# Close Encounters of a Heterogeneous Kind: Understanding the Differential Impact of Social Distancing on COVID-19 Infections and Deaths

Gibson, John and Johnson, David and Alexi, Thompson

University of Texas at El Paso, University of Central Missouri, Indiana University of Pennsylvania

1 December 2020

# Close Encounters of a Heterogeneous Kind: Understanding the Differential Impact of Social Distancing on COVID-19 Infections and Deaths *

John D. Gibson†
University of Texas at El Paso

David B. Johnson‡
University of Central Missouri

Alexi S. Thompson§
Indiana University of Pennsylvania

December 5, 2020

**Abstract**

We investigate the relationship between social distancing, as measured by encounter rates using cellphone proximity data, and COVID-19 infections and deaths. Consistent with the existing literature on the effectiveness of non-pharmaceutical interventions, we find a positive and statistically significant relationship between the encounter rate and new infections and deaths. However, the magnitude of this effect is relatively weak. One explanation for this weak effect is that the effectiveness of social distancing varies across counties due to local population heterogeneity. To this end, we interact the encounter rate with county-level characteristics and find that several of these interaction terms are statistically significant. Furthermore, after controlling for these interaction terms, we find that the effect of social distancing is around 15 times larger than the effect size found in estimates without the interactions.

**JEL classification:** D0, I12, 118, Z18
**Keywords:** COVID-19; Social Distancing; County-Level Variation

*We thank Levi Boxwell, Jacob Conway, Mathew Webb, Christopher Azevedo, Paul Chambers, Catherine Chambers, and Henry Thompson for their valuable comments and suggests. All mistakes are our own.

†Department of Economics and Finance, University of Texas at El Paso, TX, USA, 79902; `jdgibson@utep.edu`; phone (915) 747-5435.

‡Department of Economics, University of Central Missouri, MO, USA, 64093; `djohnson@ucmo.edu`; phone (336) 639-2190.

§Department of Economics, Indiana University of Pennsylvania, PA, USA, 15705; `Alexi.Thompson@iup.edu`; phone (724) 357-6485.

# 1 Introduction

We examine the effect of social distancing on the novel coronavirus (COVID-19) outcomes using daily county-level data on the number of COVID-19 infections and deaths. Similar to contemporary work, we measure social distancing with encounter rates recovered from cellphone proximity data (see Section 2.2 for more details), and consistent with these existing studies, we find a small positive yet statistically significant relationship between encounter rates and new COVID-19 infections and deaths. To investigate this effect further, we interact encounter rates with an array of county-level characteristics, with the goal being to determine if the effectiveness of social distancing is impacted by the characteristics of the underlying population. We find that many of these interactions are statistically significant, with race and political affiliation (among others) affecting the relationship between social distancing and COVID-19 outcomes. We also find that after controlling for this variation in the effectiveness of social distancing, the estimated average effect of social distancing increases by a factor of fifteen. This suggests that cellphone data as a measure of social distancing is useful but does not tell policymakers much about other suggested non-pharmaceutical interventions (NPIs) that are being used in conjunction with (or even in place of) social distancing as well as other vulnerabilities that may shape the relationship between encounters and outcomes.[1]

Understanding the impact of social distancing and other non-pharmaceutical interventions, such as mask-wearing, on COVID-19 infections and deaths is extremely important. Studies on the effectiveness of social distancing during modern pandemics suggest that non-pharmaceutical interventions are extremely effective at curbing infections and deaths. Caley et al. (2008) estimates that social distancing interventions saved 260 per 10,000 lives in Sydney, Australia during the Great Influenza Epidemic of 1918 (also erroneously known as the "Spanish Flu"). Kelso et al. (2009) models the effects of social distancing measures (e.g., school closures and workplace nonattendance) on a computer-simulated community and find that these measures reduce cases from 7% to 73% depending on the contact rate, duration, how quickly social distancing is adopted, and the average number of secondary cases generated by a single infection (R0).[2] Finally, Fong et al. (2020) presents a meta-analysis of the effectiveness of various non-pharmaceutical interventions (including quarantine, isolation, contact tracing, workplace closures, and school closures) and find that these interventions reduce viral transmission and represent a valid mitigation effort.

While our results echo that of previous studies on the effectiveness of social distancing, finding a positive association between encounter rates and COVID-19 infections and deaths, we are primarily interested in how the effectiveness of such practices are impacted by the local population. Stated differently, we also consider the extent to which the characteristics of the local population influence the success of social distancing efforts. As an illustrative example, one can imagine two counties that both close schools and shutter non-essential businesses. Given that both counties are subject to the same policy, they will look similar

[1] For example, we show that holding county-level encounters fixed, counties with a higher 2016 vote for Donald Trump see a significantly larger COVID-19 infection and death rate than counties with a lower 2016 vote for Donald Trump.

[2] For example, R0 of 5 means one infected individual infects five others. In terms of well-known diseases, Mumps has a relatively high R0 factor (Béraud et al., 2018); HIV has a low one (Holtgrave, 2010).

in terms of social distancing measures. However, if the residents of one county fully embrace mask-wearing, hand washing, and limit their informal social gatherings while the residents of the other county refuse to do the same, then the spread of the virus may differ substantially between the two counties. Understanding how variation in local characteristics can alter the effectiveness of public policy is important, especially given the strong political divide that exists within the United States.

In a closely related paper, Allcott et al. (2020a) considers the role played by both policy interventions, such as stay-at-home orders, and voluntary social distancing on economic and health outcomes at the combined statistical area (CSA) level. Allcott et al. (2020a) finds that the policy response is modest and most of the effect is driven by voluntary responses. However, one of the key characteristics of the U.S. COVID-19 pandemic, namely the wide geographic variation in severity of impact, is left unexplained by both public policy and voluntary social distancing. Instead, the authors conclude that "exogenous characteristics such as population and density" are the primary determinants of this geographic heterogeneity. The authors reach this conclusion using the SIRD (Susceptible, Infected, Recovered, Deceased) model of disease transmission to estimate the effect of stay-at-home orders on the contact rate, finding that these policies reduce the contact rate by 9 to 14 percent. Using these estimates, the authors conduct counterfactual analysis by equalizing policy and social distancing across all CSAs. Even after equalizing these two variables, they find that about half of the variation in high and low transmission CSAs is due to variation in the population and population density and that racial composition and partisanship explain a smaller share.

Our work differs from Allcott et al. (2020a) in a very important way. While both papers focus on an aspect of geographic variation during the COVID-19 pandemic, Allcott et al. (2020a) focuses on the number of cases in a given area or region, while we focus on the effectiveness of social distancing practices. Allcott et al. (2020a) attempts to uncover the determinants of infection "hot spots" but generally views social distancing scores as identical across various geographic areas. Allcott et al. (2020a) recognizes that this may not be the case by stating that their measure of social distancing cannot distinguish between high and low-risk activities. In our paper, we consider this point further by interacting our measure of social distancing (the encounter rate) with county-level characteristics. Many of these interaction terms are found to be statistically significant, suggesting that the effectiveness of social distancing is impacted by the underlying characteristics of the local population. We believe that these underlying characteristics (such as race, income, political affiliation, etc.) serve as an indirect measure of risky behavior, social trust, and access to quality health infrastructure (e.g. access to testing and treatment).[3]

We present our results in steps, focusing first on the average relationship between social distancing and COVID-19 infections and deaths, and then establishing the link between this relationship and the underlying county-level characteristics. We first estimate the relationship between lagged encounter rates and COVID-19 outcomes, controlling for local trends, day of week effects, and county fixed effects, and we find a small but statistically significant positive relationship between encounter rates and COVID-19 outcomes. We posit that the low magnitude of this relationship is due to substantial heterogeneity at the county-level.

[3] It should also be noted that our paper considers a much longer time horizon (January 22 to November 29), while Allcott et al. (2020a) focuses on the first wave of the pandemic (March 15th to April 30th).

To demonstrate this, we estimate the relationship between encounters and COVID-19 outcomes for each county in Florida and, as expected, we observe substantial heterogeneity. We then turn back to the full sample and show that county-level characteristics interacted with encounter rates greatly impact the effect of encounters on COVID-19 outcomes. Furthermore, once these interaction terms are included, the magnitude of the relationship between COVID-19 outcomes and encounters increases by over a factor of fifteen. This demonstrates that the actual relationship between social distancing and COVID-19 outcomes is much larger than previously estimated.

The paper proceeds as follows. Section 2 provides an overview of the data used in our analysis. Section 3 presents our results, which are shown in steps, and Section 4 discusses these results and concludes.

## 2 Data Description

We use both time-invariant and time-varying data in our analysis. The time-invariant data consists of county-level characteristics (e.g., percent of the county that identifies as male, etc.), while the time-varying data includes COVID-19 infections, COVID-19 deaths, and encounter rates, and is also measured at the county-level.[4] The next few subsections provide detailed descriptions of each of these data.

### 2.1 Time-Invariant Data

Summary statistics for the time-invariant data are found in Table 1 and the distributions of each of these variables can be found in Figure 6 (Appendix). All time-invariant data is measured at the county-level and is taken from 2018 population estimates/counts. Variable descriptions are as follows: %Age >75 is the percent of the population that is over the age of 75, the Infant Death Rate is the number of infant deaths per 100,000 infants, % GOP is the percent of the county that voted Republican in the 2016 election (MIT Election Data and Science Lab, 2018), income is median county income, % Black is the percent of the county residents who are Black or African American, % Essential are the number of employees working at "essential" businesses per 100 county residents,[5] % College is the percentage of the county with a college education averaged over the years 2014 to 2018, and % Male is the percent of the county that identifies as male. All of our data comes from readily accessible public sources.[6] Correlations of the time-invariant variables are found in the Appendix (Table 5)

[4]Independent cities, typically in Virginia, are merged with their neighboring counties.

[5]We consider workers employed in the following industries to be classified as essential: Construction (NAICS: 23), Health Care and Social Assistance (NAICS: 62), Administrative and Support and Waste Management and Remediation Services (NAICS: 56) and Executive, Legislative, and Other General Government Support (NAICS: 9211).

[6]Data sources include the Small Area Income and Poverty Estimates (United States Census Bureau, 2019b), the United States Department of Agriculture (United States Department of Agriculture, 2020), the CDC's WONDER databse (Centers for Disease Control and Prevention, 2019) and the County Business Patterns Survey (United States Census Bureau, 2019a).

Table 1: Time Invariant County-Level Summary Statistics

| | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| %Age > 75 | 8.25 | 2.37 | 1.17 | 27.3 |
| Infant Death Rate | 632.6 | 770.6 | 0 | 10344.8 |
| % GOP (2016) | 63.4 | 15.6 | 4.09 | 94.6 |
| Income | 52.8 | 13.7 | 25.4 | 140.4 |
| % Black | 9.85 | 14.5 | 0.097 | 86.6 |
| % Essential | 7.27 | 3.93 | 0 | 47.1 |
| % College | 21.5 | 9.32 | 0 | 74.6 |
| % Male | 50.1 | 2.27 | 43.1 | 73.2 |
| Observations | 3111 | | | |

## 2.2 Time-Varying Data

Summary statistics for our time-varying data are presented in Table 2. Variable descriptions are as follows: Total Infections is the number of cumulative cases of COVID-19 in county $i$ on day $t$, New Infections is the difference in total infections in county $i$ in day $t$ from day $t-1$,[7] and Day Number reports the day in consideration, with Day 1 corresponding to January 22 of 2020 (the day following the first positive COVID-19 test in the United States) and Day 291 corresponding to November 29 of 2020.

Table 2: Time-Varying County-Level Summary Statistics

| | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Infections | 1214.6 | 6626.4 | 0 | 408396 |
| Deaths | 37.3 | 249.5 | 0 | 7700 |
| New Infections | 13.8 | 80.0 | 0 | 14129 |
| New Deaths | 0.27 | 2.81 | 0 | 707 |
| Day Number | 158 | 90.9 | 1 | 315 |
| Encounter Rates | 0.26 | 22.2 | -1.00 | 5675.1 |
| Observations | 980280 | | | |

To measure social distancing at the county-level, we use cellphone tracking data reported by Unacast (2020). This data starts on February 22, 2020 (day 34 of our panel) and runs to November 29, 2020.[8] Typically, this type of cellphone data is used to ensure internet connectivity and provide device authentication (Sprenger, 2019). We believe cellphone data is also a reasonable proxy for social distancing as it provides an approximate location of cellphones, which are carried regularly by a majority of Americans. For example, a study by the Pew Research Center (2019) finds that 92% of American adults own a cellphone and that 67% own a smartphone. The same study finds that 90% of cellphone owners "frequently" carry their cellphone. Similar cellphone data has been used in a number of recent studies as a proxy for social distancing (Peak et al., 2018; Gollwitzer et al., 2020; Gao et al., 2020). However, several of these studies utilize data from SafeGraph INC (2020)

[7] Due to reporting errors there are sometimes negative new infections. To deal with these errors, we truncate new infections at zero.

[8] See Ngo (n.d.) for additional details.

rather than Unacast (2020). As such, it is natural to question whether these two data sources differ in meaningful ways. We do not believe this is the case. First, the Unacast (2020) data is constructed in a similar way to SafeGraph INC (2020), using anonymous data from GPS pings from millions of cellphones in the U.S. (Goolsbee and Syverson, 2020).[9] Second, the pattern of social distancing shown in studies using SafeGraph INC (2020) data looks similar to the Unacast (2020) data; encounters plummet contemporaneously with state shut down orders but also remain low after the policies expired (See Figure 4). Arguably, the primary difference between the Uncast data and SafeGraph data is that the Unacast data is reported at the county level, which maps nicely to publicly available census and COVID-19 outcome data.

The measure of social distancing we use is the encounter rate constructed by Unacast (2020). This measure is a time-varying county-level measure of social distancing. The Unacast (2020) encounter rate is constructed using the following three components; i) the number of encounters at the county-level, ii) the county land area, and iii) the average encounter rate at the national level during the pre-COVID-19 time period (February 10, 2020 - March 8, 2020). An encounter occurs when 2 devices are within 50 square meters of each other for less than hour.[10] The number of encounters within a county is then divided by the size of the county (square kilometers) and this ratio is weighted by the average pre-COVID 19 national encounter rate. Unacast then subtracts 1 from this term to express the measure as a rate, so if on day t county i has an encounter rate equal to the pre-COVID baseline their encounter rate would be 0.

### 2.3 Trends in Time-Varying data

Figure 4 presents the number of new infections/deaths in the United States (per million) and the average encounter rate by day. Corresponding figures for each US state can be found in Figures 7, 8, 9, 10, 11, and 12 - with the first two presenting new infections, the second two presenting daily deaths and the last two relating to social distancing. While these figures present an overall trend in infections, deaths and encounters, they also point to "seasonality" - with new infections peaking on Friday and social distancing generally peaking on Sunday.

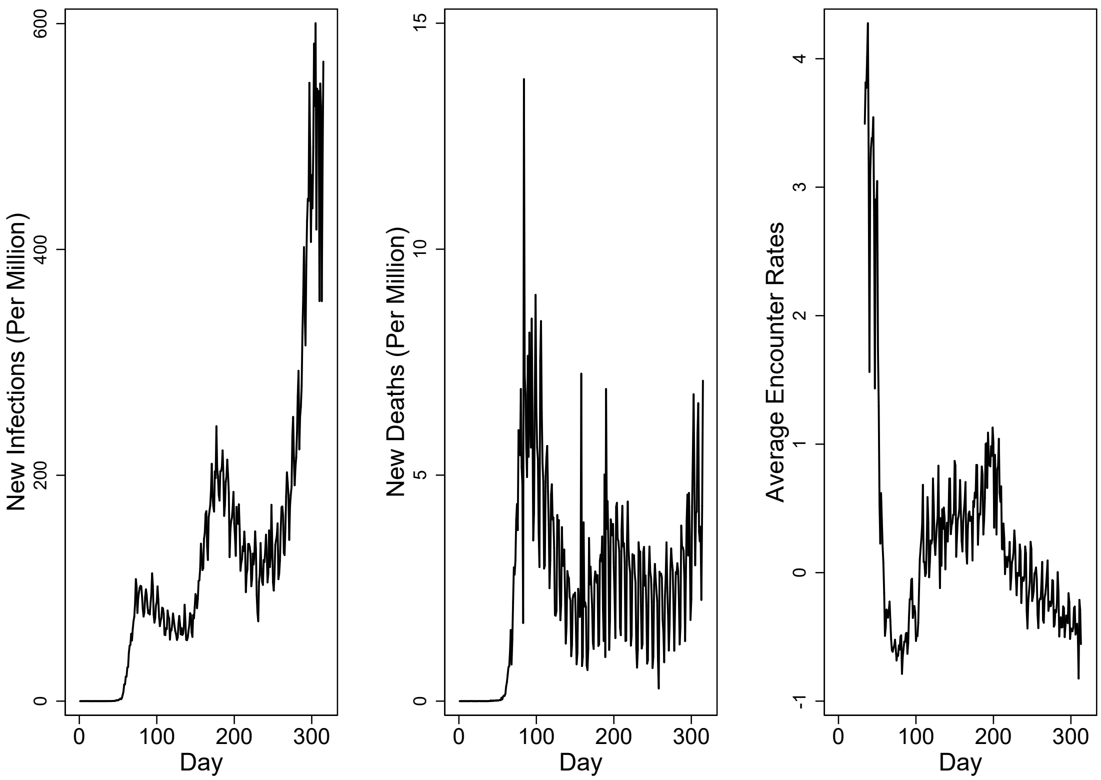
### 2.4 Geographic Variation in Social Distancing and COVID-19 Deaths

Social distancing and new infections vary across time and geography. Figure 2 presents social distance scores by US county on March 21 and the total number of infections on that day, as well as 10, 20, and 30 days after. Initially, infections appear most concentrated in densely populated areas on the East Coast (particularly around New York City and

[9] The Safegraph dataset focuses on the number of visits and the amount of time spent at points of interest; such as restaurants, bars, and hotels. Recent publications using this data include Painter and Qiu (2020); Andersen (2020); Dimke et al. (2020); Brzezinski et al. (2020); Glaeser et al. (2020).

[10] This time dimension has both strengths and weaknesses. On one hand, it prevents interactions caused by neighborhoods/families/offices but on the other, it discounts interactions at large gatherings lasting longer than an hour (e.g., a concert). While somewhat troubling we are not terribly concerned because these types of events (e.g., Concerts and Major League Baseball games) were generally canceled or took place with limited/no public attendance.

Figure 1: Trends in Daily Infections, Daily Deaths, and Social Distancing




**Notes:** New COVID-19 infections (left), deaths (center) and average encounter rates (right) by day. Day 1 corresponds to January 22, 2020, which is when the CDC started tracking data. The first reported case in the US was the day before; January 21. Social distancing data does not start until February 22, 2020 which corresponds to day 34.

Newark), as well as southern California and southern Florida. Over the course of the next 30 days, the number of cases increased dramatically in the United States, particularly in the Southeastern and Southwestern parts of the country. Many of the areas that experienced the largest increases in new COIVD-19 cases were also the areas that practiced less social distancing 30 days prior. For instance, the Southeastern portion of the United States had relatively few cases of COVID-19 in mid-March, but during this period practiced relatively less social distancing. As we will show, this led to an increase in the total number of infections weeks later.

## 3 Results

For geographic reasons, we focus our attention on the continental United States, though our results are robust to the inclusion of Hawaii and Alaska (results available upon request). In Section 3.1, we demonstrate that counties with higher encounter rates tend to have more COVID-19 infections and deaths than counties with lower encounter rates. However, we show that this relationship, while statistically significant, is relatively small. Next, in Section 3.2 we show that there is substantial heterogeneity in the correlation of encounters and COVID-19 infections and deaths. This heterogeneity is what is generating the relatively weak association. And finally, in Section 3.3, we demonstrate that some of this variation is explained by the interaction of the encounter rates with county characteristics.

In all of the presented analysis, we omit observations in which there have been no previously reported infections. These omitted observations are in the early days of the pandemic. In Figure 13 (left panel), in the Appendix, we show how the number of counties with no infections changes over the course of the pandemic. Some counties have no COVID-19 deaths - these counties are also omitted and a graph illustrating the number of counties with no COVID-19 deaths is found in Figure 13 (right panel).

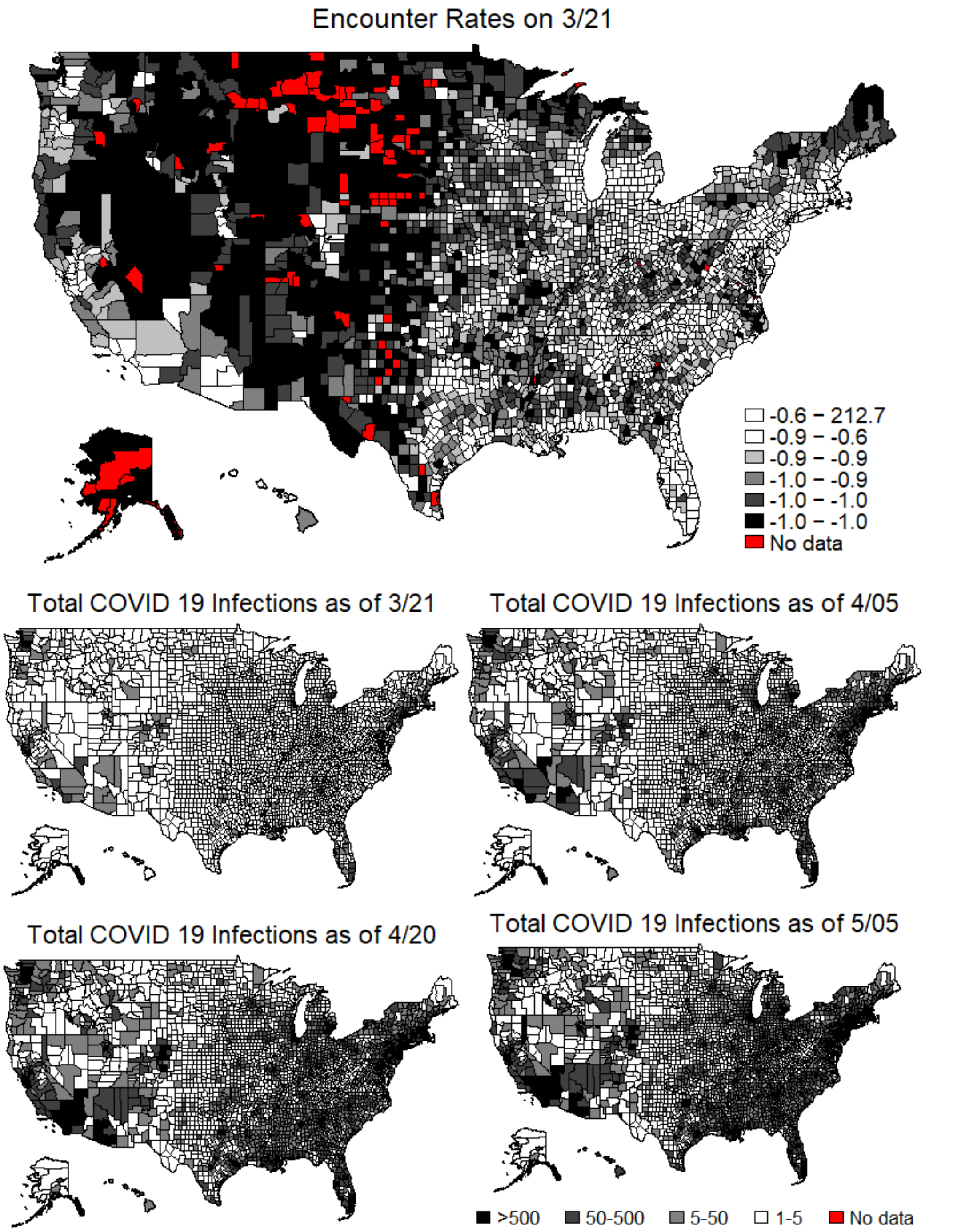
### 3.1 Encounters and COVID-19 Infections and Deaths

We begin by exploring the relationship between the encounter rate and COVID-19 infections and deaths. We do so by estimating the conditional mean of new COVID-19 infections (Equation 1) and deaths (Equation 2) in county $i$ on day $t$ with a fixed-effects Poisson ($I_{i,t}$ and $D_{i,t}$, respectively). This specification is chosen due to the relatively high variance of new infections. Many U.S. counties experience zero or very few new COVID-19 infections/deaths (count data). As such, typical linear models are not suitable.[11] Consequently, we rely on nonlinear estimation methods that are commonly applied to problems involving count data (Lovett and Flowerdew, 1989; Coxe et al., 2009; Hutchinson and Holtman, 2005). Given that our data is panel in nature, we use the fixed effects Poisson model (Hausman et al., 1984), which is commonly used in studies of infectious disease (Ma, 2020; Persico et al., 2020; Homaira et al., 2018, 2012; Nasreen et al., 2014).[12] The reported coefficients are semi-

[11] While not our chosen specification, results found using a linear county fixed effects model are presented in Table 6 of the Appendix. Here we note that the main results hold; including the increase in the magnitude of the encounter rate (though the increase is smaller).

[12] Due to the large number of fixed effects present, we use the Poisson pseudo-likelihood regression with multiple levels of fixed effects Stata package to estimate our models (Correia et al., 2019).

Figure 2: Total Infections and Social Distancing by US County




**Notes:** Social distancing scores on March 21, 2020 (top panel) and total infections on March 31, April 10, and April 20. Darker colors correspond to less social distancing (i.e., lower social distancing scores) and more infections.

elasticities - meaning they give the percentage increase in the dependent variable caused by a one-unit increase in the explanatory variable which implies that the estimated effect size is not constant.

The primary independent variables are the square root of the encounter rates + 1 (ER) in county $i$ on day $t-21$ (for infections) and the ER +1 in county $i$ on day $t-39$ (for deaths).[13] We adopt the 21 and 39 day lag structure for infections and deaths to account for common delays in viral transmission, testing, and the appearance of symptoms. We have estimated the model with a variety of other lag structures and found consistent results.[14] For the sake of readability we will refer to this transformed encounter rate as the encounter rate. We take the square root of the Unacast encounter rate for two reasons.[15] First, as shown in Table 2, there is substantial variation in encounter rates; much of this variation is caused by a handful of counties in New York.[16] Second, and most importantly, the encounter rate is based on cellphone data which, in general, has two primary issues: first, cellphone coverage is poorer in more rural areas (meaning even if two cellphones are near each other, the encounter may not register) and second, relative to people living in urban areas, people in rural populations are less likely to own a smartphone (Perrin, 2019). In practical terms, these two characteristics imply the encounter rate in rural counties will over-state the level of social distancing. Though these same issues exist in urban areas it will not be as significant due to the better coverage and higher cellphone take-up rate. In our view, taking the square root of encounters partially corrects for this bias.

Other control variables in the main regressions include county fixed effects ($\alpha_i$), day of week effects fixed effects ($\omega$), week number fixed effects ($\psi$), and state-specific trends, ($\rho_s t$). We also include an "exposure" variable, which measures the county population minus the total number of infections ($\delta_{i,t}$). This variable is included in all estimated models and is the susceptible population within the county.[17] As such, our outcome variable can be interpreted as a rate.[18]

$$E(I_{i,t}|ER) = \exp\{\beta_0 + \beta_1 ER_{i,t-21} + \alpha_i + \omega + \rho_s t + \psi + \delta_{i,t}\} \tag{1}$$

$$E(D_{i,t}|ER) = \exp\{\beta_0 + \beta_1 ER_{i,t-39} + \alpha_i + \omega + \rho_s t + \psi + \delta_{i,t}\} \tag{2}$$

Table 3 presents our initial estimation results, with the first column corresponding to infections and the third column corresponding to deaths. For all models, standard errors are clustered at the county level. We first note that the sign of the coefficients are consistent with

[13] We add one to the encounter rate to prevent taking the square root of a negative number.

[14] See Figure 14 in the Appendix for the coefficient of the encounter rate from models with alternative lag structures.

[15] Though we present results using a transformation of the encounter rate, it should be noted that our results are qualitatively robust. Similar coefficient estimates (in terms of direction and significance) are found when we use the Unacast encounter rate or different transformed encounter rate (ER +1 raised to .25, .75, and .9 power rather than .5) rather than our transformed encounter rate. These estimates are found in Tables 8, 9, 10, and 11 in the Appendix.

[16] All of the 141 observations with an encounter rate greater than 400 are from the state of New York.

[17] There have been documented cases of reinfection (c.f., Tillett et al., 2020), but reinfection is still viewed as being extremely rare.

[18] We do not include lagged infections/new cases as an explanatory variable because it is already in the exposure variable. Further we omit observations that have no infections. However, regression results including lagged infections lagged infections interacted with encounters are in the Appendix (Table 7).

expectation, indicating that encounter rates are positively associated with infections. This result is consistent with what we learned in the first months of the pandemic (January and February 2020). Namely, that several mitigation strategy studies involving China (Fan et al., 2020; Pan et al., 2020), India (Mandal and Mandal, 2020), and Italy (Guzzetta et al., 2020) demonstrated quarantine, isolation, and other NPIs (all of which raise social distancing) are effective in reducing the number of cases.[19] Second, we note that the magnitude of the estimated coefficients are relatively small and with varying levels of significance - which is similar to what is found in Allcott et al. (2020a).

There are three reasonable explanations for a statistically significant, yet small positive association between encounter rates and COVID-19 infections and deaths. First, encounter rates might not influence infections and death by much, though this seems unlikely given current research. Second, the coefficient may be downwardly biased due to reverse causality, where areas with high COVID-19 infections and deaths also limit social distancing.[20] And finally, it may be the case that the nature of the relationship between encounter rates and COVID-19 outcomes varies significantly with the underlying characteristics of the population. If this is the case, then Table 3 is actually reporting the *average* relationship between encounters and COVID-19 outcomes. Therefore, if there is variation in the use of other NPIs at the county-level, we would expect the encounter rate to have a heterogeneous relationship.

As an initial test of this "heterogeneity hypothesis", we re-estimate Equation 2 using two sub-samples. The first sub-sample ($n = 241$) consists of counties that are in the top quartile in terms of the percent of the electorate that voted for Donald Trump in 2016 election (% GOP $>$ 74.65 %) and the bottom quartile in terms of income (Income $<$ \$43,671). The second sub-sample ($n = 293$) consists of the counties that are in the top quartile in income (Income $>$ \$58,715) and the bottom quartile of 2016 Donald Trump support (% GOP $<$54.54%). In both sub-samples, we find a positive and statistically significant relationship between lagged encounters and daily COVID-19 deaths. However, the point estimate is found to be quite large in the first sub-sample ($\beta_1 = 0.38$; $p = 0.022$) and quite small in the second sub-sample ($\beta_1 = 0.023$; $p = 0.03$). The difference in coefficient estimates across the two sub-samples provides initial evidence in support of county-level heterogeneity driving the low estimated effect of encounters on COVID-19 outcomes from our full-sample regressions.

### 3.2 Variation in the Effect of Encounters

To further test this "heterogeneity hypothesis" discussed above, we re-estimate the effect of encounters, focusing on the state of Florida, but here we also interact the lagged encounter rates with the county fixed effect. This process allows us to demonstrate how the effect of encounters differs across counties. We use Florida because it has been hit relatively hard by COVID-19 and it has a large number of diverse counties.[21] We include county fixed effects

[19]Banholzer et al. (2020) conduct an analysis of the effectiveness of various NPIs using cross-country data for 20 countries. They find all NPIs lead to reductions in new cases, with venue closures being the most effective.

[20]Reverse causality is almost certainly occurring to some degree.

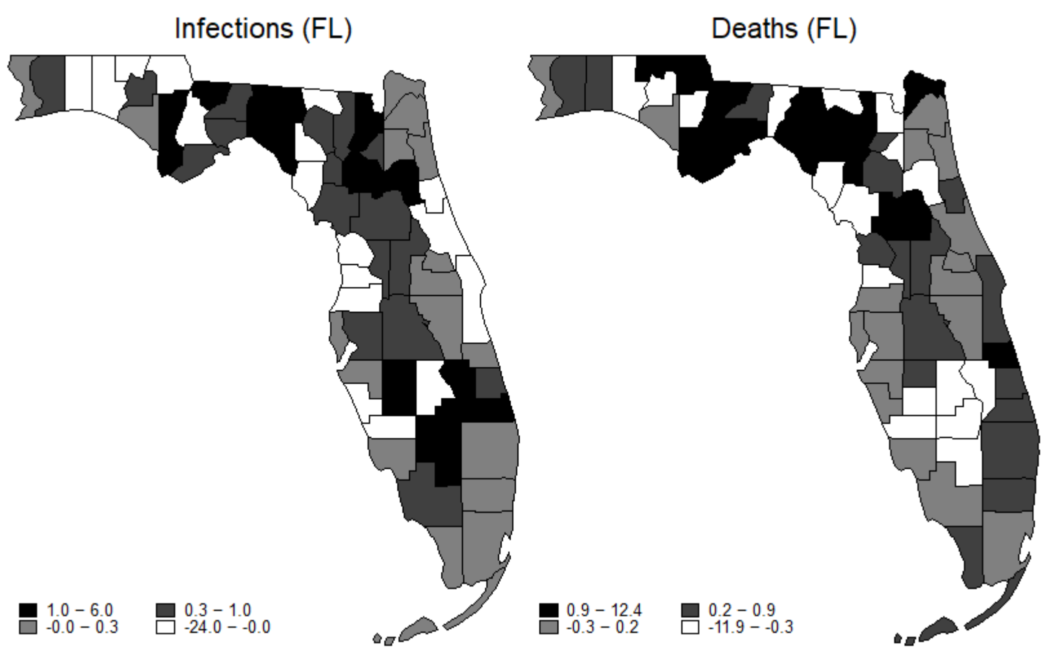
[21]In principle, we could estimate a similar model using national data but we found this estimation computationally infeasible as it required the estimation of roughly 3,000 additional coefficients.

($\alpha_i$), day-of-week effects fixed effects ($\omega$), week number fixed effects ($\psi$), and the exposure variable ($\delta_{i,t}$). The full specifications for these models are in Equations 5 and 4.

$$E(I_{i,t}|ER) = \exp\{\beta_i \alpha_i ER_{i,t-21} + \alpha_i + \omega + \psi + \delta_{i,t}\} \tag{3}$$
$$E(D_{i,t}|ER) = \exp\{\beta_i \alpha_i ER_{i,t-39} + \alpha_i + \omega + \psi + \delta_{i,t}\} \tag{4}$$

Because we are only interested in the differences in encounter rates, we present our results graphically rather than using a table - though these estimates are available upon request.[22] Graphical estimation results are presented in Figure 3. As expected, the estimated relationship between encounters and COVID-19 outcomes varies significantly at the county-level, indicating substantial heterogeneity in the estimated response. The impact of encounter rates is found to be largest in the northern panhandle.

Figure 3: Estimated County-Level Effect of Encounter Rates on New Infections and Deaths in the State of Florida




**Notes:** County specific coefficient estimates for each county in Florida. In the left-hand panel the outcome variable is COVID-19 infections. Right-hand panel is the outcome variable is COVID-19 deaths. Time period is from January 22, 2020 to November 29, 2020.

[22]We do not include these point estimates and standard errors for space reasons: there are 67 counties in Florida so presenting these results would require a large table.

### 3.3 The Joint Effect of Encounters and County Characteristics

Given the substantial heterogeneity in the effect of encounters on COVID-19 outcomes that we observed in the previous section, we now explore how this effect varies depending on the county's underlying demographic characteristics within our full sample. We take this approach because while estimates similar to those presented in the previous section could be done for each state/county, doing so does not tell us why they vary. To uncover the source of the variation, we estimate the following fixed effects Poisson:

$$E(y_{i,t}|ER,X) = \exp\left\{\beta_0 + \beta_1 ER_{i,t-p} + \sum_{j=2}^{8} \beta_j ER_{i,t-p} X_i + \alpha_i + \omega + \rho_s t + \psi + \delta_{i,t}\right\} \quad (5)$$

where $y_{i,t}$ is either the COVID-19 infections or deaths, $p$ is either 21 or 39 (in the case of infections and deaths, respectively), $X_i$ is a vector of county-level demographics/characteristics differenced from the national mean - which is done so the coefficient on the encounter rate can be interpreted as an "average" effect (i.e., the effect of encounters for a county with average characteristics). ER denotes the encounter rate, and as before, we also include the exposure variable. The county characteristics include the percent of the county population over the age of 75 (% Age > 75), the infant mortality rate (Infant Death), the percent of the county electorate who voted for Donald Trump in the 2016 election (% GOP), the percent of the county that are Black or African American (% Black), the percent of the county working in an essential industry (% Essential), the percent of the county with a bachelors degree (% College), and the percent of the county population that identifies as Male (% Male). Other control variables include county fixed effects ($\alpha_i$), day of week effects fixed effects $\omega$, week number fixed effects $\psi$, and state-specific trends, $\rho_s t$. Estimation results are presented in the second and fourth column of Table 3. For each of these models, we cluster standard errors at the county level.

First, note that as expected there is a statistically significant positive relationship between encounters and COVID-19 outcomes. However, these relationships are much stronger and more significant than those discussed in Section 3.1. In other words, once the "effect" of county characteristics are controlled for, the relationship between encounters and COVID-19 outcomes becomes substantially stronger. For a county with the average characteristics, a ten percent increase in the encounter rate from the national pre-COVID average results in a three percent increase in the number of new infections and a four percent increase in deaths per susceptible county resident.

We now discuss how the effect of encounters changes depending on the county characteristics. Several county characteristics are shown to positively affect the relationship between encounters and new COVID-19 infections. These include: % GOP, % Male, and % Black. For COVID-19 deaths, we note that almost every interacted characteristic is again statistically significant (% college and income are the only ones that are not). This finding is similar to what is found in Allcott et al. (2020b), which find that counties with a high prevalence of 2016 Trump supporters are less inclined to engage in social distancing. Our result goes one step further and shows that two counties with the same level of social distancing could have different COVID-19 death rates depending on differences in GOP support. GOP support is not the only characteristic that is found to increase COVID deaths. Other characteristics

Table 3: The Effect of Social Distancing on COVID-19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | $0.017^{***}$ | $0.29^{***}$ | $0.029^{**}$ | $0.38^{***}$ |
| | (3.13) | (3.91) | (2.12) | (7.08) |
| ER X % Age > 75 | | 0.014 | | $0.045^{***}$ |
| | | (0.91) | | (3.99) |
| ER X Infant Death | | -0.00031 | | $-0.00034^{***}$ |
| | | (-1.57) | | (-3.10) |
| ER X Income | | $-0.0080^{***}$ | | -0.0013 |
| | | (-3.28) | | (-1.14) |
| ER X % GOP | | $0.0087^{***}$ | | $0.0061^{***}$ |
| | | (2.90) | | (4.68) |
| ER X % Black | | $0.0080^{**}$ | | $0.0097^{***}$ |
| | | (2.12) | | (4.03) |
| ER X % Essential | | 0.0093 | | $0.0061^{**}$ |
| | | (1.41) | | (2.43) |
| ER X % College | | 0.0067 | | -0.0014 |
| | | (1.56) | | (-0.86) |
| ER X % Male | | $0.074^{**}$ | | $0.093^{***}$ |
| | | (2.14) | | (3.93) |
| LL | -6053705.1 | -6019371.7 | -375013.6 | -373084.9 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

$t$ statistics in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. Standard errors are clustered at the county.

include the percent of the population older than 75, the percent of the population that are Black or African American, the percent of the county working in an essential industry, and the percent of the county population that identifies as male. The opposite is found with only one characteristic: the infant death rate. The signs and coefficient estimates are consistent with the results from the infections equation and, perhaps more importantly, the precision of the effects increase.

We now ask the question, which of these interactions are *economically* significant? To answer this question, we calculate the marginal effect of encounters on COVID-19 outcomes at various levels of each characteristic (roughly 35-45 percent of the characteristic's standard deviation) while holding the other characteristics fixed at the national average.[23] These effects are presented in Figure 4 with the two left (right) columns corresponding to new infections (deaths). To give a better idea of how much the characteristic has to change to result in a meaningful difference, we also present each characteristic's mean (solid line) and first/third quartiles (dashed lines). In the top left panel of Figure 4 we show how the effect of lagged encounters on new COVID-19 cases is larger in areas with high 2016 GOP support - holding the other characteristics fixed at the national average; a 7.2 percent increase in GOP support increases the effect of encounters on infections (deaths) by about 7 (3) percentage points. Similar results are found with respect to the percentage of county residents who are African American or identify as male. Other characteristics have a more modest effect.

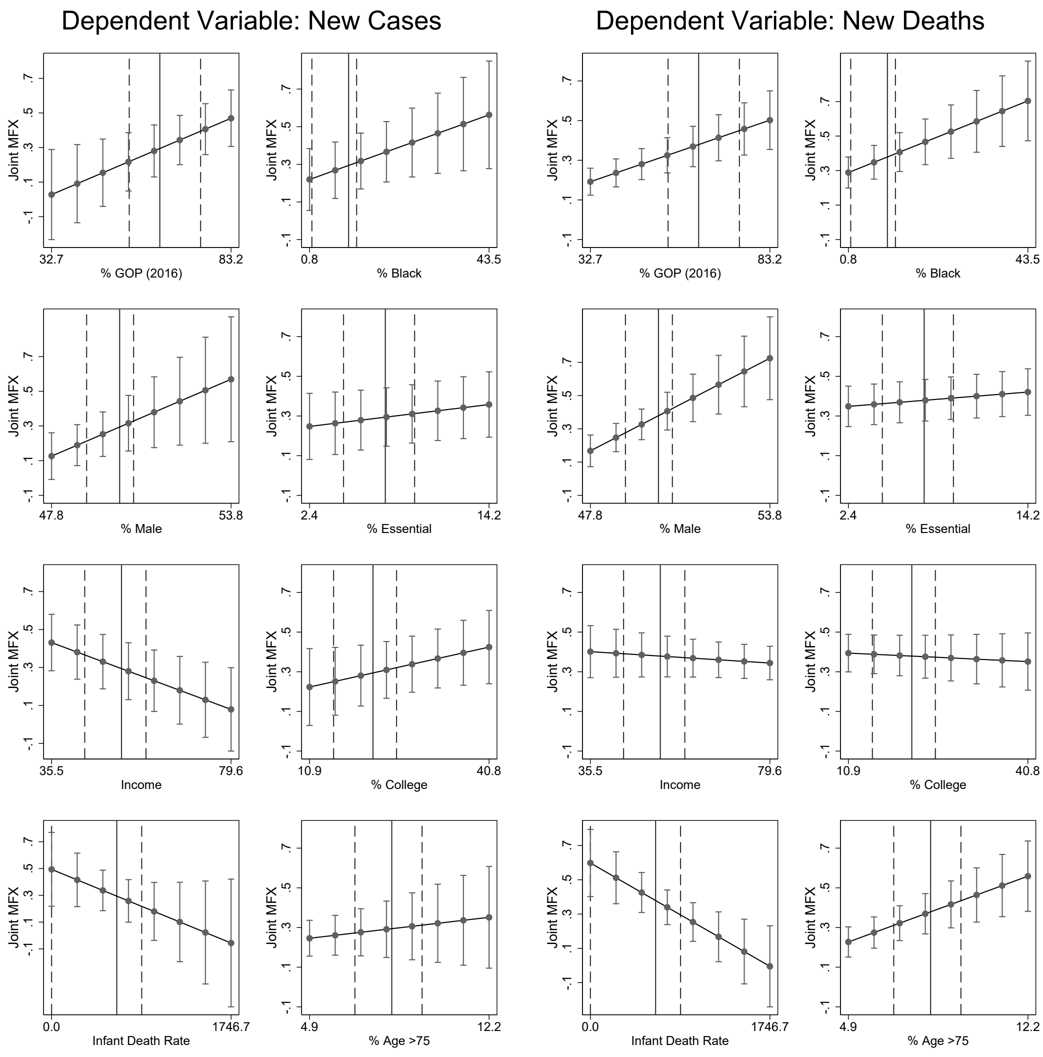
To further explain the observed heterogeneity in the effect of encounters we derive each county's estimated joint effect (EJE) of encounters on COVID-19 outcomes using the second (infections) and fourth (deaths) models presented in Table 3. The distributions of EJEs on infections and deaths are presented in Figure 5. With each of these distributions, we also include a solid "zero line" to identify the proportion of the EJEs that are negative (e.g., work in the opposite direction than we would expect) and a dashed "mean line" which corresponds to the EJE that would be observed in a county with the average characteristics. As expected, the EJEs are highly correlated ($\rho$= 0.97) and the vast majority of EJEs are positive. Overall, the estimated joint effect of encounters on infections is positive for 89% of counties in the continental United States. The figure rises to 93 % for deaths.

In Table 4, we present the mean county characteristics for counties with a positive (+ EJE) and negative EJE (- EJE). Here we find that the EJEs with signs that are contrary to expectations (i.e., negative) generally occur in counties with populations that are relatively young, heavily democratic, and that have a relatively small African American population (e.g., the counties we would expect, *ex-post*). These results highlight the nature of the effect of encounters: it is multi-dimensional. What this means is that a single quality/characteristic is not the only cause for a particular county's mortality rate. Further, this multi-dimensional nature makes it difficult to identify the average effect of encounters on COVID-19 outcomes.

As a final test of our results, we compare our EJEs with mask usage survey data (The New York Times and Dynata, 2020). This survey data is measured at the county level and is based on roughly 250,000 interviews conducted by Dynata in early July, 2020. During the survey interview, respondents were asked how often they wore a mask with the following

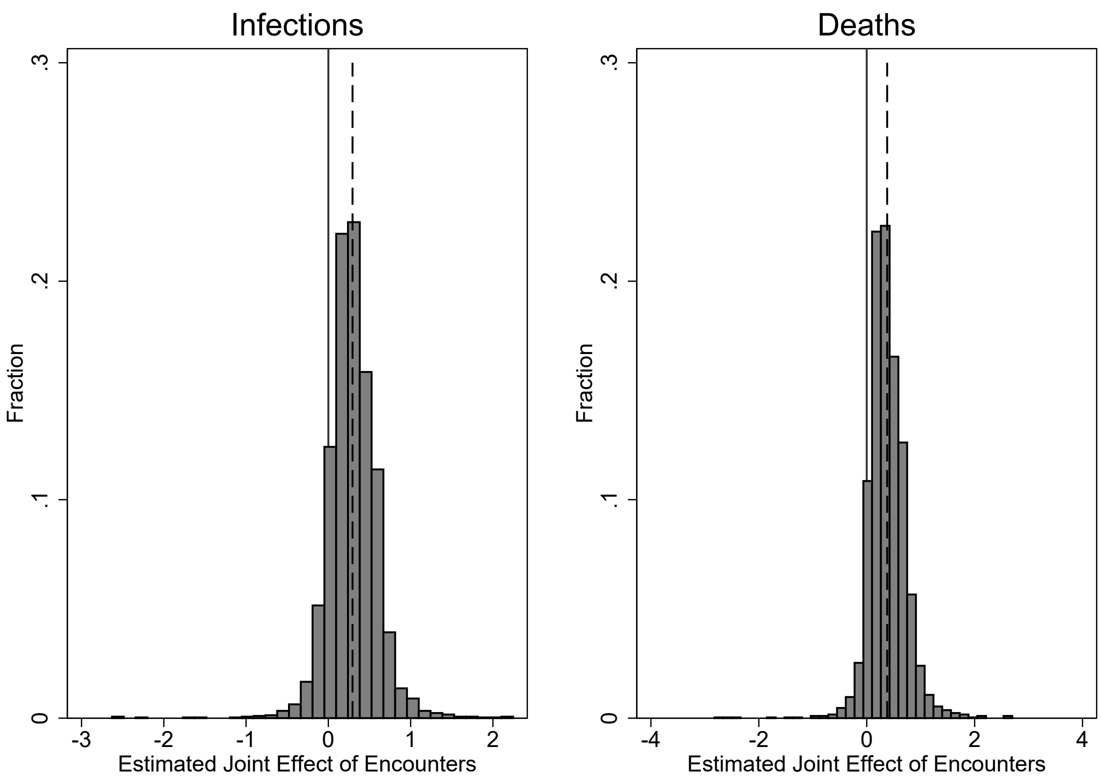
[23] These points of estimation are determined by splitting the spread between 5% and 95% percentile of the characteristic into 8 bins of equal distance.

Figure 4: Estimated Joint Effects




**Notes:** Interacted marginal effects. Solid line corresponds to the national mean. Dashed lines indicate first and third quartile. Other variables held fixed at national means.

Figure 5: Distribution of Estimated Joint Effects




**Notes:** Distribution of the estimated joint effect of encounters and county characteristics. Left panel corresponds to infections right corresponds to deaths. Solid line indicates zero while dashed line indicates the estimated effect in a county with the average characteristics.

Table 4: Average Characteristics of Counties with Positive and Negative Estimated Joint Effects

| | + EJE Inf. | - EJE Inf. | $p$ (Median) | $p$ (U-Test) | + EJE Deaths | - EJE Deaths | $p$ (Median) | $p$ (U-Test) |
|---|---|---|---|---|---|---|---|---|
| % Age >75 | 8.25 | 7.82 | .02 | 0 | 8.24 | 7.71 | .13 | 0 |
| Infant Death Rate | 516.57 | 1547.5 | 0 | 0 | 539.36 | 1949.11 | 0 | 0 |
| % College | 20.84 | 26.32 | 0 | 0 | 21.16 | 25.68 | .29 | 0 |
| % GOP (2016) | 64.544 | 52.44 | 0 | 0 | 63.84 | 53.77 | 0 | 0 |
| Income | 51.44 | 62.42 | 0 | 0 | 52.51 | 55.38 | .94 | .86 |
| % Black | 10.34 | 8.15 | .01 | .02 | 10.25 | 7.95 | 0 | 0 |
| % Essential | 7.37 | 7.44 | .11 | .35 | 7.38 | 7.37 | .94 | .91 |
| % Male | 50.11816 | 49.64 | .5 | .01 | 50.1 | 49.55 | .17 | .01 |
| Population | 91389.99 | 242945.9 | .81 | .11 | 107226.2 | 129627.7 | .02 | .01 |

**Notes:** Average characteristics of counties with positive and negative estimated joint effects. Columns 1 and 2 correspond to the groups of counties with a positive (column 1) and negative (column 2) EJE generated when estimating new cases of COVID-19. Columns 5 and 6 correspond to the groups of counties with a positive (column 5) and negative (column 6) EJE generated when estimating new deaths due to COVID-19. Columns 3, 4, 7, and 8 present p-values derived from median tests (3 and 7) and U-tests (4 and 8).

possible responses: Never, Rarely, Sometimes, Recently, and Always. If our EJEs are serving as a proxy variable for risk adjusted encounters, then we would expect to observe a significant correlation between our EJEs and reported mask usage. This is exactly what we find. Specifically, the correlation between the percent of county respondents who report to “always” wear a mask and the EJEs generated in our infections and deaths regressions are -0.2027 and -0.165, respectively (and both are highly significant; $p < 0.001$).[24]

# 4 Discussion and Conclusion

Each state (and in many cases each city and county) in the United States responded in their own way to the COVID-19 pandemic. This disorganization led to substantial variation in infections and deaths across both time and geography. Some of this heterogeneity can likely be attributed to differences in the county, state, and city-level lockdown orders and other government-mandated COVID-19 responses. As such, it is tempting to use a difference-in-difference (DiD) approach to estimate the effect of NPIs on COVID-19 infections and deaths. However, there are two main problems with such an approach: i) the policies in question were implemented at different times from different levels of administration (state, county, city), often reflecting different degrees of restriction[25] and ii) even in areas where no wide-spread policy was enacted, many steps were still taken that generate similar results (e.g., school closures or large national firms/manufacturers temporarily shuttering all locations). Furthermore, while one could use DiD with social distancing measures to separate the effect of voluntary compliance from official ordinance, doing so assumes that all encounters are roughly the same in terms of COVID-19 transmission, which our analysis calls into question.

We find, after controlling for other factors, that counties with higher encounter rates have more COIVD-19 infections and deaths than counties with lower encounter rates on average. Overall, this result is compatible with independent contemporaneous works (Friedson et al., 2020; Courtemanche et al., 2020; Dave et al., 2020). However, there is substantial variation in the *effect* of encounters, suggesting that other NPIs (such as face masks) are being more commonly used in some areas and/or there are obstacles preventing testing and delaying treatments. Our results are consistent with observable virus trends in the U.S. since the beginning of the pandemic. Males, Republicans, and African Americans are all less likely to perceive COVID-19 as a serious health risk and therefore less likely to follow public health measures (Khubchandani et al., 2020; Masters et al., 2020; Lehmann and Lehmann, 2020).

Our strongest results are observed when estimating deaths - which is expected considering the drawbacks of using infection data (e.g., depends on testing availability, willingness to be tested, and timing). We find that an increase in the number of encounters (less social distancing) results in more deaths in counties with a larger proportion of residents over age 75, a higher GOP vote in the 2016 presidential election, a larger proportion of Black or African American residents, and a larger proportion of male residents. The effect of encounters tends to be less severe in areas with a high infant mortality rate - suggesting

[24]We find similar results when comparing EJEs to the percent of the county reporting to “never” wear a mask, with $\rho = 0.168$ and $\rho = 0.137$ for EJEs recovered from the infections and deaths regressions, respectively.

[25]One good example of this is the Missouri statewide Stay at Home Order which expired on May 3, 2020; a more restrictive county order, in Jackson County Missouri, did not end until May 11, 2020.

NPIs in addition to social distancing are being used extensively in such areas.[26]

Our results present a cautionary tale for policymakers facing a future public health crisis. While a decentralized policy response may be ideal in principle, our results suggest that dealing with a highly transmittable infectious disease requires a uniform and, arguably, more heavy-handed approach. While policies limiting mobility and social gatherings represent a natural first step, these policies will have little impact if other basic guidelines are not being followed. In a world full of distrust and mixed messaging, even basic recommendations made by public health experts, such as "wear a mask in public", can be labeled as "fake news" and ignored by large segments of the population. This weakens the effectiveness of policies limiting mobility and creates an environment that is ripe for further disinformation.

[26]It is also possible that the low infant mortality rate in rural areas is driving this result. To test this possibility we estimated similar models using obesity, smoking, and diabetes rates and find a similar relationship. Therefore it seems that "low health" counties are adjusting their behaviors in ways that prevent adverse outcomes.

## References

Allcott, Hunt, Levi Boxell, Jacob C Conway, Billy a Ferguson, Matthew Gentzkow, and Benny Goldman (2020a) 'What Explains Temporal and Geographic Variation in the Early US Coronavirus Pandemic?' Technical Report, National Bureau of Economic Research

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Y Yang (2020b) 'Polarization and Public Health: Partisan Differences in Social Distancing During the Coronavirus Pandemic.' *Journal of Public Economics*

Andersen, Martin (2020) 'Early Evidence on Social Distancing in Response to COVID-19 in the United States.' *Available at SSRN 3569368*

Banholzer, Nicolas, Eva van Weenen, Bernhard Kratzwald, Arne Seeliger, Daniel Tschernutter, Pierluigi Bottrighi, Alberto Cenedese, Joan Puig Salles, Stefan Feuerriegel, and Werner Vach (2020) 'Estimating the Impact of Non-Pharmaceutical Interventions on Documented Infections with COVID-19: a Cross-Country Analysis.' *medRxiv*

Béraud, Guillaume, Steven Abrams, Philippe Beutels, Benoit Dervaux, and Niel Hens (2018) 'Resurgence Risk for Measles, Mumps and Rubella in France in 2018 and 2020.' *Eurosurveillance* 23(25), 1700796

Brzezinski, Adam, Valentin Kecht, David Van Dijcke, and Austin L Wright (2020) 'Belief in Science Influences Physical Distancing in Response to Covid-19 Lockdown Policies.' *University of Chicago, Becker Friedman Institute for Economics Working Paper*

Caley, Peter, David J Philp, and Kevin McCracken (2008) 'Quantifying Social Distancing Arising From Pandemic Influenza.' *Journal of the Royal Society Interface* 5(23), 631–639

Centers for Disease Control and Prevention (2019) 'Underlying Cause of Death, 1999-2018'

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin (2019) 'PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects.' *arXiv preprint arXiv:1903.01690*

Courtemanche, Charles J, Joseph Garuccio, Anh Le, Joshua C Pinkston, and Aaron Yelowitz (2020) 'Did Social-Distancing Measures in Kentucky Help to Flatten the COVID-19 Curve?'

Coxe, Stefany, Stephen G West, and Leona S Aiken (2009) 'The Analysis of Count Data: a Gentle Introduction to Poisson Regression and Its Alternatives.' *Journal of Personality Assessment* 91(2), 121–136

Dave, Dhaval M, Andrew I Friedson, Kyutaro Matsuzawa, and Joseph J Sabia (2020) 'When Do Shelter-In-Place Orders Fight Covid-19 Best? Policy Heterogeneity Across States and Adoption Time.' Technical Report, National Bureau of Economic Research

Dimke, Christine, Marissa C Lee, and Jude Bayham (2020) 'Working From a Distance: Who Can Afford to Stay Home During COVID-19? Evidence From Mobile Device Data.' *medRxiv*

Fan, Jingchun, Xiaodong Liu, Weimin Pan, Mark W Douglas, and Shisan Bao (2020) 'Epidemiology of 2019 Novel Coronavirus Disease-19 in Gansu Province, China, 2020.' *Emerging Infectious Diseases*

Fong, Min W, Huizhi Gao, Jessica Y Wong, Jingyi Xiao, Eunice YC Shiu, Sukhyun Ryu, and Benjamin J Cowling (2020) 'Nonpharmaceutical Measures for Pandemic Influenza in Nonhealthcare Settings Social Distancing Measures'

Friedson, Andrew I, Drew McNichols, Joseph J Sabia, and Dhaval Dave (2020) 'Did Californias Shelter-In-Place Order Work? Early Coronavirus-Related Public Health Effects.' Technical Report, National Bureau of Economic Research

Gao, Song, Jinmeng Rao, Yuhao Kang, Yunlei Liang, Jake Kruse, Doerte Doepfer, Ajay K Sethi, Juan Francisco Mandujano Reyes, Jonathan Patz, and Brian S Yandell (2020) 'Mobile Phone Location Data Reveal the Effect and Geographic Variation of Social Distancing on the Spread of the COVID-19 Epidemic.' *arXiv preprint arXiv:2004.11430*

Glaeser, Edward L, Caitlin S Gorback, and Stephen J Redding (2020) 'How Much Does COVID-19 Increase with Mobility? Evidence From New York and Four Other US Cities.' Technical Report, National Bureau of Economic Research

Gollwitzer, Anton, Cameron Martel, Julia Marshall, Johanna Marie Höhs, and John A Bargh (2020) 'Connecting Self-Reported Social Distancing to Real-World Behavior At the Individual and US State Level'

Goolsbee, Austan, and Chad Syverson (2020) 'Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020.' Technical Report, National Bureau of Economic Research

Guzzetta, Giorgio, Flavia Riccardo, Valentina Marziano, Piero Poletti, Filippo Trentini, Antonino Bella, Xanthi Andrianou, Martina Del Manso, Massimo Fabiani, Stefania Bellino et al. (2020) 'The Impact of a Nation-Wide Lockdown on COVID-19 Transmissibility in Italy.' *arXiv preprint arXiv:2004.12338*

Hausman, Jerry a, Bronwyn H Hall, and Zvi Griliches (1984) 'Econometric Models for Count Data with an Application to the Patents-R&D Relationship'

Holtgrave, David R (2010) 'Is the Elimination of HIV Infection within Reach in the United States? Lessons from an Epidemiologic Transmission Model.' *Public Health Reports* 125(3), 372–376

Homaira, N, N Briggs, J-L Oei, L Hilder, B Bajuk, T Snelling, and A Jaffe (2018) 'High Impact of Influenza on Hospitalisation Rates in Children with a Range of Chronic Lung Diseases: a Population-Based Study.' *International Journal of Infectious Diseases* 73, 36

Homaira, Nusrat, Stephen P Luby, William a Petri, Raija Vainionpaa, Mustafizur Rahman, Kamal Hossain, Cynthia B Snider, Mahmudur Rahman, ASM Alamgir, Farzina Zesmin et al. (2012) 'Incidence of Respiratory Virus-Associated Pneumonia in Urban Poor Young Children of Dhaka, Bangladesh, 2009–2011.' *PloS one*

Hutchinson, M Katherine, and Matthew C Holtman (2005) 'Analysis of Count Data Using Poisson Regression.' *Research in nursing & health* 28(5), 408–418

Kelso, Joel K, George J Milne, and Heath Kelly (2009) 'Simulation Suggests That Rapid Activation of Social Distancing Can Arrest Epidemic Development Due to a Novel Strain of Influenza.' *BMC Public Health* 9(1), 117

Khubchandani, Jagdish, Diana Saiki, and Jayanthi Kandiah (2020) 'Masks, Gloves, and the COVID-19 Pandemic: Rapid Assessment of Public Behaviors in the United States.' *Epidemiologia* 1(1), 16–22

Lehmann, Eliyahu Y, and Lisa Soleymani Lehmann (2020) 'Responding to Patients Who Refuse to Wear Masks During the Covid-19 Pandemic.' *Journal of General Internal Medicine* pp. 1–2

Lovett, Andrew, and Robin Flowerdew (1989) 'Analysis of Count Data Using Poisson Regression.' *The Professional Geographer* 41(2), 190–198

Ma, Junling (2020) 'Estimating Epidemic Exponential Growth Rate and Basic Reproduction Number.' *Infectious Disease Modeling*

Mandal, Manisha, and Shyamapada Mandal (2020) 'COVID-19 Pandemic Scenario in India Compared to China and Rest of the World: a Data Driven and Model Analysis.' *medRxiv*

Masters, Nina B, Shu-Fang Shih, Allen Bukoff, Kaitlyn B Akel, Lindsay C Kobayashi, Alison L Miller, Harapan Harapan, Yihan Lu, and Abram L Wagner (2020) 'Social Distancing in Response to the Novel Coronavirus (COVID-19) in the United States.' *PloS one* 15(9), e0239025

MIT Election Data and Science Lab (2018) 'County Presidential Election Returns 2000-2016'

Nasreen, Sharifa, Stephen P Luby, W Abdullah Brooks, Nusrat Homaira, Abdullah Al Mamun, Mejbah Uddin Bhuiyan, Mustafizur Rahman, Dilruba Ahmed, Jaynal Abedin, Mahmudur Rahman et al. (2014) 'Population-Based Incidence of Severe Acute Respiratory Virus Infections Among Children Aged Less Than 5 Years in Rural Bangladesh, June–October 2010.' *PloS one*

Ngo, Madeline 'Rounding Out the Social Distancing Scoreboard.' https://www.unacast.com/post/rounding-out-the-social-distancing-scoreboard. Accessed: 20-10-26

Painter, Marcus, and Tian Qiu (2020) 'Political Beliefs Affect Compliance with Covid-19 Social Distancing Orders.' *Available at SSRN 3569098*

Pan, an, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin et al. (2020) 'Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China.' *Journal of the American Medical Association*

Peak, Corey M, Amy Wesolowski, Elisabeth zu Erbach-Schoenberg, Andrew J Tatem, Erik Wetter, Xin Lu, Daniel Power, Elaine Weidman-Grunewald, Sergio Ramos, Simon Moritz et al. (2018) 'Population Mobility Reductions Associated with Travel Restrictions During the Ebola Epidemic in Sierra Leone: Use of Mobile Phone Data.' *International Journal of Epidemiology* 47(5), 1562–1570

Perrin, Andrew (2019) 'Digital Gap Between Rural and Nonrural America Persists.' Accessed: 2020-11-17

Persico, Claudia, Kathryn R Johnson et al. (2020) 'Deregulation in a Time of Pandemic: Does Pollution Increase Coronavirus Cases Or Deaths?' Technical Report, Institute of Labor Economics (IZA)

Pew Research Center (2019) 'Mobile Fact Sheet.' Accessed: 2020-05-13

SafeGraph INC (2020) 'The Source of Truth for POI Data & Business Listings.' https://www.safegraph.com/. Accessed: 2020-05-13

Sprenger, Florian (2019) 'The Network Is Not the Territory: on Capturing Mobile Media.' *New Media & Society* 21(1), 77–96

The New York Times and Dynata (2020) 'Mask-Wearing Survey Data'

Tillett, Richard L, Joel R Sevinsky, Paul D Hartley, Heather Kerwin, Natalie Crawford, Andrew Gorzalski, Chris Laverdure, Subhash C Verma, Cyprian C Rossetto, David Jackson et al. (2020) 'Genomic Evidence for Reinfection with SARS-Cov-2: a Case Study.' *The Lancet Infectious Diseases*

Unacast (2020) 'Unacast Social Distancing Dataset.' https://www.unacast.com/data-for-good. Accessed: 2020-04-029

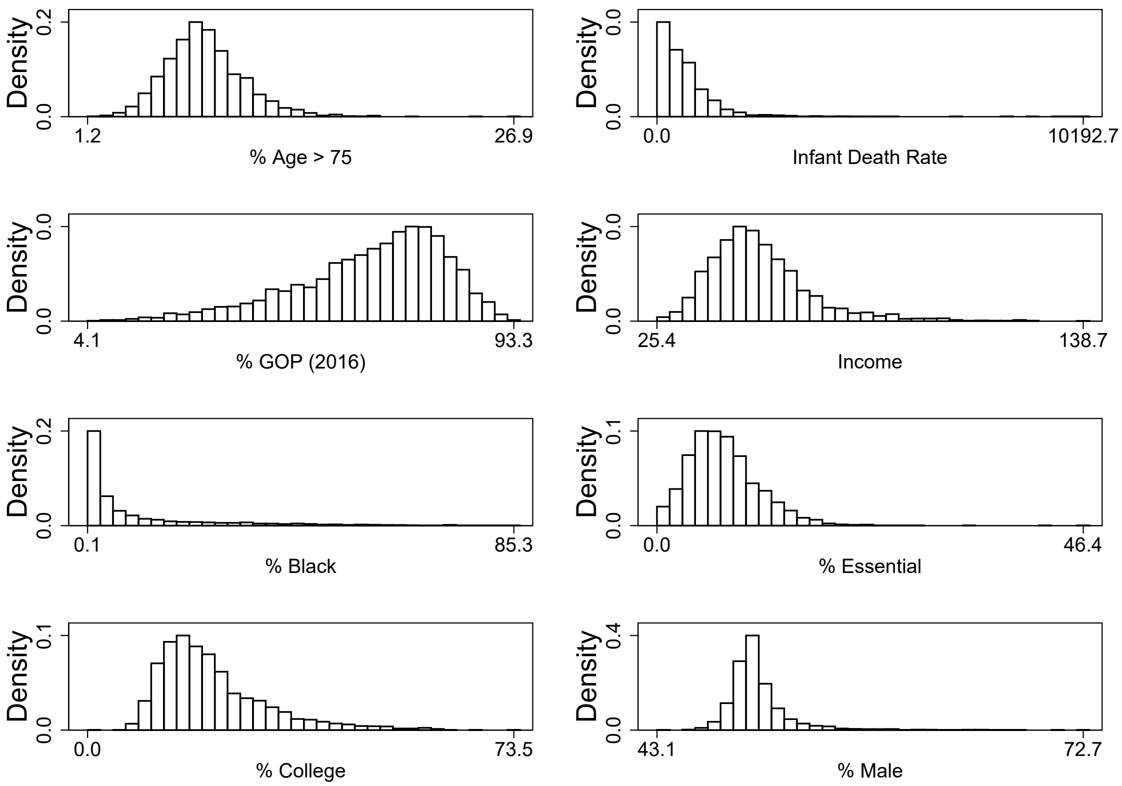
United States Census Bureau (2019a) 'County Business Patterns: 2018'

_ (2019b) 'SAIPE State and County Estimates for 2018'

United States Department of Agriculture (2020) 'County-level Data Sets'

# 5 APPENDIX

Figure 6: Distribution of the Time In-Variant Variables of Interest



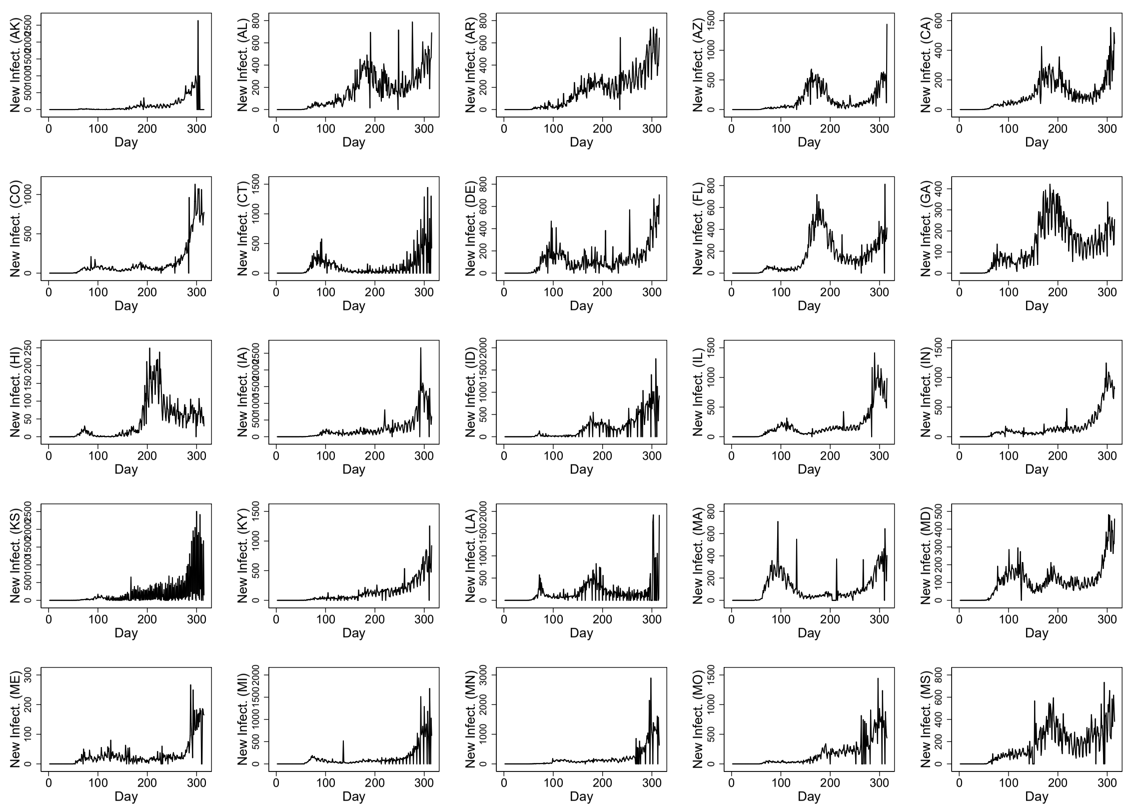

**Notes:** Distribution of the time invariant variables.

Table 5: Correlation of Time-Invariant Variables

| | %Age > 75 | Infant Death Rate | % GOP (2016) | Income | % Black | % Essential | % College | % Male |
|---|---|---|---|---|---|---|---|---|
| %Age > 75 | 1 | | | | | | | |
| Infant Death Rate | -0.00922 | 1 | | | | | | |
| % GOP (2016) | 0.318$^{***}$ | 0.0173 | 1 | | | | | |
| Income | -0.290$^{***}$ | -0.125$^{***}$ | -0.242$^{***}$ | 1 | | | | |
| % Black | -0.220$^{***}$ | 0.114$^{***}$ | -0.417$^{***}$ | -0.238$^{***}$ | 1 | | | |
| % Essential | -0.163$^{***}$ | -0.0345 | -0.363$^{***}$ | 0.287$^{***}$ | 0.0385$^{*}$ | 1 | | |
| % College | -0.209$^{***}$ | -0.102$^{***}$ | -0.489$^{***}$ | 0.719$^{***}$ | -0.0808$^{***}$ | 0.484$^{***}$ | 1 | |
| % Male | -0.0548$^{**}$ | 0.0172 | 0.158$^{***}$ | -0.0442$^{*}$ | -0.136$^{***}$ | -0.239$^{***}$ | -0.184$^{***}$ | 1 |

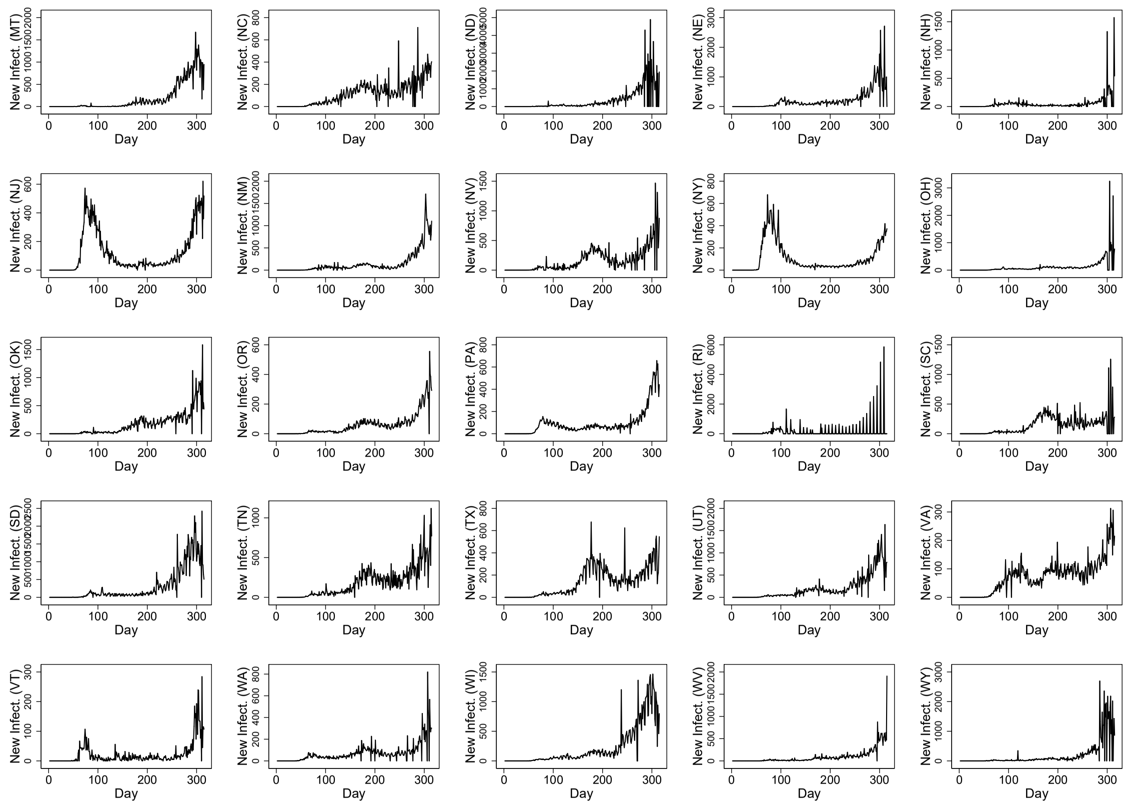
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Figure 7: New Infections by Day and State (AK - MI)



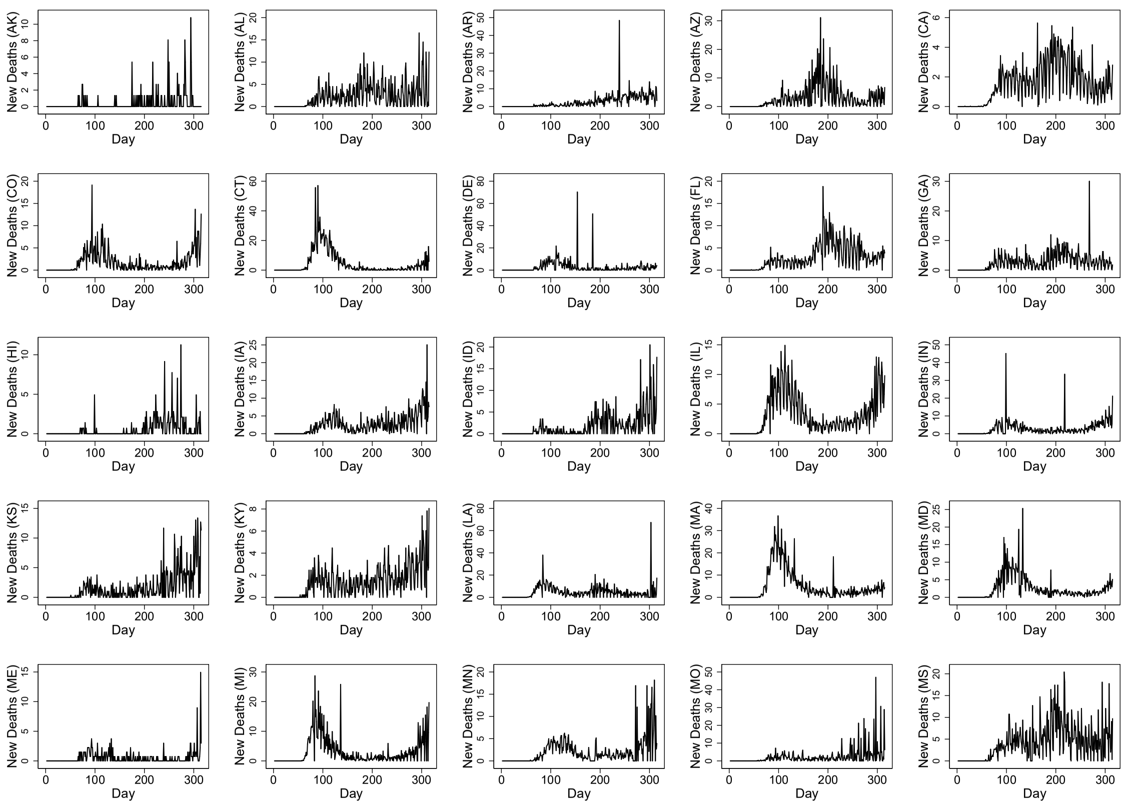
**Notes:** New COVID 19 infections by state and day. From top left to bottom right: Alabama, Alaska, Arkansas, Arizona, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Iowa, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, Michigan, Minnesota, Missouri, and Mississippi.

Figure 8: New Infections by Day and State (MT - WY)



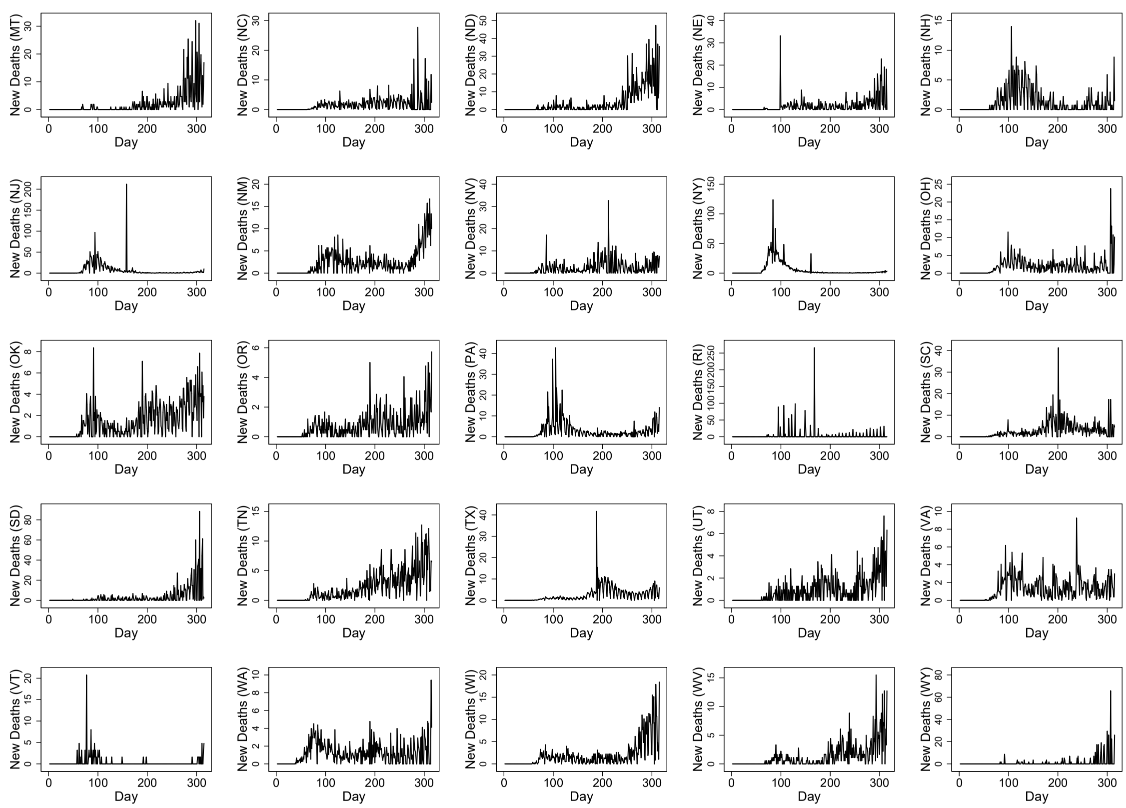

**Notes:** New COVID 19 infections by state and day. From top left to bottom right: Montana, North Carolina, North Dakota, Nebraska, New Hampshire, New Jersey, New Mexico, Nevada, New York, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Vermont, Washington, Wisconsin, West Virginia, and Wyoming.

Figure 9: New Deaths by Day and State (AK - MI)



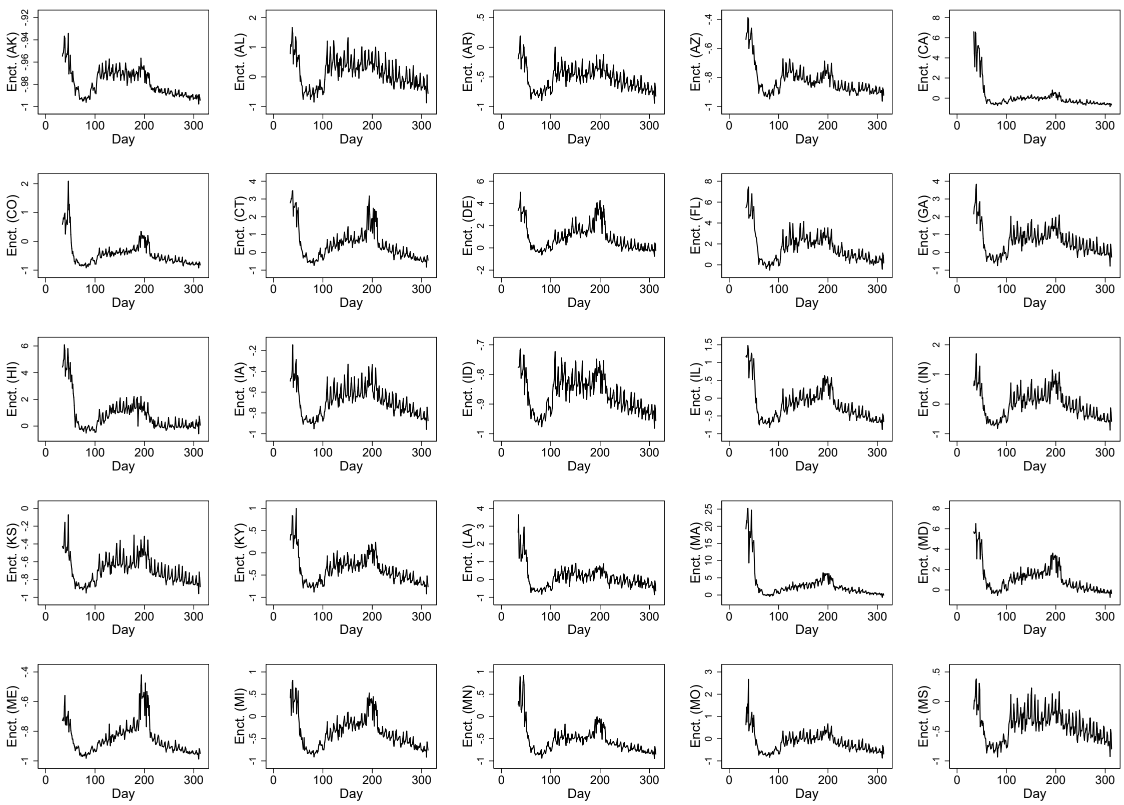

**Notes:** New COVID 19 deaths by state and day. From top left to bottom right: Alabama, Alaska, Arkansas, Arizona, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Iowa, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, Michigan, Minnesota, Missouri, and Mississippi.

Figure 10: New Deaths by Day and State (MT - WY)



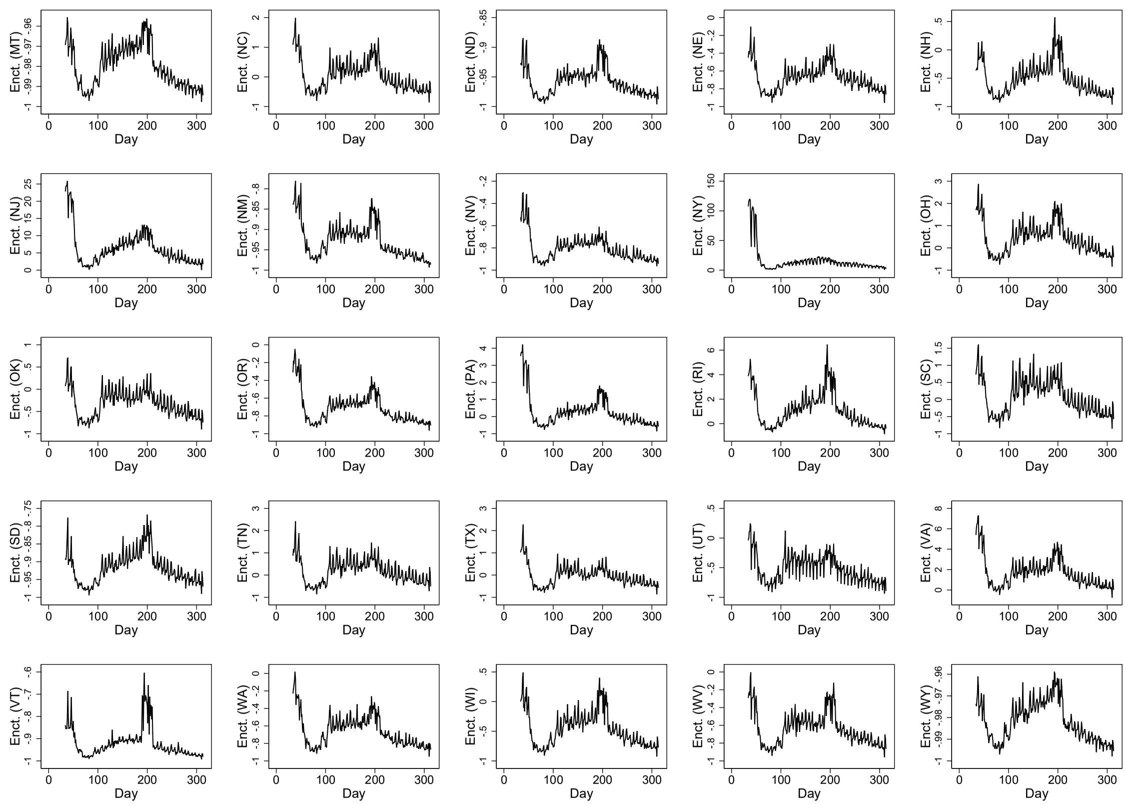

**Notes:** New COVID 19 deaths by state and day. From top left to bottom right: Montana, North Carolina, North Dakota, Nebraska, New Hampshire, New Jersey, New Mexico, Nevada, New York, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Vermont, Washington, Wisconsin, West Virginia, and Wyoming.

Figure 11: Encounter Rates by Day and and State (AK - MI)



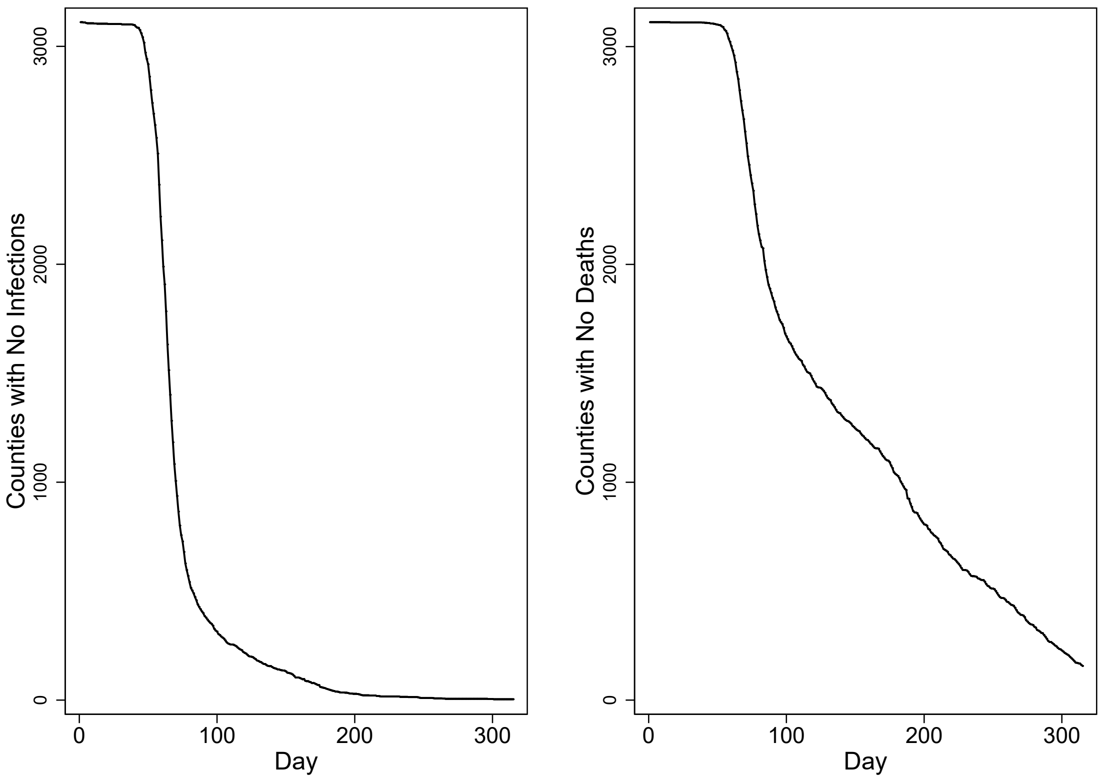
**Notes:** Social distancing by state and day. From top left to bottom right: Alabama, Alaska, Arkansas, Arizona, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Iowa, Idaho, Illinois, Indiana, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, Michigan, Minnesota, Missouri, and Mississippi.

Figure 12: Encounter Rates by Day and State (MT - WY)



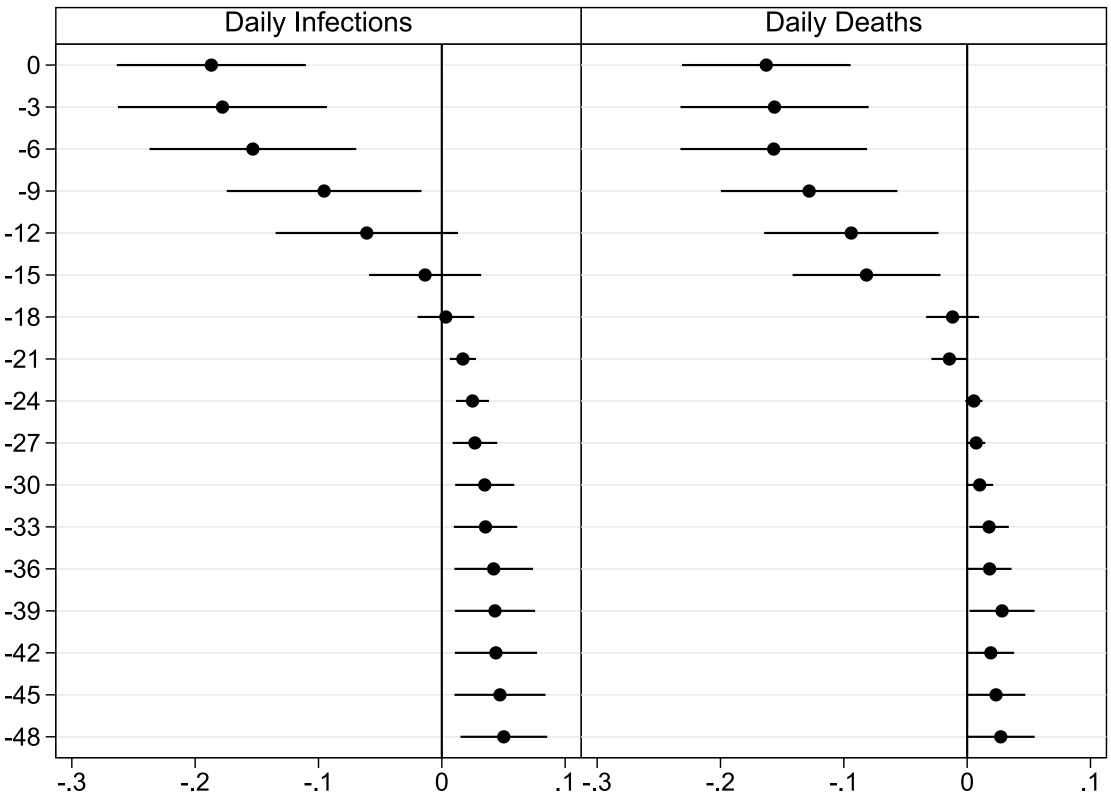

**Notes:** Social distancing by state and day. From top left to bottom right: Montana, North Carolina, North Dakota, Nebraska, New Hampshire, New Jersey, New Mexico, Nevada, New York, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, Vermont, Washington, Wisconsin, West Virginia, and Wyoming.

Figure 13: Number of Infection and Death Free Counties by Day




**Notes:** Number of US counties without any reported infections (left panel) and US counties without any reported deaths (right panel).

Figure 14: Estimated Relationship between Lagged Encounters and Daily COVID-19 Deaths and Infections




**Notes:** Estimated relationship between lagged encounters and daily COVID-19 infections (left panel) and deaths (right panel). Each dot represents a coefficient that is estimated from a model with only a single lagged encounter rate that is x days in the past (where $x \in 0, 3, 6, ...48$) and control variables. These control variables include: county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. Bars correspond to 95 % intervals using clustered standard errors. All models are fixed effects poisson.

Table 6: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | $0.027^{***}$ | 0.089 | $0.038^{***}$ | $0.14^{***}$ |
| | (3.28) | (1.62) | (4.58) | (4.82) |
| ER X % Age > 75 | | 0.0095 | | $0.019^{***}$ |
| | | (0.72) | | (3.08) |
| ER X Infant Death | | -0.00013 | | $-0.00025^{***}$ |
| | | (-1.61) | | (-4.79) |
| ER X Income | | $-0.0099^{***}$ | | -0.00068 |
| | | (-4.57) | | (-0.57) |
| ER X % GOP | | $0.010^{***}$ | | $0.0023^{***}$ |
| | | (4.62) | | (2.88) |
| ER X % Black | | $0.011^{***}$ | | $0.0046^{***}$ |
| | | (4.79) | | (3.09) |
| ER X % Essential | | $-0.0063^{*}$ | | 0.00040 |
| | | (-1.75) | | (0.21) |
| ER X % College | | $0.021^{***}$ | | -0.00089 |
| | | (5.32) | | (-0.50) |
| ER X % Male | | 0.015 | | $0.027^{**}$ |
| | | (0.45) | | (2.39) |
| LL | -583747.9 | -582307.6 | -75219.2 | -75041.2 |
| Counties | 3092 | 3091 | 2830 | 2829 |
| Observations | 465994 | 465857 | 84494 | 84474 |

$t$ statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of the natural log of daily COVID-19 infections (two left most columns)/deaths (two right most columns) per 100,000 and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are linear fixed effects panel regressions. Standard errors are clustered at the county.

Table 7: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | 0.0060 | 0.27$^{***}$ | 0.024$^{**}$ | 0.34$^{***}$ |
| | (1.20) | (3.75) | (2.23) | (6.28) |
| L. Inf. | -0.00000024 | -0.00000059 | -0.0000057$^{**}$ | -0.0000057$^{***}$ |
| | (-0.21) | (-0.49) | (-2.53) | (-2.85) |
| L. Inf. X ER | -0.0000032$^{***}$ | -0.0000031$^{***}$ | -0.0000019$^{***}$ | -0.0000017$^{***}$ |
| | (-5.95) | (-5.77) | (-2.97) | (-3.01) |
| ER X % Age > 75 | | 0.027$^{*}$ | | 0.049$^{***}$ |
| | | (1.65) | | (4.03) |
| ER X Infant Death | | -0.00026$^{**}$ | | -0.00035$^{***}$ |
| | | (-2.13) | | (-4.06) |
| ER X Income | | -0.0069$^{***}$ | | -0.00084 |
| | | (-2.95) | | (-0.73) |
| ER X % GOP | | 0.0072$^{***}$ | | 0.0059$^{***}$ |
| | | (3.20) | | (5.49) |
| ER X % Black | | 0.0080$^{**}$ | | 0.010$^{***}$ |
| | | (2.52) | | (4.53) |
| ER X % Essential | | 0.0079$^{*}$ | | 0.0068$^{***}$ |
| | | (1.78) | | (3.77) |
| ER X % College | | 0.0060 | | -0.0017 |
| | | (1.63) | | (-1.09) |
| ER X % Male | | 0.076$^{**}$ | | 0.093$^{***}$ |
| | | (2.31) | | (4.12) |
| LL | -5859788.0 | -5833748.2 | -370550.9 | -369025.6 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

$t$ statistics in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are fixed effects poisson. Standard errors are clustered at the county. Unlike results reported in Table 3, these estimates control for the include the total number of infections, L. Inf., 21 days in the past in the models estimating infections and 39 days in the past in models estimating deaths as well as the interaction of lagged infections and encounters.

Table 8: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | 0.000256$^{***}$ | 0.0483$^{***}$ | 0.000221$^{***}$ | 0.0369$^{***}$ |
| | (5.10) | (6.25) | (2.85) | (10.74) |
| ER X % Age > 75 | | 0.00670$^{***}$ | | 0.00608$^{***}$ |
| | | (4.47) | | (10.54) |
| ER X Infant Death | | -0.0000299 | | -0.0000284$^{***}$ |
| | | (-1.11) | | (-4.92) |
| ER X Income | | -0.000589$^{***}$ | | -0.000191$^{**}$ |
| | | (-3.81) | | (-2.52) |
| ER X % GOP | | 0.000853$^{***}$ | | 0.000677$^{***}$ |
| | | (3.09) | | (8.52) |
| ER X % Black | | 0.00107$^{**}$ | | 0.00114$^{***}$ |
| | | (2.13) | | (7.13) |
| ER X % Essential | | 0.000839 | | 0.000708$^{***}$ |
| | | (1.57) | | (5.05) |
| ER X % College | | 0.000426$^{*}$ | | 0.0000996 |
| | | (1.93) | | (1.26) |
| ER X % Male | | 0.0127$^{***}$ | | 0.0111$^{***}$ |
| | | (2.71) | | (8.21) |
| LL | -6053138.4 | -6031100.0 | -375508.6 | -373138.8 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

*t* statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are fixed effects poisson. Standard errors are clustered at the county. Unlike results reported in Table 3, encounters are measured using the original encounters rate as calculated by Unacast.

Table 9: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | 0.0733 | $0.411^{**}$ | $0.369^{***}$ | $1.048^{***}$ |
| | (0.95) | (2.15) | (3.08) | (5.67) |
| ER X % Age > 75 | | -0.0722 | | $0.0667^{*}$ |
| | | (-1.54) | | (1.67) |
| ER X Infant Death | | -0.000738 | | $-0.00128^{***}$ |
| | | (-1.53) | | (-3.06) |
| ER X Income | | $-0.0315^{***}$ | | 0.000644 |
| | | (-3.47) | | (0.11) |
| ER X % GOP | | $0.0279^{***}$ | | $0.0135^{**}$ |
| | | (3.08) | | (2.32) |
| ER X % Black | | $0.0211^{**}$ | | $0.0233^{***}$ |
| | | (2.06) | | (2.96) |
| ER X % Essential | | 0.0174 | | 0.0145 |
| | | (0.65) | | (1.05) |
| ER X % College | | $0.0330^{**}$ | | $-0.0191^{*}$ |
| | | (2.11) | | (-1.87) |
| ER X % Male | | 0.0603 | | $0.130^{*}$ |
| | | (0.82) | | (1.92) |
| LL | -6055235.9 | -6011975.1 | -374724.6 | -373410.2 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

$t$ statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are fixed effects poisson. Standard errors are clustered at the county. Unlike results reported in Table 3, “Encounter Rate” is Uncast’s Encounter Rate + 1 raised to the .25 power.

Table 10: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | 0.00221*** | 0.139*** | 0.00245** | 0.127*** |
| | (4.46) | (6.23) | (2.24) | (9.06) |
| ER X % Age > 75 | | 0.0157*** | | 0.0190*** |
| | | (3.34) | | (6.84) |
| ER X Infant Death | | -0.000108 | | -0.0000998*** |
| | | (-1.34) | | (-3.82) |
| ER X Income | | -0.00223*** | | -0.000643** |
| | | (-3.86) | | (-2.37) |
| ER X % GOP | | 0.00281*** | | 0.00221*** |
| | | (3.02) | | (6.90) |
| ER X % Black | | 0.00320** | | 0.00359*** |
| | | (2.19) | | (5.75) |
| ER X % Essential | | 0.00304 | | 0.00216*** |
| | | (1.59) | | (4.09) |
| ER X % College | | 0.00162* | | 0.000180 |
| | | (1.65) | | (0.58) |
| ER X % Male | | 0.0379*** | | 0.0356*** |
| | | (2.83) | | (5.96) |
| LL | -6053137.4 | -6025799.8 | -375326.4 | -373023.3 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

$t$ statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are fixed effects poisson. Standard errors are clustered at the county. Unlike results reported in Table 3,“Encounter Rate” is Uncast’s Encounter Rate + 1 raised to the .75 power.

Table 11: The Effect of Social Distancing on Covid 19 Infections and Deaths

| | Infections | Infections | Deaths | Deaths |
|---|---|---|---|---|
| Encounter Rate | 0.000609*** | 0.0761*** | 0.000574** | 0.0615*** |
| | (4.85) | (6.51) | (2.55) | (10.27) |
| ER X % Age > 75 | | 0.00994*** | | 0.00981*** |
| | | (4.25) | | (9.15) |
| ER X Infant Death | | -0.0000511 | | -0.0000471*** |
| | | (-1.19) | | (-4.48) |
| ER X Income | | -0.00101*** | | -0.000324*** |
| | | (-3.97) | | (-2.65) |
| ER X % GOP | | 0.00139*** | | 0.00111*** |
| | | (3.11) | | (7.98) |
| ER X % Black | | 0.00169** | | 0.00183*** |
| | | (2.16) | | (6.76) |
| ER X % Essential | | 0.00142 | | 0.00112*** |
| | | (1.59) | | (4.79) |
| ER X % College | | 0.000727* | | 0.000153 |
| | | (1.84) | | (1.18) |
| ER X % Male | | 0.0204*** | | 0.0180*** |
| | | (2.80) | | (7.47) |
| LL | -6053105.4 | -6029096.3 | -375449.4 | -373076.9 |
| Counties | 3092 | 3091 | 2975 | 2974 |
| Observations | 729062 | 728821 | 680456 | 680215 |

$t$ statistics in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

**Notes:** Estimated association of infections (two left most columns)/deaths (two right most columns) and encounters and the interaction of encounters with county characteristics. Other control variables include county specific fixed effects, week fixed effects, day of the week effects, and state specific time trends. All models are fixed effects poisson. Standard errors are clustered at the county. Unlike results reported in Table 3, "Encounter Rate" is Uncast's Encounter Rate + 1 raised to the .9 power.