Disentangling the Immediate Effect of Timing of Stay-home Orders and Mobility Decrease in US Counties

Michelle Audirac
audiramichelle@gmail.com

Mauricio Tec
The University of Texas at Austin
mauriciogtec@utexas.edu

Cory Zigler
The University of Texas at Austin
cory.zigler@austin.utexas.edu

Abstract

After the first wave of state and local interventions that were imposed in the US to contain the spread of covid-19, the literature on the effectiveness of such measures has been evolving rapidly with early studies measuring the state-level effect of mobility trends on cases/deaths as a proxy of the impact of government policies. Subsequent studies identified that not all mobility reductions could be attributed to the policy interventions and recognize the need and the challenges of decoupling these factors. In this article we use a Bayesian hierarchical model to study the interaction between the trend shift in disease evolution attributed to the stay-at-home orders and to mobility decrease at a county-level. By capturing county specific characteristics in the pre-intervention trajectories, and comparing the magnitude of the relative effectiveness of the timing of both measures in the post-intervention trends, we conclude that the timing of mobility reductions was more important than the timing of official stay-at-home orders in bending the daily deaths curves of the more urban counties. Our work distinguishes from others for incorporating county specific data that capture their heterogeneity, by presenting results for an array of county types with varying degrees of urbanicity, and for its use of a methodology that echos to an extent epidemiological principles unlike traditional DiD event studies.

1 Introduction

Efforts to quantify the effectiveness of the non-pharmaceutical interventions (NPIs) that were implemented to contain the evolving COVID-19 epidemic in the US must confront both the vast heterogeneity in how the epidemic has unfolded across different areas and the difficulties in decoupling the effects of explicit policy actions and coincident behavioral changes. Social and demographic characteristics generated substantial variability in epidemic growth and trajectory, complicating inferences regarding the effectiveness of measures for controlling the epidemic. In addition, the timing of these measures - be them official NPIs such as stay-at-home orders or changes in behavior - led to further heterogeneity in epidemic control; even though most NPIs occurred within a constrained window of calendar time between mid March and early April, the disparate epidemic introduction across the country generated substantial heterogeneity in the "epidemic timing" of NPIs relative to local epidemic conditions. Finally, the introduction of NPIs frequently coincided - albeit imprefectly - with behavior changes such as mobility reductions, making it difficult to disentangle policy impacts from epidemic changes attributable to coincident changes in behavior. Shedding light on the role that each of these mechanisms had in mitigating the spread of the epidemic in counties across the country can yield an improved understanding to support future intervention strategies.

We endeavor to quantify some of the heterogeneity in policy and behavioral response to the COVID-19 epidemic at a county-level, in particular that related to the timing of intervention relative to local epidemic conditions. Importantly, we permit two different definitions of "intervention": 1) a "policy intervention" that corresponds to the date of an official state- or local stay-at-home order, and 2) a "mobility intervention" that corresponds to reductions of 50% in mobility measured with the number of total visits to various points of interest, specifically: schools, colleges, restaurants, bars, museums and parks. Not only do such intervention definitions correspond to different scientific questions of interest, but, as we show, this difference has important implications with regard to the study of intervention timing, with evident misalignment between the dates of policy interventions and changes in mobility behavior. For both intervention definitions, we use a statistical model of epidemic growth fit to data from individual counties across the US.

Our work follows others' in focusing on county-level analysis of the COVID-19 epidemic in the US. Khan et al. [2020] establishes associations between county-level characteristics and COVID-19 deaths or cases with a latent class model (LCA) to extract subgroups of counties based on socio-demographic, environmental and healthcare data, and a linear regression model to find association between latent classes and COVID-19 mortality adjusted by days since first case. Desmet & Wacziarg [2020] use a series of cross-sectional linear regression coefficients to document in particular the impact of factors related to population density such as urbanicity and modes of transportation.These characterizations of heterogeneity do not explicitly consider the extent in which the timing of NPIs influenced local epidemic growth trajectories.

Other existing work shares important points of contact with ours for its specific focus on the effectiveness and timing of stay-at-home orders. Jinjarak et al. [2020] conducted a country-level analysis concluding that countries with stricter policies were better at controlling the first wave of the disease. In the US, Dave et al. [2020] adopt a state-level differences-in-differences (DiD) methodology and find large effects of stay-at-home orders on reducing deaths/cases, especially among early adopters. Similar approaches have explicitly modeled interactions between the effects of stay-at-home orders and actual observed mobility, with Lin & Meissner [2020] finding evidence that stay-at-home orders resulted on reduced mobility but not reduced COVID-19 cases when using a matching methodology along neighboring counties which implemented different policies, and Kapoor et al. [2020] showing that strong rainfall prior to the enactment of a stay-a-home order showed a reduction in deaths/cases. Admittedly, purely statistical models that utilize the strength of the DiD design or event studies must balance their ability to capture epidemic dynamics when anchored to cross-sectional linear trends only; alternatively, several other researchers have used SEIR-type epidemic models to evaluate intervention impacts. These approaches rely on a more detailed specification of underlying disease mechanics, but require a variety of assumptions and input sources of information[1] . Dehning et al. [2020] combine a Bayesian change-point methodology with SEIR models to detect the points where the benefits of stay-at-home orders in Germany become apparent; another example is due to Pei et al. [2020] who also use a SEIR model and find evidence supporting a strong effect of the intervention timing on the number of deaths in metropolitan areas.

Even with many existing results showing uncertainty intervals of averted cases/deaths or of waning reproduction numbers validate that either government imposed measures or social distancing limited the spread of disease, there remains no consensus about which of such measures better captures the effectiveness for impacting epidemic growth. Unwin et al. [2020] uses changes in mobility as a proxy for the impact of NPIs, while other authors (Abouk & Heydari [2020], Courtemanche et al. [2020]) point out the challenges of decoupling the impact of interventions and coincide on the need to unwind mobility from policy measures to better understand the difference in their effects.

We offer a statistical model of county-level time series of COVID-19 deaths that specifies a negative binomial log-linear trend for the 7-day moving average death rate in which each county is permitted its own quadratic time trend via county-specific random effects that capture heterogeneity above and beyond that explained by measured covariates and echo the epidemiological principle that features of infectious disease evolution such as reproduction number and death rate are, to an extent, county-specific [Desmet & Wacziarg, 2020]. We deploy the statistical model among 440 counties during the period of mid March to mid May to address the primary hypothesis of whether differences in the timing of the US's first wave of NPIs were associated with different levels of effectiveness epidemic control during the first wave of US interventions. We conduct parallel analyses for both the policy-

[1] **From Cory: Need to refine this point on what SEIR models are less good at**

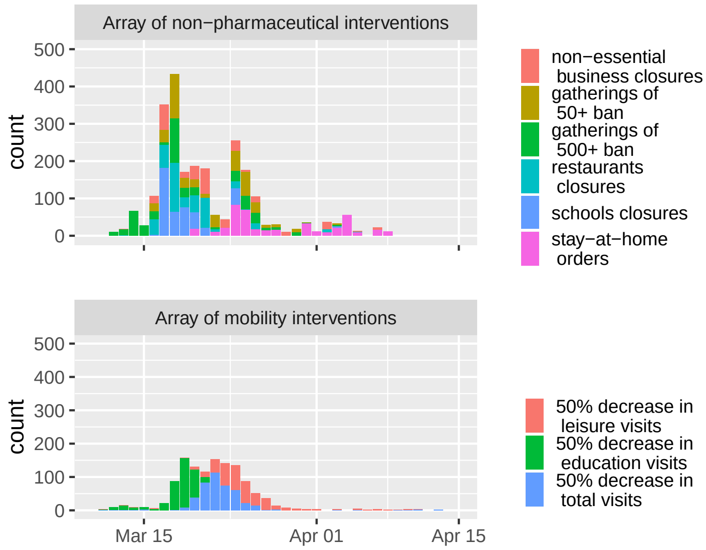


Figure 1: Adoption timeline of an array of both policy and mobility interventions. On the top, counts of dates for six different NPIs, such as stay-home orders, school closures and other more lenient measures, are stacked. On the bottom, counts of days when total visits reached a 50% decrease are included. To illustrate the components that make up total visits, the days "education" and "leisure" visits decreased 50% are also shown; with the education classification covering schools and colleges, and leisure covering restaurants, bars, parks and museums.

and mobility-interventions and compare inferences. The models devote particular care to allow certain county demographic characteristics to dictate local pre- and post-intervention trends over the 7-day moving average of COVID-19 deaths, corresponding to statistical interactions with, for example, the NCHS Urban-Rural Classification Scheme for Counties, a six-category classification meant to capture essential features of urbanicity beyond just population density. The basic structure of the model specification with a quadratic time function has been shown to reasonably forecast deaths during the time frame of study with uncertainty [Woody et al., 2020], particularly for the first stages of the US epidemic. Changes in trends according to the timing of stay-at-home interventions are modeled to depend on the timing of the intervention relative to the local epidemic growth of the county, characterizing the extent to which earlier versus later action in "epidemic time" (vs. calendar time) may have altered the course of a county's epidemic growth. The simple quadratic time function also offers the benefit of readily-available and interpretable summaries of epidemic trajectory such as the time until peak death daily death rate and the number of daily deaths at peak related to certain county characteristics.

2 Data Sources

2.1 County-Level COVID Death and Demographic Data

Our analysis relies on three distinct data sources. The first, which is itself a combination of data sources compiled and reported by Killeen et al. [2020], comprises data on socioeconomic factors that may affect the spread of epidemiological outbreaks, along with confirmed COVID-19 deaths at the county level from the COVID-19 Data Repository compiled by the the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

Also comprised in this dataset are the dates in which six different interventions were put in place in each county: bans of 500+ and 50+ people gatherings; closure mandates of schools, non-essential businesses and restaurants; and orders to stay at home. Intervention policies were most-often put in place state wide, but there are cases in which counties adopted such measures before their respective states. Figure 1 shows the number of counties that adopted a policy in a given day. Most bans and

	Mobility-intervention		
Policy-intervention	Yes	No	Total
Yes	336	84	420
No	20	0	20
Total	356	84	440

Table 1: Counties in our final analysis data sets having either a policy-intervention, a mobility-intervention, or both.

closures occurred in the same small window of time and were preceded by stay-at-home orders, which lagged other policies by 8 days in average. No statewide official orders to shelter-in-place were issued in North Dakota, Nebraska and Arkansas.

We augment this county-level data set with data from the National Center for Health Statistics (NCHS) on the NCHS Urban-Rural Classification Scheme for Counties [Dd & Sj, 2012], which classifies each US county to be in one of the following six categories: 1) large central metro; 2) large fringe metro; 3) medium metro; 4) small metro; 5) micropolitan; 6) non-core.

2.2 Human Mobility Data from SafeGraph

The final data source comes from SafeGraph, a company that provides anonymized population mobility datasets representing 45 million smartphone devices. SafeGraph data aggregates visit counts to numerous points of interest (POIs) classified into categories.

As a proxy measure for overall mobility behavior in each county, we extracted data on the number of visits per day to POIs; in particular to schools, colleges, restaurants, bars, parks and museums, and obtained time series of daily total visits for each county. Although Safegraph provides visits data for several other types of POIs, as well as data on the number of minutes devices remain at home or at work, we limited our analysis only to those categories that had a comprehensive coverage across all the range of counties. The average total visits per day between Jan 15, 2020 and Feb 15, 2020 was used to establish county-baseline levels of mobility. We define the date of the mobility-based intervention to be the date on which the right-aligned ten-day moving average of total visits decreased 50% relative to baseline.

2.3 Analysis Data Set

County-specific daily counts of COVID19 cases/deaths were extracted from Jan 21, 2020 through May 10, 2020. During that period, 1802 counties, from a total of 3221, had reached a threshold of 3 deaths per 10 million residents. To smooth day of week reporting inconsistencies and other sources of noise, rolling averages of seven days were applied over daily deaths and counties having less than 5 cumulative deaths were dropped. Finally, we limited our window of observations after an intervention for each county to 29 days since we are interested in the short-term impact of interventions.

The final analysis consists of two data sets: one containing the 420 counties that put a stay-at-home order in place regardless of their mobility patterns, and the other with the 356 counties whose mobility data reflect a 50% decrease in total visits regardless of whether a stay-home order was put in place. In total, these two datasets comprised data on 440 distinct counties, with 336 having both policy and mobility interventions, and 84 (20) having only a policy (mobility) intervention (Table 1).

It is well established that there is a strong positive relationship between the reproduction number and any factor having the potential to influence the contact rate, such as population density, frequent modes of travel (transit, personal car, pedestrian), distance to major airports, and other factors expected to vary across the spectrum of rural and urban areas. We use the NCHS urban-rural classification categories as a proxy for such county features.

We also include other covariates that have been reported to have a strong relationship with COVID-19 death rate. The percent of black residents and percent of hispanic residents are included to account for the apparent disparities between infection and comorbidity and death rates among these populations relative to other races and ethnicities. To account for the age-related risk of death, we include the percentage of residents that are 65 years or older. We also include the percentage of residents attending college, as students are young people whose main activity became completely remote throughout the epidemic in contrast with others of the same age whose activities might require them to remain exposed. Differences among these features were not captured by the rural-urban NCHS classification.

3 Methods

3.1 Statistical Model

The definition of death trajectories relies on two important features. First, we define local "epidemic time" in terms of the number of days elapsed since local arrival of the epidemic, defined as the date at which a county reached a deaths threshold of 3 deaths per 10 million residents. Second, because reports of daily deaths exhibit great variation across counties (e.g., one county may report most cases on Monday while another county on Wednesday), we fit the statistical models described below to the 7-day centered rolling average of daily deaths, effectively removing day-of-week reporting effects which are not important for the purpose of this analysis.

To model the death trajectories, we use a Bayesian hierarchical model for the time series of each county's death trajectory. The first level of the hierarchy is a log-linear model expressing the expected number of deaths using a negative binomial distribution and a polynomial function of time that changes after the introduction of an intervention. The second level models latent county heterogeneity with county-specific random effects for each degree of the polynomial time function.

More precisely, let y_{it} be the 7-day moving average of daily deaths observed in county i at time t, where t represents the number of days since the deaths threshold was reached. First, we assume

$$y_{it} \sim \mathrm{NegBin}(\lambda_{it}, r), \tag{1}$$

where the parameterization is such that $\mathbb{E}[y_{it}] = \lambda_{it}$ and $\mathbb{V}[y_{it}] = \lambda_{it}(1 + \lambda_{it}/r)$; the unknown parameter r indicates over-dispersion: the smaller it is, the higher the variance. For λ_{it} we use a log link and a per capita normalization. Let N_i be the population of county i, then

$$\log\left(\frac{\lambda_{it}}{N_i}\right) = f(\boldsymbol{x}_i, d_i, t) + r_i(t)$$

where t is the number of days since the threshold date, $\boldsymbol{x}_i$ are the county-level features, and d_i represents intervention timing as the number of days between the threshold date and the date the intervention was enacted. r_i is a county-specific residual time function that captures the portion of the curve not explainable by the county-level features and intervention contained in f.

To specify the form of f we use a quadratic polynomial in t that changes shape after introduction of the intervention and across different levels of county-level features. Specifically,

$$\begin{aligned} f(\boldsymbol{x}_i, d_i, t) := &\sum_{k=0}^{2} \left[\alpha_k + \boldsymbol{x}_{i,pre}^\top \boldsymbol{\beta}_k\right] t^k \\ &+ \sum_{k=1}^{2} \left[\eta_k^I + d_i \eta_k^d + \boldsymbol{x}_{i,post}^\top \boldsymbol{\eta}_k^c\right] \mathbb{1}(t \geq d_i)(t - d_i + 1)^k. \end{aligned} \tag{2}$$

The first summation in (2) specifies the polynomial time trend during the pre-intervention period: $(\alpha_0, \alpha_1, \alpha_2)$ can be thought of as "baseline" intercept, linear, and quadratic polynomial parameters, which are each shifted by $(\beta_0, \beta_1, \beta_2)$ according to county-level features in $\boldsymbol{x}_{i,pre}$. That is, each county's pre-intervention deaths trajectory is dictated in part by the features of that county, where the features contained in $\boldsymbol{x}_{i,pre}$ are: NCHS classification, the percent of college attendees, the percent aged 65 and older, the percent of black, and the percent hispanic of the county.

The second summation in (2) specifies the bend in the quadratic death trajectory after the introduction of the intervention. Here, d_i represents the date that an intervention occurred in county i plus a lag of 12 days[2]. The term $\mathbb{1}(t \geq d_i)$ represents the indicator equal to 1 when epidemic time surpasses the time after the intervention lag, and $(t - d_i + 1)$ is the days elapsed since the intervention lag. The parameters η_k^I for $k = 1, 2$ can be thought of (respectively) as the "baseline" shift in the linear, quadratic terms of the polynomial after the introduction of the intervention. Analogously η_k^d can be thought as shifts in the linear and quadratic terms according to the timing of the intervention, implying different post-intervention trajectory shapes for counties that intervened at different points in their local epidemic time. Similarly, η_k^c dictate shifts in the linear and quadratic terms according to the covariates in $\boldsymbol{x}_{i,post}$, where $\boldsymbol{x}_{i,post}$ in the present analysis includes only the NCHS county classification.

Note the omission of the zero-degree polynomial terms in the post-intervention terms of the second summation in (2), intentionally omitting mean shifts in the polynomial function upon the introduction of the intervention. This is done to specify continuity in the polynomial trend across the introduction of the intervention, as we would not expect the intervention to initiate an immediate jump or drop in daily deaths.

Finally, the idiosyncratic factor r_i is specified using county-specific random effects [Laird & Ware, 1982] for each polynomial coefficient

$$r_i(t) := \sum_{k=0}^{2} \xi_k^{(i)} t^k, \quad (\xi_0^{(i)}, \xi_1^{(i)}, \xi_2^{(i)})^\top \sim N(\mathbf{0}, \Sigma),$$

implying that each county is modeled to have it's own pre-intervention polynomial time trend.

In the above specification, the use of quadratic time polynomials was chosen to parsimoniously capture the nonlinearity of daily death curves with a small enough number of parameters to explore the interactions entailed in (2). Furthermore, a related quadratic trend model was shown to be useful in forecasting the first wave of the disease in US metropolitan areas [Woody et al., 2020]. The explicit modeling of nonlinear trends in daily deaths is an important distinction with other approaches (e.g., difference-in-difference models [Imbens & Rubin, 2015]) that rely on linear approximations to trends over time.

We fit three variations of model (equation 2):

1. *Stay-at-home model*: Here, d_i refers to the days elapsed between a county's death threshold and the day that county instituted a stay-at-home order plus a 12-day lag.
2. *Mobility model*: Here, d_i refers to the days elapsed between a county's death threshold and the day that county reached a 50% decrease in total visits plus a 12-day lag.
3. *Double intervention model*: This model has a three-fold specification, where equation (2) is augmented with a third summation that has the same construction as the second summation. The second term specifies the bend after the mobility intervention, and the third term captures the additional or residual effect of the stay-at-home order that is not captured by the change in mobility.

Note that, since model (2) is specified to borrow information across all counties, results for a given county may be different in all three of the above models, even if the mobility and stay-at-home intervention dates coincide. We fit all models using the `R` language (`3.6.3`) with the package `rstanarm` (`2.19.3`) [Goodrich et al., 2018]. After fitting the models, we base inference on posterior simulations from the models. To evaluate impacts of intervention timing, we simulate posterior predicted deaths from each county's time polynomial, but where the intervention timing, d_i, is replaced with alternative hypothetical timing corresponding to earlier or later intervention. Specifically, under different values of d_i, we take a sample from the Bayesian posterior distribution for each county and take group averages by NCHS for each time t. In addition, we compute the number of days since the threshold at the peak and the number of per capita deaths at the peak for each one of the average curves per NCHS. We repeat this process for 1,000 posterior samples to

[2] A delay consistent with the first quartile of the distribution of time between infection and death. [Lauer et al., 2020] estimate the first quartile of time between infection to symptoms to be 3.8 days, while [Yang et al., 2020] estimate the first quartile of days between symptoms to death at 10. By adding both we obtain an approximation of the left tail of the distribution of time lag of deaths. A similar calculation is done by [Wilson et al., 2020] using IQR values.

produce a set of 1,000 aggregate statistics and curves for each NCHS. The results and further details are presented in section 4.

4 Results

4.1 Descriptions of Intervention Timing

Amid the initial spread of COVID-19, an array of non-pharmaceutical interventions including bans on gatherings, closures of schools and restaurants, and stay-at-home orders were implemented from mid-March to early April across the US. In most cases, these interventions applied to entire states, which further limited the variability in calendar timing across policies across metro, micropolitan and non-core counties. Among the 420 counties in the present analysis, the average date for bans and closures was March 19 and for stay-home orders it was March 28. Figure 1 shows the distribution of dates on which counties implemented various policy NPIs, including the stay-at-home orders that are the focus of the present analysis.

These relatively uniform implementation of policies was echoed by an equally uniform calendar timing of mobility changes in rural and urban counties. Most of the 356 counties with mobility interventions exhibited a steady decrease in mobility within a window of ten days, with March 22 being the average date at which counties reached the 50% decrease from baseline visits to all POIs, corresponding to the definition of a mobility-based intervention. Figure 1, shows the distribution of dates on which counties mobility-intervention, depicted alongside the dates that counties achieved a 50% decrease in visits to two sub-categories of POIs: 1) schools and colleges and 2) leisure destinations such as restaurants, bars, parks and museums.

Despite the relatively uniform calendar timing of policy and mobility interventions, the " epidemic timing" of these interventions relative to the date at which a county exhibited a death threshold of 3 per 10 million residents varied considerably. Figure 2 summarizes the epidemic timing of intervention for each NCHS group. The left panel (Figure 2a) shows the distribution of dates at which counties in each NCHS category reached the threshold of 3 deaths per 10 million, establishing the local arrival of the epidemic. Death thresholds were reached sooner in large central metro areas, followed by a similar timing distribution among large fringe metro areas, and medium/small metro areas, with micropolitan and non-core areas showing the latest distribution of epidemic arrival. Given the relatively homogeneous calendar timing of interventions, this led to the opposite ordering in epidemic timing of interventions, with large central metro areas intervening later in both policy and mobility interventions (Figures 2b,c). Note from Figure 2 that the distribution of mobility intervention timing is shifted approximately earlier than that of policy interventions, almost always occurring before a county even reached the 3/10million death thresholds in all NCHS categories except large central metro counties. Indeed, the median number of days between the policy-intervention and the threshold was 0, both lagging the mobility-intervention by 5 days as summarized in Table 3.

To further illustrate how the timing of mobility interventions compared to that of policy interventions among the 336 counties with both, Figure 3 depicts the decrease in mobility (relative to baseline) that was observed on the day a stay-at-home order took effect. All counties had already shown a marked reduction, with micropolitan and non-core counties showing a reduction of 66% on the day of the policy, and other more urban counties having even more reduced mobility, with an average reduction of 70% mobility when the a stay-at-home order took effect. Thus, all counties indicate the potential for behavior and mobility changes to have impacted epidemic dynamics in advance of official stay-home.

Along with this, reductions in mobility were close to a level of saturation when stay-home went into effect with total visits to POI not reducing much more. The distribution on the right panel in Figure 3 shows that for 80% of the counties in our dataset, the mobility levels observed the day stay-home stay-home orders were adopted, were at most 5% above their absolute minimum levels. After the minimum levels were reached, mobility slowly started to move upwards with 80% of counties lingering no more than 20% above those levels on May 1st.

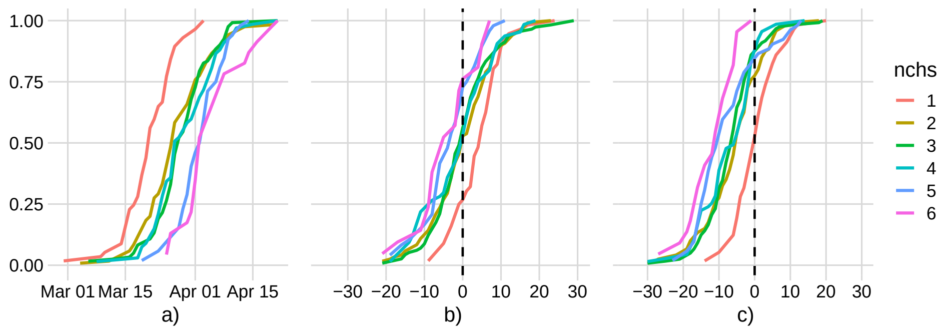


Figure 2: The introduction of covid varied across NCHS groups, leading to differences in the timing of interventions relative to the local epidemic progression. Grid contains the distribution of: a) death-thresholds in calendar timing, b) policy-interventions in epidemic timing where 0 is the day the death-threshold is reached, c) mobility-interventions in epidemic timing. For instance, 25% (75%) of counties in NCHS 1 (6) implemented stay-home before hitting the death threshold, while 50% (100%) in NCHS 1 (6) had already achieved a mobility reduction of 50%.

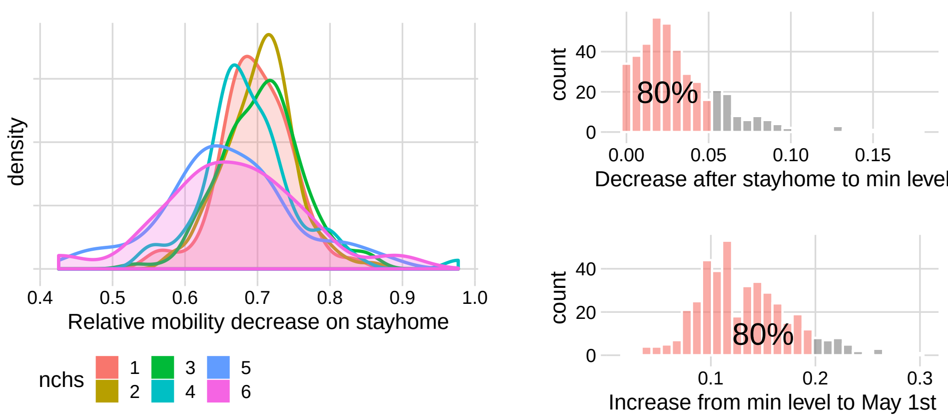


Figure 3: Left: Shows the relative decrease in mobility with respect to the baseline on the day the stay-at-home order was enacted. Right top: Shows the additional decrease in mobility since stay-home up to each county's minimum mobility level. Right bottom: Contains the distribution of the mobility increase beginning at each county's minimum levet up to May 1st.

4.2 Drivers of Pre-intervention Trajectory Heterogeneity

With the rationale that different county characteristics are associated with differently-shaped local epidemic death curves, the model from Section 3.1 permits individual counties to have pre-intervention daily death trajectories that explicitly depend on the county-level covariates: NCHS Urban-Rural Classification Scheme for Counties and percent of residents that are Black, Hispanic, aged 65 years and older, and attending college. Note that given the specification of the model, the differences that are due to the timing of the intervention influence are captured separately.

For both the stay-at-home and mobility models, we find that NCHS category, percent of college student residents, and percent of Hispanic residents had significant impact on death trajectories. In the double intervention model, these same characteristics plus the percent of Black residents had significant impact. A table presenting the coefficient estimates associated with each county characteristic and its interaction with the linear and quadratic time polynomial terms appears in Table 4 in the Appendix. Note that the coefficient estimates in the table are not easily interpreted nor directly comparable in magnitude across the three models since the time polynomial in each model is orthogonalized using the `poly` function in `R`, but judgments of statistical significance remain valid. Figure 7 summarizes the covariate impacts across the different models. The figure shows posterior estimates and uncertainty intervals of the interaction between time polynomials and coefficients effects. More precisely, for each covariate j with corresponding polynomial coefficients $\boldsymbol{\beta}^j = (\beta^j_0, \beta^j_1, \beta^j_2)$ in the regression model (2), we take 1,000 samples from their posterior distribution and for each sample compute an estimate of the total contribution of covariate j through time given by $c^j(t) = \sum_{k=0} t^k \beta^j_k$. Figure 7 is showing the pointwise median and quantiles at each time t of the posterior samples of the total contribution $c^j(t)$. Observe that all models yield similar estimates for the confounding variables effects, but the percent of Hispanic residents and college students are the most statistically significant variables, as evidenced by the fact that the curves do not contain the zero line. Note that these results should be interpreted as descriptive summaries of how these daily death trajectories varied across levels of these characteristics. For example, the estimates for Hispanic population are first negative but gradually increase towards zero. One could hypothesize that this relationship could be attributed to the fact that counties with higher Hispanic population, particularly southern counties, also had more extant COVID safety awareness since the epidemic took hold later in calendar time.

Importantly, while we only evaluated explicit pre-intervention heterogeneity with respect to this relatively small subset of covariates that were included in the model, the inclusion of county-specific random effects is designed to capture additional heterogeneity in death curves attributable to other factors.

4.3 Effectiveness of Intervention Timing

Using the model specification in Section 3.1, we offer comparisons between the modeled trajectory of daily deaths in each county and a model-based prediction of what the daily death trajectory would have been under different intervention timing. Specifically, we offer posterior predictions for each county under the hypothetical scenario where the intervention (policy or mobility) had been enacted 10 days before or 10 days after the observed date. Figure 4 shows the average fitted curves of daily deaths (per capita) under the observed intervention scenarios and hypothetical intervention timings, separated by NCHS category. These curves are calculated by averaging point-wise posterior predictive quantities among counties within the same NCHS category.

Figure 4(a) depicts modeled daily death trajectories for different timing of stay-at-home orders. While average trajectories are largely overlapping between the observed and late intervention timing, there is evidence that intervening 10 days earlier would have significantly impacted the trajectory of daily deaths, particularly for the more urban counties classified as large central, large fringe, or medium metro areas. Table 2 provides numerical estimates (median and inter-quartile range (IQR) intervals) of the average number of days since local epidemic initiation (3/10 million death threshold) to the peak in daily deaths and the height of the daily death peak under each intervention timing. The table suggests statistically significant differences in NCHS 1-3 between the actual and early estimates as well as between actual and late estimates. For example, for both models early intervention in NCHS 1 is associated with a change in deaths at the peak of approximately 2 less deaths per 1 million, and an earlier peak of 5 days for the stay-home model and 8 days for the

	Days since threshold at peak			Deaths per 1 million at peak		
NCHS	*Early*	*Actual*	*Late*	*Early*	*Actual*	*Late*
1	21 (2)	26 (1)	28 (2)	1.59 (0.42)	3.50 (0.16)	4.94 (0.89)
2	19 (4)	23 (2)	23 (1)	1.56 (0.49)	2.95 (0.29)	3.59 (0.52)
3	15 (2)	21 (2)	23 (2)	1.02 (0.25)	1.99 (0.16)	2.47 (0.41)
4	14 (2)	19 (2)	20 (2)	1.77 (0.58)	2.46 (0.33)	2.19 (0.49)
5	13 (4)	18 (2)	20 (3)	3.08 (1.40)	4.97 (0.79)	4.84 (1.57)
6	22 (16.25)	23 (14)	18 (4)	18.47 (54.90)	12.12 (9.74)	6.65 (4.17)

(a) Stay-at-home order model

	Days since threshold at peak			Deaths per 1 million at peak		
NCHS	*Early*	*Actual*	*Late*	*Early*	*Actual*	*Late*
1	16 (2)	24 (1)	27 (1)	1.30 (0.37)	3.20 (0.19)	5.66 (1.13)
2	11 (2)	19 (2)	24 (1)	1.02 (0.28)	3.33 (0.29)	7.74 (1.50)
3	10 (2)	18 (1)	21 (2)	0.92 (0.25)	2.31 (0.18)	4.15 (0.88)
4	11 (3)	16 (1)	18 (2)	2.75 (1.14)	2.96 (0.41)	2.41 (0.80)
5	7 (2)	16 (3)	25 (5)	3.27 (1.20)	6.28 (1.03)	9.65 (3.44)
6	11 (12)	21 (25.25)	33 (25)	35.13 (72.47)	16.97 (49.00)	8.34 (71.33)

(b) Mobility decrease model

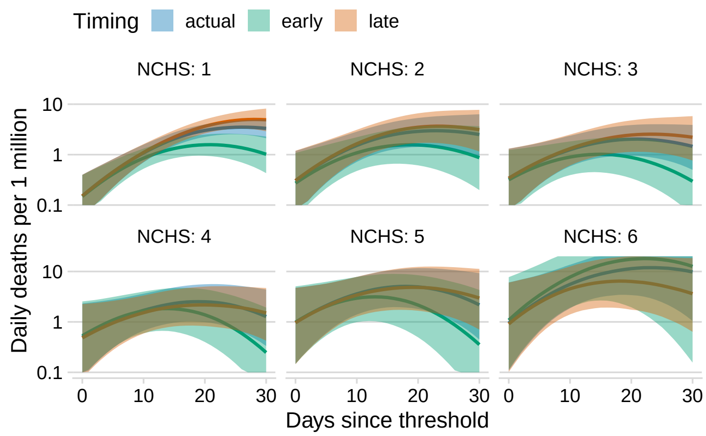
Table 2: Summary statistics for fitted daily deaths curves per NCHS. The table shows the median and interquartile range (IQR) in parenthesis for the peak of the average curve per NCHS from 1000 posterior samples. Significant effects can be seen in NCHS 1-3.

mobility model. Differences in NCHS 4-6 are not statistically significant for either model since the IQR intervals strongly overlap.

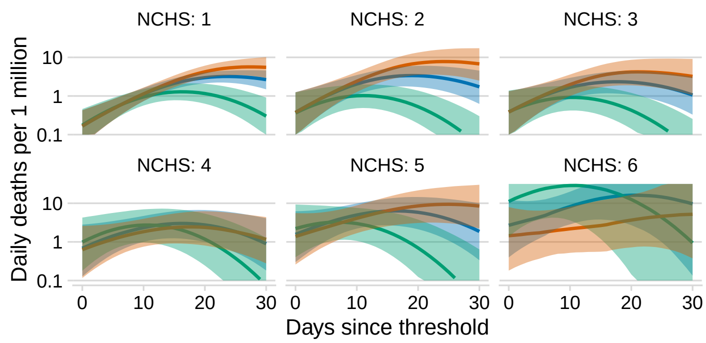
Figure 4(b) depicts estimated curves for the models for mobility intervention timing. While the general shape of trajectories is similar to those for the policy interventions, important differences emerge. For large central, large fringe, and medium metro areas, there is more pronounced evidence that earlier mobility intervention impacted the daily death trajectories; intervening earlier is predicted to have reduced the time until peak death rate and the height of peak death rate. In addition, and unlike in the policy-intervention models, having intervened later than observed is estimated to impact the daily death trajectories among these counties, with longer times to peak and higher peak death rates. See numeric summaries in Table 2. As in the policy intervention analysis, there is evidence of similar patterns in the small metro and micropolitan counties, but the associated uncertainty renders the evidence inconclusive. In fact, as evident from Figure 2, most mobility interventions in these less urban counties occurred several days before the threshold case, with a median value of ten days (Table 3), so it is expected that an earlier or later intervention would have little or no effect because the epidemic conditions in many such counties were practically the same as when the observed mobility change occurred.

In total, the results of the mobility-intervention analysis relative to the policy-intervention analysis match expectations, since stay-at-home orders typically happened after a significant decrease in mobility had already taken place (*c.f.* Table 3), with mobility drops persisting beyond the first date reaching a 50% reduction from baseline. In fact, the "double intervention" model from Section 3.1 that includes terms to bend trajectories based on both the mobility- and policy-based intervention definition estimates that, when accounting for mobility reduction timing in the model, the impact of stay-at-home order was negligible (i.e., trend-altering parameters not significantly different from zero).

While Figure 4 and Table 2 summarize results averaged across counties in each NCHS category, an important feature of the underlying model from Section 3.1 is that it permits multiple sources of county-level heterogeneity, with each county's estimated daily death trajectory dependent on both it's characteristics and county-specific random effects for trajectory parameters. To highlight this heterogeneity, Figure 6 shows the observed and hypothetical trajectories for three counties: King County, Washington; Kings County, New York; and Jefferson, Louisiana. In King, Washington, a hypothetical late adoption of a stay-home order does not appear to have a significant impact on the post-intervention death curve trajectory, suggesting that its curve had already flattened by the time the policy was adopted, likely as a result of awareness brought by national attention as the first US



(a) **Stay-at-home order model**



(b) **50% mobility decrease model**

Figure 4: For each time t, the figure shows the median and 90% credible intervals for the estimated timing effect averaged by NCHS and different counterfactual levels for early (green) and late (red) intervention considering a 10-day difference from the actual (blue) intervention.

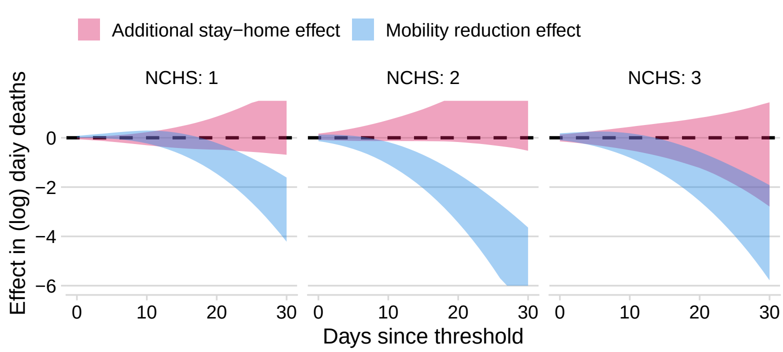


Figure 5: Shows 90% credibility intervals for the estimated effects from the **double intervention** model. We observe that the main effect in decreasing deaths comes from the reduction in mobility, and it does not appear to be an additional effect from the stay-home intervention (it contains the zero line).

cases emerged in this county. In Kings, New York, the 50% mobility decrease was reached three days after the stay-home was adopted and yet the models' estimation show that a late mobility-intervention would have a significantly higher peak than a late policy-intervention and that it would more dramatically bend the post-intervention trend. Finally, in the case of Lousiana's Jefferson, even if both the policy and mobility interventions happened exactly on the same day, the late policy-intervention counterfactual has a considerable overlap with the original fit contrasting with the late mobility-intervention counterfactual.

5 Discussion

We have offered a rigorous statistical model that expresses curves of daily COVID-19 deaths in terms of a pre- and a post intervention trends, where the pre-intervention portion is determined by the average effect of different county-level demographics, and the post-intervention by the timing interventions across US counties. While anchored to time quadratic functions that approximate epidemic growth patterns, this formulation allowed us to make inferences on the relative effectiveness of the timing of policy vs mobility interventions. We refer to a "policy intervention" to the date of a stay-at-home order, and a "mobility intervention" to mobility reductions measured in total visits to points of interest collected by SafeGraph.

The descriptive analysis of the timing of stay-at-home orders and total visits to points of interest clarified both the heterogeneity in intervention timing relative to local epidemic conditions and that mobility reductions often occurred prior to stay-home-orders. By-and-large, more urban counties classified as large central, large fringe, or medium metro areas tended to have the epidemic arrive earlier in calendar time, with mobility and policy interventions in these areas tending to be later in epidemic time relative to less urban counties. Also, counties in our dataset reveal that the total visits to POI did not reduce much more after stay-home orders were adopted across all the rural-urban spectrum, as the total visits had often reached close to its minimum observed level at the time a stay-at-home order was instituted.

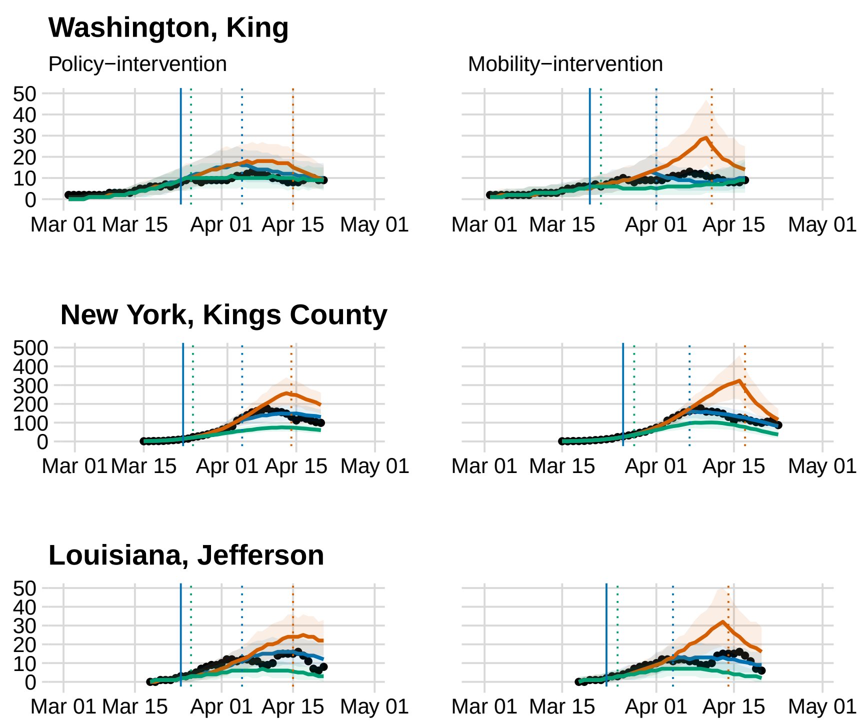


Figure 6: The early and late intervention timing counterfactuals for three counties. We simulate 1000 draws from the Bayesian posterior predictive distribution of the (log) expected counts for each county, and take point-wise medians and 5%/95% quantiles for each t: in blue the fitted credible intervals; in green (red) the early (late) credible bands around the counterfactual medians. The three plots on the left show estimates using the stay-at-home model, and on the right estimates from the mobility model. The vertical dotted lines depict the actual and hypothetical date of the intervention plus the intervention lag.

With model specifications specifically tailored to various dimensions of county-level heterogeneity, the suite of statistical models fit indicated that the timing of mobility reductions was more important for dictating changes in the daily deaths trajectories than the timing of official stay-at-home orders for more urban counties, while the uncertainty of the results for counties rural counties are rendered as inconclusive. The analysis of mobility interventions indicated that hypothetically shifting mobility decreases ten days earlier would have reduced deaths, with hypothetically later mobility intervention resulting in more deaths. In contrast, the analysis of stay-at-home orders showed less pronounced impact of intervention timing, indicating some benefit of early action but little or no impact of delayed action relative to the observed timing of the stay-at-home order. These results were corroborated by the "double intervention" model that estimated little or no impact of stay-at-home timing after accounting for the timing of mobility decrease. Taken together, these results point towards the conclusion that the timing mobility drops have more influence over the dynamics in county death trajectories, with stay-home orders having little benefit above-and-beyond that induced by the persistent drop in mobility.

However, interpretation of the relative effectiveness of mobility versus policy interventions generally relies on an interplay between these two types of "interventions" that is not fully resolved in the present work. The relative timing of mobility decreases relative to timing of stay-at-home orders makes clear that there were other drivers of behavior change. We point out that since our mobility measure is determined by visits to all points of interest, other polices such as school and restaurant closures inevitably impacted this measure to some degree that is separate from the stay-at-home order. Additional influence of non-policy drivers of behavior change such as awareness-driven voluntary actions on visits and other mobility metrics is likely to have also played a role in a manner not captured by the present analysis. This is not to say that the stay-at-home orders had no effect, even though total visits had often decreased more than 60% and close to its minimum in advance of official stay-at-home orders, these orders were likely to have served maintain mobility low after its initial decrease. Work in [Abouk & Heydari, 2020] describes how, at the state level, stay-home orders increased the actual presence at home, a metric of mobility that did not seem as sound for use at the county level. What's more, the timing of official policy interventions may well have been influenced by observed mobility if, for example, policy makers were prompted to adopt an official order to continue behavior changes that were already occurring. An analysis that fully resolves the potential confounding and mediating effects that may result from the influence of mobility reductions on the timing of stay-at-home orders (or *vice versa*) is beyond the scope of this analysis.

An important feature of the analysis is its focus on the relative effectiveness of intervention timing cannot necessarily correspond to the relative effectiveness of the policies as a whole. Measuring the change in deaths after a policy does not necessarily capture the policy impact relative to what would have occurred without the policy. The method pursued here attempts to characterize how different intervention timing bent death trajectories post-intervention by learning the full functional form of quadratic curves, which contrasts with DiD designs that rely on time cross-sectional death drops captured by linear trends, yet arguably a simpler task than characterizing what would have happened without any intervention whatsoever. Furthermore, the analysis is unable to explicitly account for differences in the execution of the official orders. In truth, execution of policies or implications of mobility changes may well vary across the US, with the present analysis unable to capture such variation beyond that which may be captured by NCHS urban-rural classification.

Finally, the inferences here are valid insofar as the statistical model with quadratic time trends represents a reasonable approximation to the shape of the death trajectories. While it has been observed that the daily number of deaths can be approximated with such quadratic curves after the initial introduction of cases and for a limited time frame afterward[3], this parametric specification is only an approximation of the more detailed nuances of epidemic growth. The county-level heterogeneity afforded by the random effects specification is expected to improve this approximation at the county level. We should add soemthign here about why this approximation is worth it, i.e., why we were able to answer things that an SIER model couldn't answer at all or couldn't answer without being encumbered by other assumptions that we may or may not wish to make in every analysis. We should get some help with this point.

[3] In the present work the time frame is capped to 29 days after an intervention in each county as mentioned before in section 2.3. Of these 29 days, 12 days correspond to the intervention lag, and 17 days to the post-intervention trend.

Our work was specifically designed to surmount the challenges of disentangling the intertwined local characteristics and events that unfolded around the time of the first wave of COVID-19 interventions. The limitations of this analysis notwithstanding, we move a step closer into parsing these events by providing evidence for the timing of mobility-related behavior changes as an important determinant of local daily COVID-19 deaths. These results point towards the need to investigate how official reopening policies and other policies that varied across counties interplay with changes in mobility beyond the time frame considered here and into the later phases of the US COVID-19 epidemic.

References

Rahi Abouk and Babak Heydari. The immediate effect of covid-19 policies on social distancing behavior in the united states. *Available at SSRN*, 2020.

Charles Courtemanche, Joseph Garuccio, Anh Le, Joshua Pinkston, and Aaron Yelowitz. Strong social distancing measures in the united states reduced the covid-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed covid-19 cases across the united states. *Health Affairs*, pp. 10–1377, 2020.

Dhaval M Dave, Andrew I Friedson, Kyutaro Matsuzawa, and Joseph J Sabia. When do shelter-in-place orders fight covid-19 best? policy heterogeneity across states and adoption time. Technical report, National Bureau of Economic Research, 2020.

Ingram Dd and Franco Sj. NCHS urban-rural classification scheme for counties. *Vital and Health statistics. Series 2, Data Evaluation and Methods Research*, (154):1–65, January 2012. ISSN 0083-2057, 2333-0872. URL `https://europepmc.org/article/med/22783637`.

Jonas Dehning, Johannes Zierenberg, F Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. *Science*, 2020.

Klaus Desmet and Romain Wacziarg. Understanding spatial variation in covid-19 across the united states. Technical report, National Bureau of Economic Research, 2020.

Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via stan. *R package version*, 2(4):1758, 2018.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Yothin Jinjarak, Rashad Ahmed, Sameer Nair-Desai, Weining Xin, and Joshua Aizenman. Accounting for global covid-19 diffusion patterns, january-april 2020. Technical report, National Bureau of Economic Research, 2020.

Rolly Kapoor, Haedong Rho, Kinpritma Sangha, Bhavyaa Sharma, Ajay Shenoy, and Guanghong Xu. God is in the rain: The impact of rainfall-induced early social distancing on covid-19 outbreaks. *Available at SSRN 3605549*, 2020.

Sadiya Khan, Megan McCabe, Amy Krefman, Lucia C Petito, Xiaoyun Yang, Kiarri Kershaw, Lindsay Pool, and Norrina B Allen. A county-level susceptibility index and coronavirus disease 2019 mortality in the united states: A socioecological study. *medRxiv*, 2020.

Benjamin D. Killeen, Jie Ying Wu, Kinjal Shah, Anna Zapaishchykova, Philipp Nikutta, Aniruddha Tamhane, Shreya Chakraborty, Jinchi Wei, Tiger Gao, Mareike Thies, and Mathias Unberath. A County-level Dataset for Informing the United States' Response to COVID-19. April 2020.

Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pp. 963–974, 1982.

Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

Zhixian Lin and Christopher M Meissner. Health vs. wealth? public health policies and the economy during covid-19. Technical report, National Bureau of Economic Research, 2020.

Sen Pei, Sasikiran Kandula, and Jeffrey Shaman. Differential effects of intervention timing on covid-19 spread in the united states. *medRxiv*, 2020.

H Juliette T Unwin, Swapnil Mishra, Valerie C Bradley, Axel Gandy, Thomas A Mellan, Helen Coupland, Jonathan Ish-Horowicz, Michaela Andrea Christine Vollmer, Charles Whittaker, Sarah L Filippi, et al. State-level tracking of covid-19 in the united states. *medRxiv*, 2020.

Nick Wilson, Amanda Kvalsvig, Lucy Telfar Barnard, and Michael G Baker. Case-fatality risk estimates for covid-19 calculated by using a lag time for fatality. *Emerging infectious diseases*, 26(6):1339, 2020.

Spencer Woody, Mauricio Garcia Tec, Maytal Dahan, Kelly Gaither, Michael Lachmann, Spencer Fox, Lauren Ancel Meyers, and James G Scott. Projections for first-wave covid-19 deaths across the us using social-distancing measures derived from mobile phones. *medRxiv*, 2020.

Xiaobo Yang, Yuan Yu, Jiqian Xu, Huaqing Shu, Hong Liu, Yongran Wu, Lu Zhang, Zhui Yu, Minghao Fang, Ting Yu, et al. Clinical course and outcomes of critically ill patients with sars-cov-2 pneumonia in wuhan, china: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*, 2020.

Appendix

NCHS	Total number of counties in US	Number of counties in dataset	Population covered in dataset	Median days between policy intervention and threshold	Median days between mobility intervention and threshold	Median days between mobility and policy intervention
1	68	57	93 M	5	0	-4
2	368	120	48 M	0	-5	-4
3	373	121	41 M	0	-5	-6
4	358	67	9 M	0	-5	-4
5	641	52	3 M	-3	-10	-6
6	1339	23	0.72 M	-5	-11	-7
Total	3147	440	194.72 M	0	-5	-5

Table 3: Description of the counties coverage in the analysis grouped by NCHS along with median number of days between interventions and death threshold.

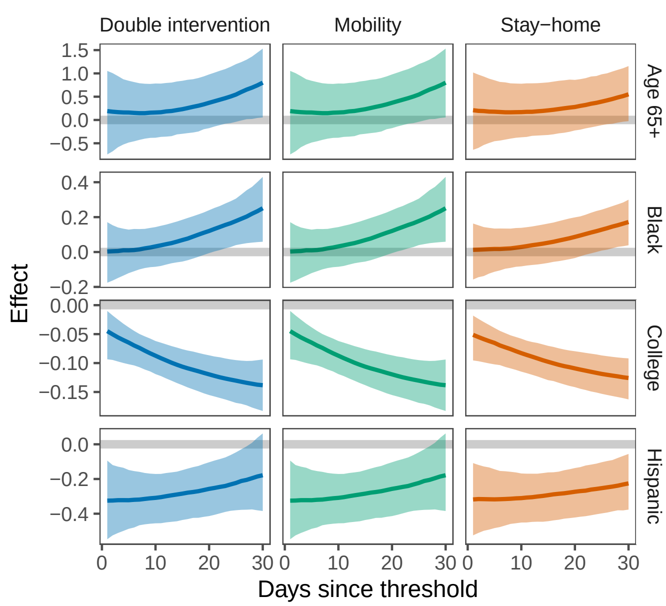


Figure 7: Shows the time-varying of the confounding variables. For each covariate the figure shows median and 90% posterior credible intervals. The zero-th line is shown in gray in each subplot. Effects that do not consider contain the zero-th line are statistically significant.

Parameter	Stay-home only *Mean*	*[5%, 95%]*	Mobility only *Mean*	*[5%, 95%]*	Double intervention *Mean*	*[5%, 95%]*
Degree=0						
(Intercept)	-11.3*	[-15.8, -6.79]	-9.71*	[-14.4, -5.42]	-10*	[-14.6, -5.38]
NCHS-2	0.27	[-0.11, 0.67]	0.87*	[0.44, 1.26]	1*	[0.54, 1.43]
NCHS-3	0.01	[-0.42, 0.5]	0.55*	[0.11, 0.99]	0.65*	[0.2, 1.09]
NCHS-4	-0.17	[-0.74, 0.4]	0.03	[-0.63, 0.65]	0.25	[-0.37, 0.9]
NCHS-5	0.4	[-0.34, 1.08]	0.9*	[0.29, 1.54]	1.04*	[0.28, 1.74]
NCHS-6	0.12	[-0.96, 1.18]	-0.64	[-2.48, 1.09]	0.22	[-1.77, 2.2]
% in college	-0.09*	[-0.12, -0.07]	-0.08*	[-0.11, -0.06]	-0.08*	[-0.11, -0.06]
% black	0.07	[-0.03, 0.16]	0.07	[-0.02, 0.18]	0.11*	[0.01, 0.21]
% hispanic	-0.29*	[-0.45, -0.15]	-0.36*	[-0.5, -0.22]	-0.34*	[-0.47, -0.2]
% +65 age	0.28	[-0.26, 0.82]	0.07	[-0.47, 0.57]	0.04	[-0.54, 0.62]
Degree=1						
(Linear)	-19.7	[-313, 274]	20.7	[-217, 266]	-46.8	[-287, 168]
NCHS-2	-37.6*	[-70.6, -3.77]	8.29	[-19.6, 35.1]	11.5	[-18.7, 42.7]
NCHS-3	-50*	[-84.3, -18]	-15.5	[-44.3, 14]	-14.8	[-49.5, 15.3]
NCHS-4	-102*	[-148, -61.4]	-91.2*	[-134, -49]	-86.3*	[-130, -42.5]
NCHS-5	-98.5*	[-154, -41.7]	-42.7*	[-81.4, -2.42]	-63.3*	[-116, -13.1]
NCHS-6	-140*	[-229, -52.1]	-151*	[-252, -59.9]	-62.2	[-223, 95.7]
% in college	-2.56*	[-4.2, -0.87]	-1.89*	[-3.24, -0.43]	-1.74*	[-3.09, -0.43]
% black	7.39	[-0.14, 14.6]	1.85	[-3.64, 7.67]	5.43	[-0.38, 11.1]
% hispanic	4.62	[-4.96, 14.2]	5.4	[-1.18, 12.6]	4.6	[-2.4, 11.8]
% +65 age	15.9	[-19, 48.7]	13	[-16.3, 40.3]	19.3	[-6.67, 46.6]
Degree=2						
(Quadratic)	-172	[-407, 52.9]	-20.7	[-176, 130]	-4.05	[-176, 154]
NCHS-2	-7.46	[-28.2, 12]	0.43	[-14.5, 14.7]	0.38	[-14.3, 14.3]
NCHS-3	4.2	[-17.2, 25.5]	-2.64	[-17.3, 11.2]	0.51	[-14.7, 16]
NCHS-4	-10.4	[-38.2, 16.7]	-15.7	[-35.3, 3.34]	-15.6	[-36.5, 4.3]
NCHS-5	-13	[-45.3, 20.9]	9.42	[-9.01, 28]	-4.78	[-30, 18.8]
NCHS-6	-47.2*	[-94.3, -0.89]	-2.15	[-52, 46.9]	36.6	[-42.5, 116]
% in college	0.61	[-0.63, 1.9]	-0.04	[-0.99, 0.83]	-0.07	[-0.95, 0.85]
% black	2.65	[-3.14, 8.56]	-2.77	[-6.7, 1.12]	-1.08	[-5.09, 3.28]
% hispanic	2.32	[-4.39, 9.12]	1.37	[-3.42, 5.89]	0.93	[-3.68, 5.96]
% +65 age	9.06	[-17.2, 36.8]	1.14	[-16.6, 19.7]	-2.31	[-21.7, 17.7]

Table 4: Shows posterior summaries for the model coefficients of each model. The asterisk highlights mean estimates that fall outside the 95% credible interval. Note that for each model a different polynomial orthogonalization is used.