table A5.

denser counties, as predicted by the standard epi-models (Section 2).[45] The per capita infection rate tends to be lower in more populous counties (noting that the elasticity with respect to population size is less than unity). Note that the coefficient on (log) population controlling for (log) population density is closer to, but still less than, unity. Cumulative infections rise with days since the first with an elasticity of a little below unity. A higher proportion of the county population 65 years and older is associated with fewer cases, suggesting that younger people are spreading the virus more.

We find a significantly negative effect of median income when we do not control for the poverty rate (Columns 2-4, Table 3). The negative elasticity implies that it is a (decreasing) convex function, such that the inequality in median incomes among counties is increasing the overall infection rate. Adding median income alone does not have much effect on the epidemiological covariates for density, population and the epi curve. We also find that counties with higher income inequality (as measured by the Gini index) have higher infection rates (Column 4).

However, the effect of higher median income switches sign when we include the county poverty rate, and the coefficient on the Gini index drops substantially (Column 5). It is clear that the negative median-income effect in Columns 2 and 3 is picking up the effect of poverty (recalling that the two variables are negatively correlated). As noted, when one controls for the median, the poverty rate will reflect income inequality, suggesting again that more unequal counties see higher infection rates. However, neither the positive effects of median income nor the poverty rate are significantly different from zero when we add state effects, indicating that these effects are largely driven by inter-state differences.

The appearance of a positive median-income effect on infections when we control for the poverty rate is consistent with the idea of an adjustment-cost effect whereby people in richer places have norms of interaction that cannot be costlessly adjusted downward during the epidemic. By this interpretation, we see two opposing effects, namely the protection effect (captured in our empirical analysis by the poverty rate) and an adjustment effect (picked up by median income once one controls for the poverty rate), although, on balance, the median-income effect is negative (as we see in Columns 2 and 3 of Table 3).

[45] Recall that this is the log of PID, given by (log) population squared per unit area; the coefficient for log population controlling for log population density is the sum of those on log PID and log population.

On the surface, this interpretation does not appear to fit well with our results in Table 2 on social distancing. There we see a greater social distancing response in counties with higher median income, also with an indication that this is also true in more unequal counties (as indicated by the significantly positive coefficient on the poverty rate). However, it should be recalled that the social distancing score is for the improvements relative to a baseline. Richer counties can improve their performance at social distancing in response to the epidemic, though the attained level of social distancing can still end up lower than elsewhere, given the adjustment costs. Thus, our results can be interpreted in a way that is consistent with the existence of both the adjustment-cost effect and the protection effect.

We find a strong positive effect of a higher population share of Black Americans, which is robust to including state fixed effects. It is also robust to dropping all other socioeconomic covariates; if simply add the share of Black Americans to the basic specification in Column (1) the elasticity is very similar (0.38, with a standard error of 0.02, as compared to 0.36 in Column (6)). A one standard deviation difference between counties in the log of the population share of Black Americans would yield on average about a 40% difference in the count of infections ($dln\mathrm{C} = 0.41$). The elasticity for the share of Hispanic people is only slightly lower at 0.30 when added to Column (6), with a standard error of 0.02 (Appendix, Table A4).

Once we control for the population share of Black Americans, the effects of the poverty rate and the Gini index drop substantially and are no longer significantly different from zero (comparing Columns (5) and (6) in Table 3).[46] The poverty and inequality variables are clearly picking up the difference in racial composition.

It is also notable from Table 3 how much the introduction of the share of Black Americans reduces the coefficient on the classic epidemiological factor, population density (comparing Columns 5 and 6). The fact that Black Americans tend to live in more dense counties and are also more exposed to the virus (presumably given their work) accounts for a large share of the presumed effect of density on the infection rate in models that ignore race.

The parameter estimates in Table 3 are averages. As the infection spreads, the effects may well change, to the extent that different socioeconomic groups mix. To test for this, Table 4 interacts the socioeconomic characteristics from Table 3 with the number of days since the first

[46] This is an effect of the Black American share, not the Hispanic share alone; if we only control for the latter then the inequality and poverty measures remain significant.

confirmed case of COVID-19.[47] We find little sign that the socioeconomic effects fade over time, the one exception being for the share of the population over 65. Comparing Columns (6) in Table 3 with Column (5) in Table 4, we see that the elasticity w.r.t. this variable starts off at about twice its average level, but then fades over time, though remaining negative though the entire range of the data. The effect of a higher share of Black Americans starts off negative but becomes positive within one week, and increases as more days pass since the first infection. The median and the poverty rate show a similar pattern.

Deaths: Table 5 gives the estimates of Equation (11), for COVID-19 deaths. (Table A7 in the Appendix gives the reduced-form estimates.) As expected, deaths rise with the number of cases. The elasticity is significantly greater than unity, implying that the ratio of fatalities to cases tends to rise with the number of cases. This is a strong (and statistically significant) pattern in our results. This could reflect a strain on local health care staff and infrastructure at high case loads. It could also reflect lags in reporting.[48] Comparing counties, and not allowing for state effects, the crude death rate (deaths per case) tends to be higher in denser counties. Death rates tend to rise with higher density though the elasticity is small, around 0.03.

Interestingly, pre-existing health conditions and health care capacity have rather weak effects. (Note that those variables in the $\mathbf{H}_i$ vector that are at state level drop out when we allow for state effects.) The health conditions in particular have effects that are close to zero, with only a higher proportion of asthmatics and those with COPD having a weakly significant effect on deaths though with a small elasticity of 0.05. Of course, this doesn't preclude an individual-level effect, as documented in the medical and epidemiological literature; here we are looking at averages. These comorbidity effects are weaker when we extend the range of socio-economic covariates.

Higher median incomes are associated with lower death rates, but this is not robust to including other socio-economic covariates. The share of elderly knocks out the income effect on fatalities. As expected, we tend to see higher death rates in counties with a higher population share over 65 years of age, with an elasticity around unity, and this is a robust feature of the data across different specifications for the regressors. The income-related variables play a more minor

47 Table A6 in the Appendix reports analogous results using the cases data from the New York Times.

48 For example, suppose that extra deaths are only recognized as being due to COVID-19 once a critical minimum number of cases is identified. Then the ratio of recorded deaths to cases rises with the number of cases.

role, and are insignificant and close to zero when we add the race and state fixed effects. A higher poverty rate is associated with a higher death rate, though the effect is not statistically significant when we include our full set of covariates and state effects.

Combining Tables 3 and 5, we see opposing effects of a higher share of the elderly. As expected, death rates are higher for the elderly; the elasticity of deaths (controlling for the other variables, including cases) to the population share of the elderly is close to unity. However, as we have seen, this reverses when we turn to the infection rate. On its own, the finding that the medical and epidemiological effects go in opposite directions is possibly not too surprising (as discussed in Section 2). On balance, with the full set of controls, we find a small but statistically insignificant effect of a higher proportion of the elderly on death rates in the reduced-form model (substituting Equation (10) into (11), as found in the Appendix, Table A5). So, the positive effect on deaths conditional on infections is counter-balanced by the effect on the infection rate.

Testing and excess deaths: Given that testing and excess deaths data are at the state-level, there is little we can say about heterogeneity in testing capacity within states. The Appendix (top row of Figure A2) provides scatter plots between testing rates and poverty or income across states, suggesting that wealthier states tend to have slightly higher testing rates, though the t-statistic for the fitted lines are statistically insignificant. When we add a control for testing at state level (Appendix Table A4) our main results turn out to be robust, while the incidence of testing comes in with a strong positive effect, as expected.

In the Appendix we also compare how the estimated parameters on the state fixed effects (Column 7, Table 3) vary with testing rates (bottom row of Figure A2 in the Appendix). We find no obvious relationship (the test statistic for the fitted line is statistically insignificant). Nor do we find that the state fixed effects in the regressions in Table 5 reflect excess deaths.

7. Conclusions

There are theoretical ambiguities in the influence of socioeconomic covariates on the spread of infections, given the induced behavioral responses, also allowing for costs of adjusting behavior to the threat of infections and the potential lags and nonlinearities. In the absence of effective pharmaceutical interventions, personal social-distancing choices are a plausible channel linking socioeconomic factors to the spread of infections. Our analysis of the bivariate relationship in the data is consistent with the view that social distancing lowers the COVID-19 infection rate in the US. However, the marginal cost of greater social distancing is likely to be higher for poorer families, who cannot easily maintain their consumption in isolation; this is what we dub the protection effect. Against this, the pre-epidemic levels of social and economic interactions are likely to be higher for wealthier people and there are costs of adjustment to a lower level during the epidemic—the adjustment effect. Similarly, there are a priori reasons why a more elderly population can yield lower infection rates but higher death rates conditional on infections, and our results support the view that both these opposing, and (it turns out) roughly equal, channels of impact are present in the US data.

We see signs of both the protection and adjustment effects in the relationship between COVID-19 outcomes across US counties and incomes. Without controlling for the incidence of poverty, a higher median income tends be associated with greater improvements in social distancing and lower infection rates. However, this is due to the fact that countries with a higher median income tend to have a lower poverty rate; controlling for poverty a higher median tends to come with higher infection rates and death rates, while a higher poverty rate does the same, reflecting the positive impact on infection rates of a less pro-poor distribution at a given median. The median-income effect on social distancing and infection rates is interpreted as indicative of the adjustment cost effect, while the positive effect is interpreted as reflecting the protection effect. The overall negative effect of higher median income, allowing the poverty rate to vary, suggests that the protection effect outweighs the adjustment-cost effect. The (independent) data on the social distancing response to the epidemic also support our argument that the protection effect dominates, with richer counties, and also more unequal counties, seeing stronger social distancing responses. Once one controls for the median, the poverty rate is reflecting relative inequality. Counties with higher overall income inequality tend to have higher infection rates, which is in part due to the fact that higher inequality comes with higher poverty rates. Similarly, higher inequality between counties increases the overall (national) infection rate.

We also find a strong effect of race, separately to poverty and inequality: a higher population share of Black Americans is associated with higher infection rates at county level. The effects of income inequality and poverty within counties largely vanish when one controls for the Black American population share, indicating that the directly relevant factor is race not income inequality or poverty per se. Also, without controlling for the racial composition of counties one substantially over-estimates the viral impact of higher population density.

Our interpretation is that poorer people are less able to protect themselves, which leads them to different choices—in essence, a steeper trade-off between their health and their economic welfare in the context of the threats posed by COVID-19. This points to a potential role for anti-poverty policy as a complement to health policy in combating this infectious disease. However, the infections are clearly spread through the distribution of income. Alongside the poverty effect, our results are consistent with the view that richer people tend to interact more (in both their income-earning and consumption choices). They reduce these interactions in the epidemic (relative to the pre-epidemic levels) but the costs of adjustment still leave richer counties with higher infection rates once one controls for the poverty rate and/or the share of Black Americans.

These socioeconomic effects on the spread of the virus do not fade over time since the first infection; rather, the effects tend to become even stronger. Thus, there is little to suggest that the mixture of different socioeconomic groups dulls the impacts of the underlying inequalities. What we see in the data is more consistent with a model of stronger socioeconomic segmentation as the virus spreads, probably reflecting a learning process in combination with the differences in economic constraints on social distancing. An exception to this pattern is found in how infection rates vary with an elderly population, which tends to matter less over time, probably reflecting younger families adopting greater social distancing (which comes more naturally for many of the elderly).

Controlling for the socio-economic characteristics that we have studied, we still find signs of the effects that have been more prominent in the epidemiological and medical literatures, though those effects become weaker. Population density remains a significant predictor of infection rates, but the coefficient is greatly attenuated once we control for socio-economic characteristics, especially racial composition. The partial correlations with the incidence of pre-existing health conditions are generally weak when one controls for socioeconomic covariates.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, Romain Wacziarg, 2003, "Fractionalization," *Journal of Economic Growth* 8: 155-194.

Allcott, Hunt, Levi Boxell, Jacob C. 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*, 26946.

Anderson Roy, and Robert May, 1991, *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press.

Anderson, Roy M, Hans Heesterbeek, Don Klinkenberg, and T Déirdre Hollingsworth, 2020, "How will country-based mitigation measures influence the course of the covid-19 epidemic?" *The Lancet* 395 (10228): 931–934.

Bardhan, Pranab, Samuel Bowles and Herbert Gintis, 2000, "Wealth Inequality, Wealth Constraints and Economic Performance," in A.B. Atkinson and F. Bourguignon (eds) *Handbook of Income Distribution Volume 1*, Amsterdam: North-Holland.

Bargain, Olivier, and Ulugbek Aminjonov, 2020, "Poverty and COVID-19 in Developing Countries," Bordeaux University.

Braveman, Paula A., Catherine Cubbin, Susan Egerter, David R. Williams, and Elsie Pamuk, 2010, "Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us," *American Journal of Public Health* 100(S1): S186-S196.

Brzezinski, Adam, Guido Deiana, Valentin Kecht, and David Van Dijcke, 2020, "The Covid-19 pandemic: government vs. community action across the United States". *Covid Economics: Vetted and Real-Time Pape*rs 7, 115–156.

Brown, Caitlin, Martin Ravallion and Dominique van de Walle, 2020, "Can the World's Poor Protect Themselves from the New Coronavirus?" NBER Working Paper 27200.

CBS News, 2020, "Pandemic's Front-line Work Falls on Women, Minorities," 1 May.

CDC COVID-19 Response Team, 2020, "Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19)," Centers for Disease Control and Prevention, United States, February 12–March 16.

Celentano, David, and Moyses Szklo, 2019, *Gordis Epidemiology Sixth Edition*. Elsevier.

Chen, Jarvis T., and Nancy Krieger, 2020, "Revealing the Unequal Burden of COVID-19 by Income, Race/Ethnicity, and Household Crowding: US County vs. ZIP Code Analyses," HCPDS Working Paper Volume 19, Number 1.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler, 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014," *Journal of the American Medical Association* 315(16): 1750-1766.

Chin, Taylor, Rebecca Kahn, Ruoran Li, Jarvis T. Chen, Nancy Krieger, Caroline O. Buckee, Satchit Balsari and Mathew V. Kiang, 2020, "U.S. County-Level Characteristics to Inform Equitable COVID-19 Response," medRxiv preprint.

Cornwell, Benjamin, 20100, "Age Trends in Daily Social Contact Patterns," *Research on Aging* 33(1): 598-631.

Du, Rong-Hui, Li-Rong Liang, Cheng-Quing Yang, Wen Wang, Tan-Ze Cao, Ming Li, Guang-Yun Guo, Juan Du, Chun-Lan Zheng, Ming Hu, Xu-Yan Li, Peng Peng, and Huan-Zhong Shi, 2020, "Predictors of Mortality for Patients with COVID-19 Pneumonia Caused by SARS-CoV-2: A Prospective Cohort Study," *European Respiratory Journal 55*(5).

Economist, The, 2020, "Many Poor Americans Can't Afford to Isolate Themselves," 24 April.

Ellison, George, 2020, "Implications of Heterogeneous SIR Models for Analysis of COVID-19," *NBER Working Paper*, 27373.

Ettner, Susan L., 1996, "New Evidence on the Relationship Between Income and Health," *Journal of Health Economics* 15(1): 67-85.

Fenichel, Eli P., Carlos Castillo-Chavez, M. G. Ceddia, Gerardo Chowell, Paula A. Gonzalez Parra, Graham Hickling, Garth Holloway, Richard Horan, Benjamin Morin, Charles Perrings, Michael Springborn, Leticia Velazquez, Cristina Villalobos, 2011, "Adaptive Human Behavior in Epidemiological Models," *Proceedings of the National Academy of Sciences* 108 (15): 6306-6311.

Ferguson, Neil, 2007, "Capturing Human Behavior," *Nature* 446: 733.

Gaskin, Darrell, Roland Thorpe Jr, Emma McGinty, Kelly Bower, Charles Rhode, J Hunter Young, Thomas LaVeist, and Lisa Dubay, 2014, "Disparities in Diabetes: The Nexus of Race, Poverty, and Place," *American Journal of Public Health* 104(11): 2147-2155.

Gordis, Leon, 2013, *Epidemiology* 5th Edition, Elsevier.

Hasell, Joe, Edouard Mathieu, Diana Beltekian, Bobbie Macdonald, Charlie Giattino, Esteban Ortiz-Ospina, Hannah Ritchie and Max Roser, 2020, "Coronavirus (COVID-19) Testing," Our World of Data, Oxford University.

Honjo Kaori, 2004, "Social Epidemiology: Definition, History, and Research Examples," *Environmental Health and Preventative Medicine* 9(5):193-199.

Ioannidis, John P.A., Cathrine Axfors, and Despina G. Contopoulos-Ioannidis, 2020, "Population-level COVID-19 Mortality Risk for Non-elderly Individuals Overall and for Non-elderly Individuals Without Underlying Diseases in Pandemic Epicenters," medRxiv preprint.

Jacob, Rebekah, Lauren D. Arnold, Jean Hunleth, K. Allen Greiner, and Aimee S. James, 2014, "Daily Hassles' Role in Health Seeking Behavior Among Low-income Populations," *American Journal of Health Behavior* 38(2): 297-306.

Kirby, James B., and Toshiko Kaneda, 2010, "Unhealthy and Uninsured: Exploring Racial Differences in Health and Health Insurance Coverage Using a Life Table Approach," *Demography* 47(4): 1035-1051.

Lu, Jianyun, Jieni Gu, Kuibiao Li, Conghui Xu, Wenzhe Su, Zhisheng Lai, Deqian Zhou, Chao Yu, Bin Xu, and Zhicong Yang, 2020, "COVID-19 Outbreak Associated with Air Conditioning in Restaurant, Guangzhou, China, 2020," *Emerging Infectious Diseases* 26(7), 1 April.

Marmot, Michael, 2002, "The Influence of Income on Health: Views of an Epidemiologist," *Health Affairs* 21(2), March/April.

McLaren, John, 2020, "Racial Disparity in Covid-19 Deaths: Seeking Economic Roots with Census Data," NBER Working Paper 27407

National Center for Health Statistics, 2016, *Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities.* Hyattsville, Maryland.

Nordling, Linda, 2020, "Scientific Networks are Helping African Countries to Access Coronavirus Lab Supplies," *Nature* 26 May.

O'Connor, Alane and Wellenius, Gregory, 2012. "Rural–Urban Disparities in the Prevalence of Diabetes and Coronary Heart Disease," *Public Health* 126(10): 813-820.

Papageorge, Nicholas, Matthew Zahn, Michèle Belot, Eline van den Broek-Altenburg, Syngjoo Choi, Julien Jamison, and Egon Tripodi, 2020, "Socio-Demographic Factors Associated with Self-Protecting Behavior during the Covid-19 Pandemic," *NBER Working Paper,* 27378.

Park Shin Young, Young-Man Kim, Seonju Yi, Sangeun Lee, Baeg-Ju Na, Chang Bo Kim, Jung-il Kim, Hea Sook Kim, Young Bok Kim, Yoojin Park, In Sil Huh, Hye Kyung Kim, Hyung Jun Yoon, Hanaram Jang, Kyungnam Kim, Yeonhwa Chang, Inhye Kim, Hyeyoung Lee, Jin Gwack, Seong Sun Kim, Miyoung Kim, Sanghui Kweon, Young June Choe, Ok Park, Young Joon Park, and Eun Kyeong Jeon**,** 2020, "Coronavirus Disease Outbreak in Call Center, South Korea," *Emerging Infectious Diseases* 26(8).

Ravallion, Martin, 2016, *The Economics of Poverty: History, Measurement, and Policy*. Oxford University Press.

Ravallion, Martin, 2020, "Pandemic Policies in Poor Places," Center for Global Development, Washington DC.

Sajadi, Mohammad, Parham Habibzadeh and Augustin Vintzileos, 2020, "Temperature, Humidity, and Latitude Analysis to Estimate Potential Spread and Seasonality of Coronavirus Disease 2019 (COVID-19)," *JAMA Network Open* June 11.

Ssentongo, Paddy, Anna E. Ssentongo, Emily S. Heilbrunn, Djibril M. Ba, Vernon M. Chinchilli, 2020, "The Association of Cardiovascular Disease and Other Pre-existing Comorbidities with COVID-19 Mortality: A Systematic Review and Meta-analysis," medRxiv 2020.05.10.20097253.

Semega, Jessica, Melissa Kollar, John Creamer and Abinash Mohanty, 2019, "Income and Poverty in the United States: 2018," US Census Bureau.

Stewart, Miriam J., Edward Makwarimba and Linda I. Reutter, 2009, "Poverty, Sense of Belonging and Experiences of Social Isolation," *Journal of Poverty* 13:173–195.

Tarwater, Patrick M., and Clyde F. Martin, 2001, "Effects of Population Density on the Spread of Disease," *Complexity* 6(6): 29-36.

Valentino-DeVries, Jennifer, Denise Lu and Gabriel J.X. Dance, 2020, "Location Data Says It All: Staying at Home During Coronavirus Is a Luxury," *New York Times* 3 April.

Verity, Robert, Lucy Okel, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu, Amy Dighe, Jamie T Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes

Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson, 2020, "Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis," *Lancet*, published online March 31, 2020.

Wu, Xiao., Rachel Nethery, Benjamin Sabath, Danielle Braun and Francesca Dominici, 2020, "Exposure to Air Pollution and COVID-19 Mortality in the United States," medRxiv, 2020.04.05.20054502.

Xu, Ran, Hazhir Rahmandad, Marichi Gupta, Catherine DiGennaro, Navid Ghaffarzadegan, Heresh Amini, and Mohammad Jalali, 2020, "The Modest Impact of Weather and Air Pollution on COVID-19 Transmission," medRxiv, 2020.05.05.20092627.

Yancy, Clyde W., 2020, "COVID-19 and African Americans," *Journal of the American Medical Association* Viewpoint, April 15.

Zheng, Meijuan, Yong Gao, Gang Wang, Guobin Song, Siyu Liu, Dandan Sun, Yuanhong Xu, and Zhigang Tian, 2020, "Functional Exhaustion of Antiviral Lymphocytes in COVID-19 Patients," *Cellular & Molecular Immunology* 17(5): 533–535.

Table 1: Summary statistics for key variables

	N	Mean	Std Dev.	Min	Max
Cases	3,143	688.41	3547.20	0.00	86179.00
Cases, log	3,143	4.12	2.15	0.00	11.36
Deaths	3,143	37.19	256.17	0.00	6965.00
Deaths, log	3,143	1.34	1.62	0.00	8.85
Days	3,143	72.53	22.94	0.00	143.00
Days, log	3,143	4.10	0.97	0.00	4.97
Population	3,142	104127.10	333486.30	88.00	10100000.00
Population, log	3,142	10.27	1.49	4.48	16.13
Population density	3,143	670.97	4465.49	0.11	179922.30
Potential interaction density, log	3,142	15.00	3.14	3.67	26.40
Population share 65+	3,142	19.27	4.71	4.83	57.59
Population share 65+, log	3,142	2.93	0.25	1.57	4.05
Median income	3,141	52794.41	13880.12	25385.00	140382.00
Median income, log	3,141	10.84	0.24	10.14	11.85
Poverty rate	3,141	15.16	6.13	2.60	54.00
Poverty rate, log	3,141	2.64	0.40	0.96	3.99
Gini index	3,128	44.55	3.65	25.67	66.47
Gini index, log	3,128	-0.81	0.08	-1.36	-0.41
Share of Black Americans	3,142	9.34	14.47	0.00	86.07
Share of Black Americans, log	3,142	1.61	1.14	0.00	4.47
Hospitals per 100,000	3,142	0.61	0.94	0.00	10.56
Hospitals per 100,000 (IHS)	3,142	0.47	0.50	0.00	3.05
ICU beds per 1,000	3,142	0.13	0.54	0.00	27.45
ICU beds per 1,000 (HIS)	3,142	0.12	0.19	0.00	4.01
Percent with diabetes	3,142	10.38	3.80	1.50	33.00
Percent with asthma	3,141	9.16	1.20	7.30	13.20
Percent diabetics with high BP	3,141	72.08	3.97	65.00	79.50
Percent with COPD	3,141	7.13	2.20	3.60	15.00
Share Hispanic	3,142	9.65	13.84	0.61	96.36
Share non-white	3,142	15.53	16.38	0.96	96.16
Republican	3,143	0.73	0.44	0.00	1.00
Temperature March	3,135	48.71	11.62	-4.20	76.40
Temperature April	3,135	52.22	9.35	12.00	80.20
Temperature May	3,135	61.64	7.47	28.70	83.00

Notes: Data at county-level. COVID-19 cases and deaths come from the CDC; total counts as of 18th June 2020. Days are the number of days since the first case confirmed. Log cases, deaths, and days are equal to the variable + 1 logged. Demographic variables are drawn primarily from the US Census and the CDC. Share of Black American refers to the share of the county population that identify as Black or African American only. Share of Hispanic refers to the share of the population that identify as Hispanic or Hispanic and another race. Share of non-white is 1 minus the share of the population that identifies as white only. Potential interaction density is equal to population density (persons per square kilometer) multiplied by population. The poverty rate for the US is based on the US poverty line. Number of hospitals and ICU beds are from Johns Hopkins University. Prevalence of diabetes is at county level; asthma, high blood pressure, and COPD is at state level. Temperature is monthly average from the NOOA.

Table 2: Regressions for social distancing in response to the COVID-19 epidemic

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	-1.09***	-1.09***	-1.12***	-1.12***	-1.10***	-1.02***	-1.10***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)
Population	-0.04	-0.08	0.02	0.01	-0.08	-0.23**	-0.80***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.16)
Days since first case	0.02	0.02	0.05	0.04	0.03	0.06	0.12*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Median household income		0.45**	0.56***	0.70***	3.50***	3.95***	2.65***
		(0.18)	(0.19)	(0.22)	(0.50)	(0.51)	(0.57)
Share 65 and older			1.15***	1.15***	1.69***	1.57***	0.53**
			(0.18)	(0.19)	(0.21)	(0.21)	(0.24)
Gini index				0.69	-0.56	-0.07	0.46
				(0.58)	(0.62)	(0.62)	(0.69)
Poverty rate					1.81***	2.34***	2.25***
					(0.29)	(0.31)	(0.34)
Share of Black Americans						-0.24***	0.03
						(0.04)	(0.06)
State fixed effects	No	No	No	No	No	No	Yes
N	3054	3054	3054	3042	3042	3042	3042

Note: Data are at the county level. The dependent variable is the social distancing grade for the county from Unacast's Social Distancing Scoreboard. Grades are as of May 28 2020. Ordinal logistic regression used for estimation. All covariates are logged. Variable descriptions can be found in Table 1 notes. Standard errors in parentheses. * prob<0.10 ** prob<0.05 *** prob<0.01. N=3054.

Table 3: Regressions for reported COVID-19 cases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.19***	0.20***	0.20***	0.19***	0.19***	0.03	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Population	0.64***	0.68***	0.62***	0.61***	0.58***	0.82***	0.87***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Days since first case	0.95***	0.92***	0.88***	0.88***	0.88***	0.81***	0.73***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Median household income		-0.57***	-0.67***	-0.37***	0.77***	0.45**	0.17
		(0.08)	(0.08)	(0.09)	(0.21)	(0.20)	(0.21)
Share 65 and older			-1.52***	-1.59***	-1.37***	-1.22***	-1.35***
			(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Gini index				1.64***	0.95***	0.20	0.14
				(0.27)	(0.29)	(0.28)	(0.27)
Poverty rate					0.76***	0.12	0.19
					(0.13)	(0.13)	(0.13)
Share of Black Americans						0.36***	0.40***
						(0.02)	(0.02)
Constant	-8.84***	-3.13***	3.16***	1.80*	-13.51***	-9.81***	-6.33**
	(0.22)	(0.84)	(0.89)	(0.94)	(2.71)	(2.61)	(2.70)
Shape parameter	0.08***	0.07***	-0.02	-0.03	-0.04	-0.12***	-0.32***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
State fixed effects	No	No	No	No	No	No	Yes
N	3142	3141	3141	3128	3128	3128	3128

Note: Data for US counties. Negative binomial used for estimation. All covariates are logged. The dependent variable is cases. Variable descriptions can be found in Table 1 notes. Standard errors in parentheses. * prob<0.10 ** prob<0.05 *** prob<0.01.

Table 4: Incorporating interaction effects with time since the first infection

	(1)	(2)	(3)	(4)	(5)	(6)
Density	0.20***	0.20***	0.19***	0.19***	0.03	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Population	0.69***	0.62***	0.60***	0.58***	0.81***	0.87***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Days since first case	1.65	0.49	0.65	-11.39**	-11.77**	-9.27*
	(2.18)	(2.22)	(2.31)	(5.50)	(5.32)	(5.06)
Median household income	-0.27	-0.78	-0.81	-3.39*	-3.66**	-3.04*
	(0.88)	(0.86)	(0.96)	(1.87)	(1.80)	(1.72)
Median income # days	-0.07	0.03	0.10	0.96**	0.94**	0.74*
	(0.20)	(0.20)	(0.22)	(0.43)	(0.41)	(0.39)
Share 65 and older		-1.65***	-1.55***	-2.38***	-2.62***	-2.79***
		(0.59)	(0.59)	(0.73)	(0.73)	(0.69)
Share 65 and older # days		0.03	-0.01	0.23	0.32*	0.33**
		(0.14)	(0.14)	(0.17)	(0.17)	(0.16)
Gini index			-2.70	-0.99	0.54	0.13
			(1.90)	(1.97)	(1.90)	(1.79)
Gini index # days			1.01**	0.45	-0.10	-0.01
			(0.44)	(0.46)	(0.44)	(0.42)
Poverty rate				-1.89*	-2.05**	-1.46
				(1.02)	(0.99)	(0.96)
Poverty rate # days				0.61***	0.50**	0.38*
				(0.24)	(0.23)	(0.22)
Share of Black Americans					-0.32	0.02
					(0.27)	(0.24)
Share of Black Americans # days					0.16**	0.09
					(0.06)	(0.06)
Constant	-6.33	4.83	2.91	39.83*	45.25*	37.25*
	(9.55)	(9.72)	(10.14)	(23.99)	(23.22)	(22.03)
Shape parameter	0.07***	-0.02	-0.03	-0.04*	-0.13***	-0.33***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
State fixed effects	No	No	No	No	No	Yes
N	3141	3141	3141	3128	3128	3128

Note: Data for US counties. Negative binomial used for estimation. All covariates are logged. The dependent variable is cases. Variable descriptions can be found in Table 1 notes. The symbol # represents an interaction term. Standard errors in parentheses. * prob<0.10 ** prob<0.05 *** prob<0.01.

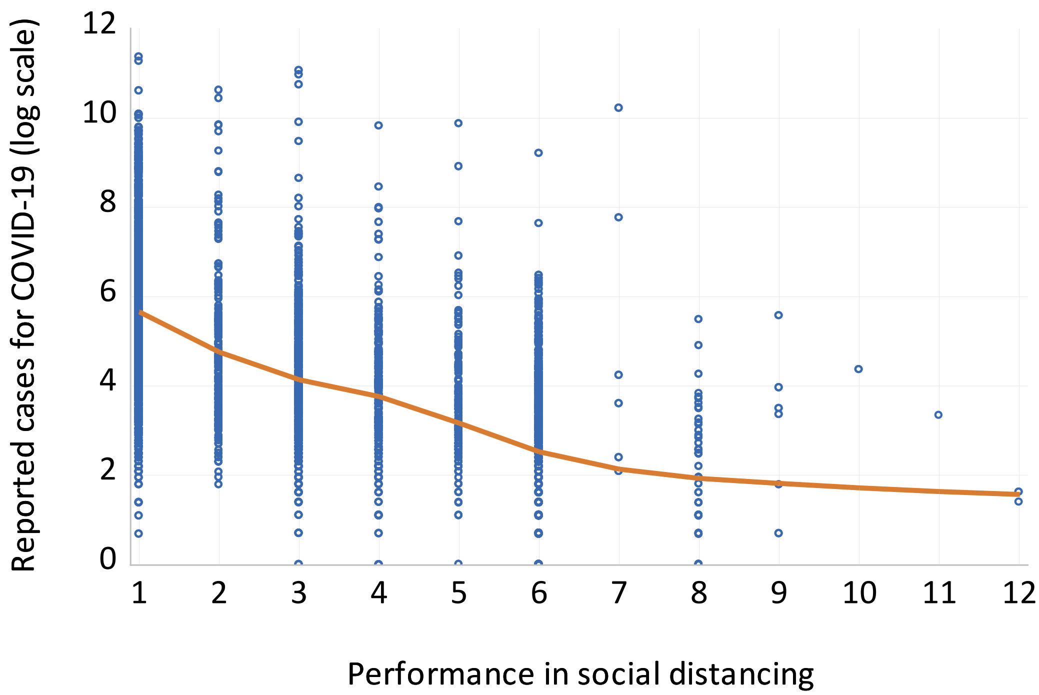
Table 5: Regressions for COVID-19 deaths conditional on reported infections

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cases	1.12***	1.11***	1.15***	1.15***	1.15***	1.13***	1.09***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Density	0.01	0.02	0.04**	0.03**	0.03*	0.03**	0.02
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Hospitals per 100,000	-0.04	-0.05	-0.03	-0.04	-0.04	-0.03	-0.06
	(0.08)	(0.07)	(0.08)	(0.07)	(0.09)	(0.10)	(0.07)
ICU beds per 1,000	-0.06	-0.13	-0.14	-0.15	-0.14	-0.15	-0.08
	(0.13)	(0.14)	(0.10)	(0.13)	(0.11)	(0.11)	(0.10)
Proportion with diabetes	0.01	0.00	0.01	0.01	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Proportion diabetic with high BP	-0.00	-0.00	-0.01	-0.01	-0.01	-0.01**	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Proportion with asthma	0.06***	0.07***	0.03	0.03	0.03*	0.05**	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Proportion with COPD	0.02	0.02	0.03**	0.03*	0.03**	0.04**	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Median household income		-0.22*	-0.04	0.04	0.27	0.20	0.23
		(0.12)	(0.12)	(0.14)	(0.26)	(0.26)	(0.20)
Share 65 and older			1.05***	1.04***	1.08***	1.09***	1.06***
			(0.10)	(0.10)	(0.11)	(0.13)	(0.12)
Gini index				0.42	0.25	0.10	0.39
				(0.35)	(0.42)	(0.37)	(0.36)
Poverty rate					0.16	0.08	0.07
					(0.16)	(0.16)	(0.15)
Share of Black Americans						0.06**	0.05*
						(0.03)	(0.03)
Residual	-0.00**	-0.00	-0.00***	-0.00***	-0.00**	-0.00***	-0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-4.70***	-2.30	-7.19***	-7.63***	-10.83***	-9.78***	-10.05***
	(0.72)	(1.44)	(1.47)	(1.75)	(3.54)	(3.51)	(2.71)
Shape parameter	-0.39***	-0.40***	-0.47***	-0.47***	-0.47***	-0.48***	-0.80***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
State fixed effects	No	No	No	No	No	No	Yes
N	3141	3141	3141	3128	3128	3128	3128

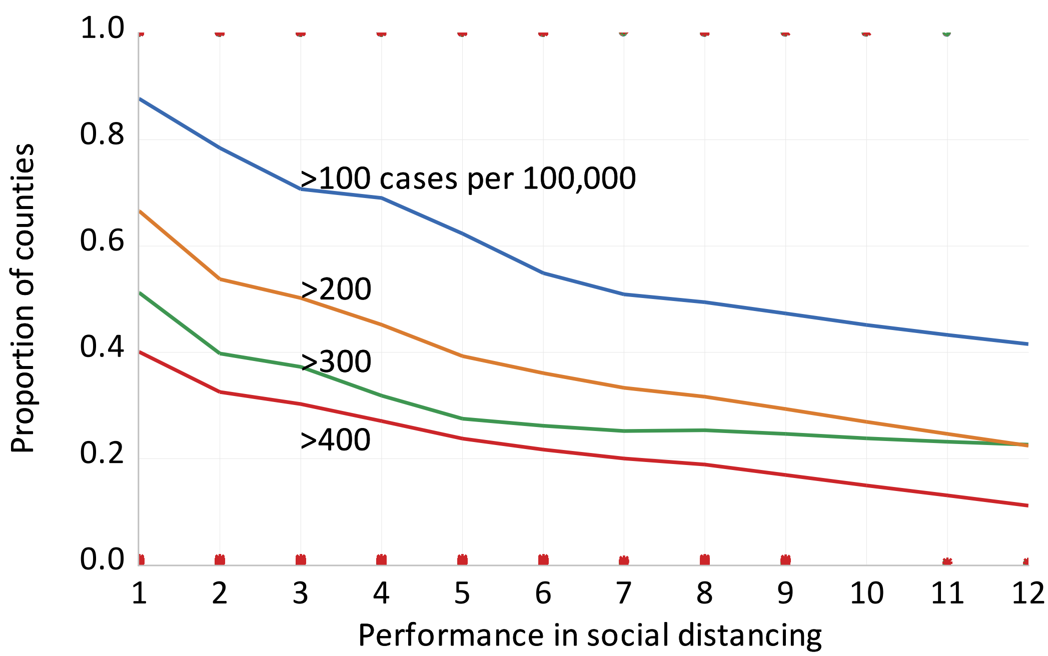
Note: All covariates excluding the health variables are logged. Hospitals and ICU beds are transformed using an inverse hyperbolic sine function. A negative binomial (NB) model is used for estimation. Variable descriptions can be found in Table 1 notes. Bootstrapped standard errors in parentheses. * prob<0.10 ** prob<0.05 *** prob<0.01.

Figure 1: COVID-19 cases across counties plotted against performance in social distancing

(a) Cases



(b) High incidence



Note: The figures provide nonparametric regression functions, giving the conditional mean at each point, based on a locally smothered scatter plot. Each point on the x-axis corresponds to Unacast's social distancing grade, ranging from F (x = 1) to A (x = 12).

APPENDIX

Table A1: Correlation coefficients between socio-economic variables

	Density	Median income	Share 65 and older	Poverty rate
Median income	0.376***			
Share 65 and older	-0.416***	-0.276***		
Poverty rate	-0.176***	-0.888***	0.027	
Share Black American	0.414***	-0.167***	-0.296***	0.356***

Note: Data are at the county level. The table shows the correlation coefficients and p-values for the socio-economic variables included in our model. Variable descriptions can be found in Table 1 notes. * prob<0.10 ** prob<0.05 *** prob<0.01.

Table A2: COVID-19 cases, deaths and testing by US States

	Total cases	Cases per 100,000	Total deaths	Deaths per 100,000	Excess deaths	Total tests	Tests per 100,000
Alabama	28206	577.06	810	16.57	389	322534	6598.66
Alaska	793	107.53	10	1.36	56	81185	11009.06
Arizona	43731	609.78	1287	17.95	1016	379732	5294.91
Arkansas	13493	447.70	208	6.90	241	228434	7579.54
California	167135	422.52	5359	13.55	2487	3074530	7772.40
Colorado	29886	524.72	1638	28.76	626	262216	4603.86
Connecticut	45170	1264.32	4226	118.29	397	372585	10428.77
Delaware	10485	1084.09	431	44.56	158	88684	9169.42
District of Columbia	9903	1409.77	527	75.02	285	72199	10278.10
Florida	85808	402.87	3060	14.37	1191	1512769	7102.43
Georgia	56893	540.84	2564	24.37	996	663204	6304.54
Hawaii	750	52.80	17	1.20	54	64374	4532.09
Idaho	3744	213.43	90	5.13	127	68012	3877.08
Illinois	135220	1061.29	6556	51.46	2868	1284693	10083.08
Indiana	42480	634.80	2491	37.22	614	384722	5749.09
Iowa	24951	790.55	680	21.55	141	240931	7633.71
Kansas	11886	408.24	252	8.66	124	142124	4881.46
Kentucky	13391	299.68	540	12.08	252	306380	6856.59
Louisiana	48176	1033.83	2950	63.31	942	545221	11700.08
Maine	2876	214.88	102	7.62	114	74060	5533.46
Maryland	63839	1056.46	2990	49.48	1298	447608	7407.40
Massachusetts	106114	1537.41	7759	112.41	1144	741260	10739.55
Michigan	66825	668.52	6062	60.64	2920	850186	8505.33
Minnesota	31617	563.46	1343	23.93	247	460879	8213.59
Mississippi	20641	691.14	938	31.41	758	238715	7993.06

Missouri	15657	255.56	933	15.23	523	297342	4853.41
Montana	655	61.66	20	1.88	130	66870	6294.80
ND	3193	420.09	75	9.87	97	90654	11926.95
Nebraska	17376	900.65	244	12.65	82	144813	7506.11
Nevada	12164	400.87	475	15.65	114	230796	7606.01
New Hampshire	5449	401.71	331	24.40	222	101984	7518.41
New Jersey	167424	1879.37	12800	143.68	4289	1171734	13152.96
New Mexico	10153	484.53	456	21.76	137	275897	13166.62
New York	174523	893.06	9094	46.54	3034	3179660	16270.73
North Carolina	48426	466.37	1209	11.64	269	693678	6680.50
Ohio	43122	368.90	2633	22.52	860	600024	5133.04
Oklahoma	9354	237.23	366	9.28	168	269553	6836.10
Oregon	6373	152.07	187	4.46	309	189136	4513.22
Pennsylvania	84780	661.98	6417	50.11	5092	624068	4872.84
Rhode Island	14524	1373.67	797	75.38	126	211593	20012.29
South Carolina	21548	423.83	621	12.21	781	280523	5517.62
South Dakota	6109	692.45	78	8.84	72	70353	7974.41
Tennessee	32604	481.59	504	7.44	413	652394	9636.53
Texas	102677	357.74	2137	7.45	2735	1407741	4904.71
Utah	16015	506.63	143	4.52	201	282685	8942.60
Vermont	1130	180.43	56	8.94	301	54745	8741.03
Virginia	56180	659.57	1585	18.61	903	502327	5897.46
Washington	28663	380.37	1246	16.53	78	470043	6237.64
West Virginia	2418	133.90	88	4.87	253	144429	7997.92
Wisconsin	24043	413.57	719	12.37	530	457963	7877.49
Wyoming	1144	198.01	18	3.12	111	36154	6257.87
Total	70677.92	739.44	3272.07	43.50	1184.08	1071765.00	8788.26

Note: Data on cases and deaths come from the CDC. Cases and deaths are as of 18th of June 2020 and are at the county level. Excess deaths are at the state level and exclude Covid-19 cases. Estimates are from the CDC and are the upper bound of the 95% confidence interval from 23rd of May 2020. Total tests are from the COVID-19 Tracking Project and are at the state level. Testing data is up to 18th of June 2020.

Table A3: Social distancing regression results using Google Mobility data

	Retail, grocery and recreation		Transit stations		Workplaces		Residential	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Density	-0.94***	-1.36***	-3.88***	-5.86***	-1.54***	-1.59***	0.80***	0.81***
	(0.34)	(0.43)	(0.76)	(1.01)	(0.11)	(0.15)	(0.07)	(0.08)
Population	0.34	1.98**	1.04	5.97***	0.92***	1.29***	-0.49***	-0.63***
	(0.70)	(0.86)	(1.51)	(1.99)	(0.23)	(0.30)	(0.15)	(0.17)
Days since first case	0.78	0.08	10.60***	10.49***	-0.01	0.15	-0.83	0.26
	(2.52)	(2.39)	(2.02)	(1.97)	(0.21)	(0.20)	(0.73)	(0.65)
Median h'hold income	-30.24***	-18.31***	-56.48***	-41.15***	-21.85***	-19.17***	12.81***	10.80***
	(3.33)	(3.48)	(7.02)	(8.03)	(1.14)	(1.21)	(0.66)	(0.65)
Share 65 and older	9.16***	16.61***	4.73	15.02***	6.16***	7.95***	-2.31***	-4.02***
	(1.46)	(1.57)	(3.29)	(3.92)	(0.49)	(0.53)	(0.29)	(0.30)
Gini index	-35.59***	-39.97***	-59.60***	-57.49***	-11.33***	-12.35***	7.07***	8.46***
	(5.33)	(5.19)	(11.54)	(11.90)	(1.62)	(1.59)	(1.12)	(1.03)
Poverty rate	-12.83***	-5.94***	-23.15***	-14.58***	-6.98***	-5.27***	2.80***	2.01***
	(2.11)	(2.15)	(4.61)	(5.16)	(0.72)	(0.73)	(0.42)	(0.40)
Share of Black Americans	-1.42***	-1.03**	0.97	1.04	0.61***	0.28**	0.19***	-0.20**
	(0.35)	(0.44)	(0.80)	(1.04)	(0.11)	(0.14)	(0.07)	(0.08)
Constant	323.89***	143.49***	620.46***	382.57***	215.26***	173.41***	-132.14***	-105.10***
	(43.71)	(45.91)	(93.92)	(108.53)	(14.79)	(15.54)	(8.63)	(8.67)
State fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
R^2	0.299	0.439	0.387	0.481	0.596	0.657	0.717	0.806
N	1689	1689	982	982	2725	2725	1372	1372

Notes: Data for US counties. The dependent variable comes from Google's COVID-19 Mobility Reports and is of 18th of June 2020. Given that the data are changes from the baseline, a larger negative number indicates more social distancing.

Table A4: Additional regressions for social distancing and reported COVID-19 cases

	Social Distance Grade				COVID-19 Cases				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Density	-1.10***	-0.92***	-1.02***	-0.92***	0.12***	0.08***	0.02	0.04*	0.10***
	(0.06)	(0.06)	(0.05)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Population	-0.09	-0.42***	-0.39***	-0.66***	0.64***	0.73***	0.84***	0.82***	0.71***
	(0.11)	(0.11)	(0.10)	(0.11)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Days since first case	0.07	0.06	0.10*	0.09	0.82***	0.82***	0.82***	0.81***	0.83***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Median household income	4.27***	3.55***	3.04***	4.04***	-0.01	0.23	0.21	0.42**	0.38*
	(0.52)	(0.51)	(0.51)	(0.52)	(0.20)	(0.21)	(0.20)	(0.20)	(0.20)
Share 65 and older	1.42***	1.61***	1.11***	1.43***	-0.86***	-1.12***	-1.26***	-1.22***	-1.10***
	(0.21)	(0.21)	(0.21)	(0.21)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
Poverty rate	2.46***	2.05***	2.02***	3.06***	-0.02	-0.04	0.03	0.14	0.23**
	(0.30)	(0.30)	(0.30)	(0.31)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)
Share of Black Americans	-0.23***	-0.53***	-0.30***	0.13***	0.37***	0.23***	0.36***	0.36***	0.45***
	(0.04)	(0.08)	(0.04)	(0.05)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Share Hispanic	-0.17***				0.30***				
	(0.04)				(0.02)				
Share non-white		0.40***				0.21***			
		(0.09)				(0.04)			
Republican governor			-1.09***					-0.04	
			(0.10)					(0.04)	
Temperature March				-0.10***					-0.05***
				(0.02)					(0.01)
Temperature April				0.08**					0.02
				(0.03)					(0.01)
Temperature May				-0.06***					0.04***
				(0.02)					(0.01)
Total tests per 100,000							0.39***		
							(0.06)		
Constant					-4.37*	-7.58***	-10.61***	-9.78***	-10.60***
					(2.46)	(2.54)	(2.49)	(2.54)	(2.50)
Shape parameter					-0.20***	-0.13***	-0.13***	-0.12***	-0.15***
					(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	3054	3054	3054	3048	3141	3141	3141	3141	3135

Notes: Data for US counties. Ordered logistic for social distancing; negative binomial used for cases. Republican governor is an indicator equal to one if the state has a Republican governor. Temperature is the average temperature for the month listed and is drawn from the National Oceanic and Atmospheric Administration (NOAA). All covariates excluding the republican governor indicator and the temperature variables are logged. See main text and Tables for further details.

Table A5: Regressions for reported COVID-19 cases using the New York Times data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.17***	0.17***	0.18***	0.16***	0.17***	0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Population	0.67***	0.71***	0.64***	0.63***	0.60***	0.85***	0.90***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Days since first case	1.23***	1.19***	1.15***	1.15***	1.15***	1.06***	0.97***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Median h'hold income		-0.52***	-0.63***	-0.35***	0.73***	0.37*	0.18
		(0.08)	(0.08)	(0.10)	(0.21)	(0.21)	(0.21)
Share 65 and older			-1.51***	-1.57***	-1.36***	-1.22***	-1.36***
			(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
Gini index				1.51***	0.85***	0.07	0.06
				(0.27)	(0.29)	(0.28)	(0.27)
Poverty rate					0.73***	0.07	0.17
					(0.13)	(0.13)	(0.13)
Share of Black Americans						0.36***	0.40***
						(0.02)	(0.03)
Constant	-9.98***	-4.70***	1.67*	0.42	-14.27***	-10.03***	-7.44***
	(0.26)	(0.86)	(0.92)	(0.96)	(2.73)	(2.65)	(2.74)
Shape parameter	0.08***	0.07***	-0.02	-0.03	-0.03	-0.11***	-0.31***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
State fixed effects	No	No	No	No	No	No	Yes
N	3142	3141	3141	3128	3128	3128	3128

Note: Data for US counties. Negative binomial used for estimation. All covariates are logged. The dependent variable is cases as reported by the New York Times as of 18 June 2020, available through their GitHub repository. Variable descriptions can be found in Table 1 notes. Standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A6: Interaction effects with time since the first infection using cases data from the New York Times

	(1)	(2)	(3)	(4)	(5)	(6)
Density	0.17***	0.17***	0.16***	0.17***	0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Population	0.71***	0.64***	0.62***	0.59***	0.82***	0.87***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Days since first case	0.79	0.02	-0.12	-18.15***	-13.64**	-12.60**
	(2.55)	(2.68)	(2.57)	(5.56)	(5.60)	(5.21)
Median household income	-0.68	-1.38	-2.04**	-6.39***	-4.54**	-4.57***
	(1.05)	(1.04)	(1.03)	(1.81)	(1.88)	(1.75)
Median income # days	0.04	0.17	0.38	1.62***	1.10***	1.06***
	(0.24)	(0.24)	(0.23)	(0.41)	(0.43)	(0.40)
Share 65 and older		-0.55	-0.41	-1.75**	-2.56***	-2.72***
		(0.77)	(0.72)	(0.86)	(0.85)	(0.80)
Share 65 and older # days		-0.22	-0.27	0.09	0.30	0.30
		(0.18)	(0.17)	(0.20)	(0.19)	(0.18)
Gini index			-8.89***	-4.26*	-0.45	-2.18
			(1.94)	(2.36)	(2.30)	(2.14)
Gini index # days			2.37***	1.16**	0.10	0.50
			(0.44)	(0.54)	(0.53)	(0.49)
Poverty rate				-3.55***	-2.78**	-2.20**
				(1.13)	(1.14)	(1.08)
Poverty rate # days				0.98***	0.65**	0.53**
				(0.26)	(0.26)	(0.25)
Share of Black Americans					-1.04***	-0.91***
					(0.29)	(0.27)
Share of Black Americans # days					0.32***	0.30***
					(0.06)	(0.06)
Constant	-2.93	6.80	6.46	70.92***	55.38**	53.62**
	(11.32)	(11.90)	(11.45)	(24.44)	(24.64)	(22.91)
Shape parameter	0.07***	-0.02	-0.03	-0.05*	-0.13***	-0.33***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
State fixed effects	No	No	No	No	No	Yes
N	3141	3141	3141	3128	3128	3128

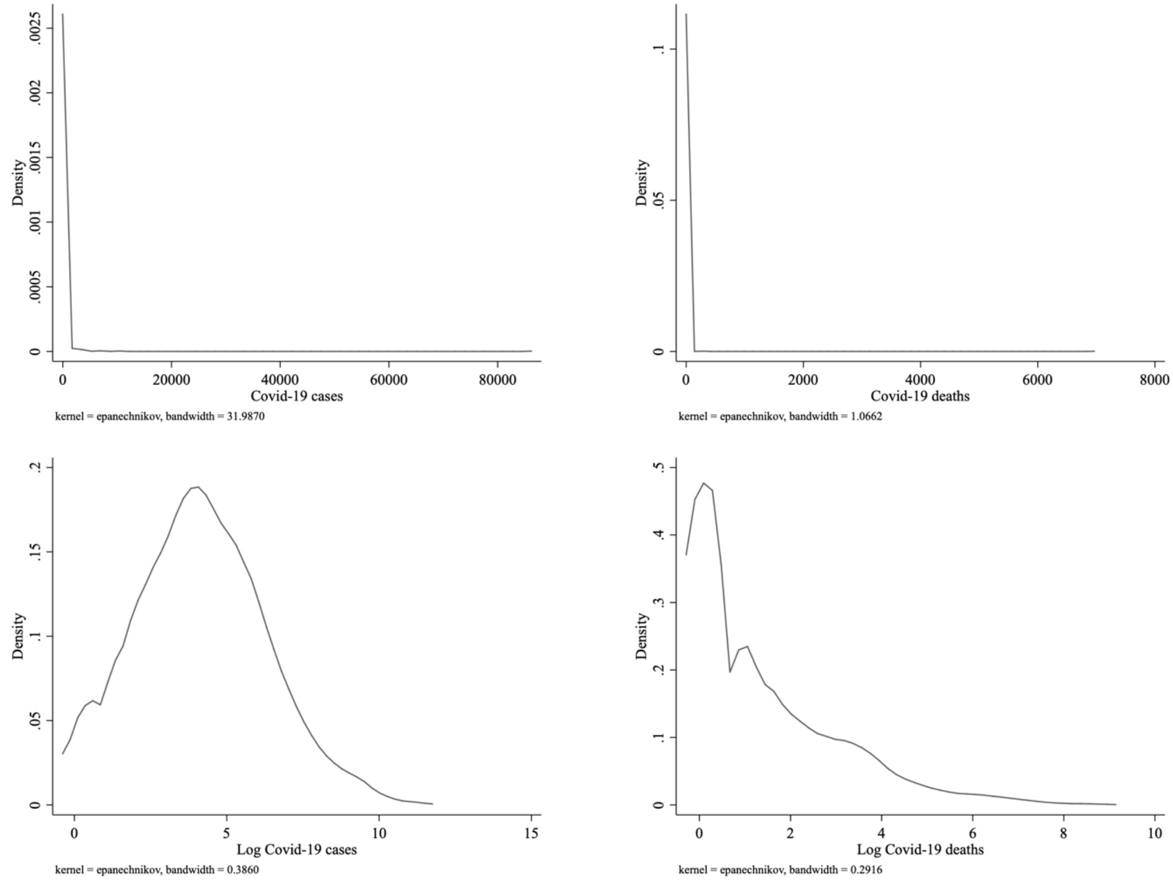
Note: Data for US counties. Negative binomial used for estimation. All covariates are logged. The dependent variable is cases as reported by the New York Times as of 18 June 2020, available through their GitHub repository. Variable descriptions can be found in Table 1 notes. The symbol # represents an interaction term. Standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A7: Reduced form regressions for COVID-19 deaths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Density	0.36***	0.42***	0.43***	0.40***	0.41***	0.23***	0.14***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Population	0.39***	0.36***	0.34***	0.33***	0.29***	0.59***	0.73***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Days	3.26***	3.31***	3.29***	3.12***	3.08***	2.50***	2.27***
	(0.26)	(0.26)	(0.26)	(0.26)	(0.26)	(0.25)	(0.24)
Hospitals per 100,000	0.16*	0.12	0.11	0.03	0.05	0.08	-0.06
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
ICU beds per 1,000	-0.43***	-0.70***	-0.70***	-0.74***	-0.68***	-0.67***	-0.51***
	(0.16)	(0.17)	(0.17)	(0.17)	(0.16)	(0.16)	(0.15)
Proportion with diabetes	0.07***	0.04***	0.04***	0.04***	0.03***	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Proportion diabetic with high BP	-0.00	-0.01	-0.01	-0.02	-0.01	-0.07***	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Proportion with asthma	-0.01	0.00	0.01	0.03	0.04	0.15***	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Proportion with COPD	-0.02	-0.04	-0.05	-0.04	-0.04	0.01	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Median household income		-0.97***	-1.00***	-0.56***	0.91***	0.18	0.10
		(0.16)	(0.16)	(0.18)	(0.35)	(0.34)	(0.33)
Share 65 and older			-0.23	-0.31**	-0.02	0.11	0.01
			(0.15)	(0.15)	(0.16)	(0.15)	(0.15)
Gini index				2.47***	1.45***	0.15	0.48
				(0.44)	(0.48)	(0.48)	(0.46)
Poverty rate					0.99***	0.24	0.33*
					(0.20)	(0.20)	(0.20)
Share of Black Americans						0.47***	0.39***
						(0.04)	(0.04)
Constant	-22.47***	-11.88***	-10.79***	-11.83***	-31.79***	-19.39***	-20.68***
	(1.20)	(2.09)	(2.21)	(2.23)	(4.62)	(4.58)	(4.29)
Shape parameter	0.59***	0.57***	0.57***	0.56***	0.54***	0.46***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
N	3141	3141	3141	3128	3128	3128	3128

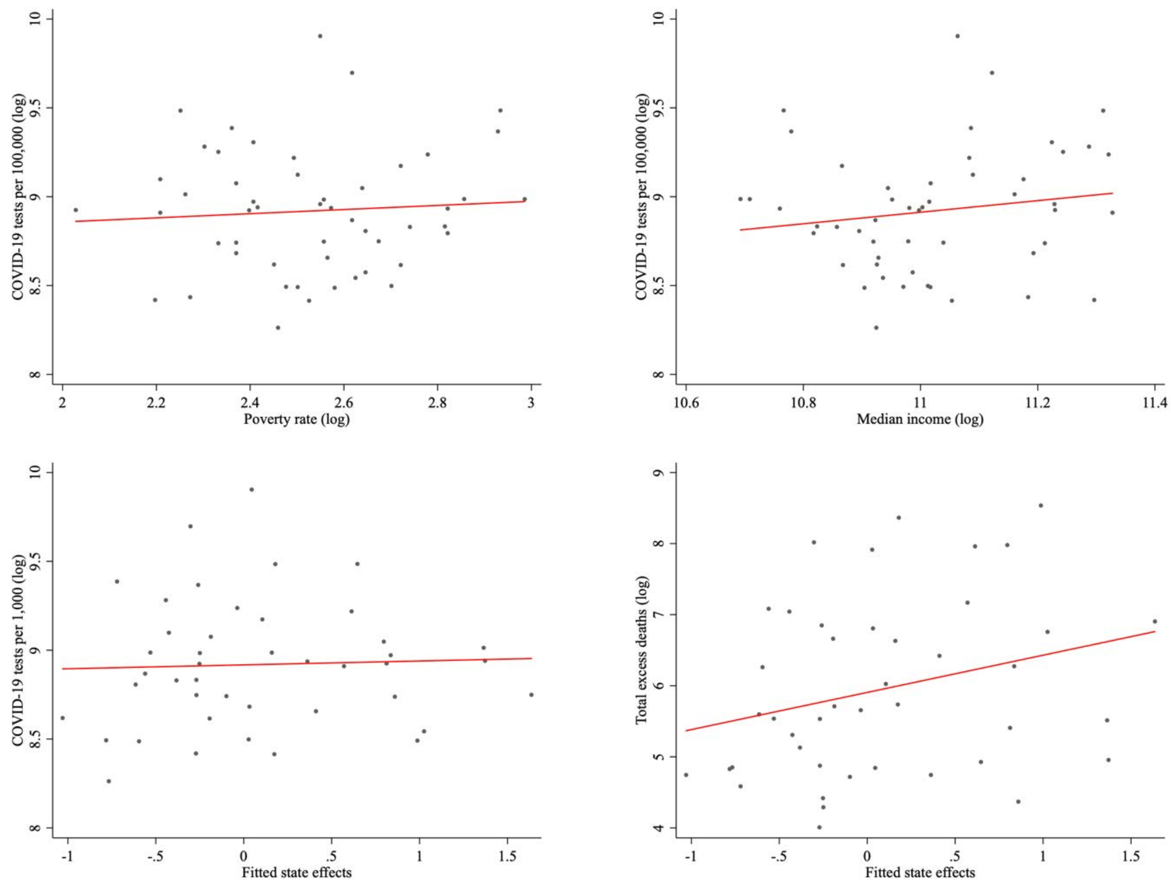
Note: All covariates excluding the health variables are logged. Hospitals and ICU beds are transformed using an inverse hyperbolic sine function. A negative binomial (NB) model is used for estimation. Variable descriptions can be found in Table 1 notes. Standard errors in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Figure A1: Density functions for COVID-19 cases across US counties



Note: Data on cases and deaths is from the CDC. Cases and deaths are cumulative until 18th of June 2020. Given that some counties have zero cases and deaths, the log variables in the second row have been calculated by adding one and then applying a log transformation.

Figure A2: COVID-19 testing and excess deaths across states



Note: Data is at the state level. Testing totals are from the COVID-19 Tracking Project and are cumulative until the 18th of June 2020. Fitted state fixed effects are estimated for the regression model with cases as the dependent variable with the full set of socio-economic variables and state fixed effects (with no constant terms), as in Column 6 in Table 2. The t-statistics for the fitted lines are 0.5 for the poverty rate (log), 1.11 for median income (log), 0.25 for state fixed effects on COVID-19 tests, and 1.83 for state fixed effects on excess deaths (+1, log).