Evaluating Public Masking Mandates on COVID-19 Growth Rates in U.S. States

EVALUATING PUBLIC MASKING MANDATES ON COVID-19 GROWTH RATES IN U.S. STATES

A Thesis Presented

by

ANGUS K. WONG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

May 2021

Public Health
Biostatistics and Epidemiology

EVALUATING PUBLIC MASKING MANDATES ON COVID-19 GROWTH RATES IN U.S. STATES

A Thesis Presented

by

ANGUS K. WONG

Approved as to style and content by:

Laura B. Balzer, Chair

Nicholas G. Reich, Member

Zhichao Jiang, Member

Lisa Chasan-Taber, Department Chair
Biostatistics and Epidemiology

ACKNOWLEDGMENTS

First, I would like to thank my mentor Dr. Laura Balzer for constantly guiding and motivating me throughout my thesis. Her approach to prioritize my learning allowed me to absorb as much knowledge on causal inference as possible. Without her encouragement, I could not imagine this thesis being as interesting and relevant as it is. She has taught me to be a better critically thinker, statistician, and above all a more balanced person. I am very grateful for her support both inside and outside of the classroom.

Secondly, I would like to thank Professor Nicholas Reich for introducing me to the world of biostatistics. Without him, I would not be in this field that I now enjoy so much. His dedication and unconventional path into the field inspire me to fearlessly pursue my interests. I am very thankful to have him as a mentor and committee member.

Thirdly, I would like to thank Professor Zhichao Jiang for being a knowledgeable teacher and showing me a different perspective on causal inference. I appreciate having him as a thesis committee member.

I would also like to thank everyone in the Department of Biostatistics and Epidemiology who have helped, taught, or worked with me during my time at UMass. This journey would not have been possible without everyone's help.

Lastly, I would like to thank my family and friends for being very supportive, understanding, and encouraging while I pursue a Masters degree.

ABSTRACT

EVALUATING PUBLIC MASKING MANDATES ON COVID-19 GROWTH RATES IN U.S. STATES

MAY 2021

ANGUS K. WONG

B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Laura B. Balzer

Background: U.S. state governments have implemented numerous policies to help mitigate the spread of COVID-19. While there is strong biological evidence supporting the wearing of face masks or coverings in public spaces, the impact of public masking policies remains unclear.

Methods: We aimed to evaluate how early versus delayed implementation of state-level public masking orders impacted subsequent COVID-19 growth rates. We defined "early" implementation as having a state-level mandate in place before September 1, 2020, the approximate start of the school-year. We defined COVID-19 growth rates as the relative increase in confirmed cases 7, 14, 21, 30, 45, 60-days after September 1. Primary analyses used targeted maximum likelihood estimation (TMLE) with Super Learner and considered a wide range of potential confounders to account for differences between states. In secondary analyses, we took an unadjusted approach and calculated the average COVID-19 growth rate among early-implementing states divided by the average COVID-19 growth rate among late-implementing states.

Results: At a national level, the expected growth rate after 14-days was 4% lower with early vs. delayed implementation (aRR: 0.96; 95%CI: 0.95-0.98). Associations did not plateau over time, but instead grew linearly. After 60-days, the expected growth rate was 16% lower with early vs. delayed implementation (aRR: 0.84; 95%CI: 0.78-0.91). Unadjusted estimates were exaggerated (e.g. 60-day RR: 0.72; 95%CI: 0.60-0.84). Sensitivity analyses varying the timing of the masking order yielded similar results.

Conclusion: In both the short and long term, state-level public masking mandates were associated with lower COVID-19 growth rates. Given their low-cost and minimal (if any) impact on the economy, masking policies are promising public health strategies to mitigate further spread of COVID-19.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

By early November 2020, the coronavirus disease (COVID-19) pandemic had resulted in over 9 million confirmed positive cases and 225,000 deaths in the United States (US) [10]. Previously in the pandemic, both federal and state governments had implemented numerous public health measures in an effort to "flatten the curve" and minimize further transmission [47]. However, the timing and strictness of each policy has differed and continues to differ between states. For example, a stay-at-home (a.k.a. shelter-in-place) order was implemented as early as March 19, 2020 in California, while some states have never implemented one as of December 2020 [39]. Other major non-pharmaceutical interventions (NPIs) include non-essential business closures, school closures, large gathering bans, and public masking. However, as the pandemic unfolds, the effectiveness of these NPIs remain unknown and difficult to examine.

There has been an increasing amount of scientific evidence supporting the hypothesis that NPIs can help mitigate the COVID-19 growth rate. For example, using an event-study econometric regression model, Courtemanche et al. evaluated the impact of strong social distancing policies, defined as large social gatherings ban, school closures, public facilities closures, and shelter-in-place orders, on the daily case growth rate in 3,138 U.S. counties [15]. They defined daily growth rate as natural log of total cases in a day minus the log of total cases in previous day, and assumed exponential growth in the absence of the intervention. Using this approach, the researchers estimated that from March 1 - April 27, 2020 the county-level daily growth rate would

have been reduced by 5.4, 6.8, 8.2, and 9.1 percent after 1-5, 6-10, 11-15, 16-20 days of the government-imposed policy, respectively. Courtemanche et al. also predicted that confirmed case count would have been ten times higher without shelter-in-place orders and thirty-five times higher without the other four strong policies. These NPIs, however, are highly disruptive to both social and economic well-being.

Policies to encourage or enforce wearing face masks or coverings while in public (hereafter called "public masking order") are promising approaches that have been hypothesized to help mitigate the spread of COVID-19 for the following reasons. First, many densely populated Asian regions with high level of mask compliance such as Hong Kong, Japan, and Singapore have demonstrated success in combating the pandemic [19, 27, 55]. Second, in comparison to other more restrictive social distancing measures, a public mask order is relatively inexpensive to implement and does not have as much of a negative impact on the US economy [54]. Third, as various re-opening plans are considered and implemented, public masking can be used as a compliment, instead of an alternative, to other social distancing policies [17]. Nonetheless, public masking orders remain controversial, and official recommendations have been inconsistent [43].

At the beginning of the pandemic, the US Centers for Disease Control and Prevention (CDC) did not recommend public masking due to concern of masks shortage for health professionals [13]. Although the CDC updated the recommendation on April 3, 2020 to advise the use of cloth face cover in public places, many remain skeptical of the policy's necessity and effectiveness [18]. State-specific recommendations and policies have also varied [39]. In Massachusetts, for example, the initial public masking order, issued on May 1, 2020, requires everyone aged 2+ years to wear a mask or a face covering at both indoor and outdoor public places where social distancing of six feet is not possible. It was subsequently revised on November 2, 2020 to require

face masks or cloth face coverings in all public places for all aged 5+ years even when six feet social distance could be maintained [37, 41, 40].

Although the biological efficacy of masks use has been demonstrated by many studies [3, 5, 25], there is limited scientific evidence to demonstrate the effectiveness of mask policy at a population-level. In this dissertation, we aim to help fill this knowledge gap by evaluating the impact of public masking on mitigating the COVID-19 cases by using a formal framework for causal inference [6, 45]. Before doing so, we briefly review other studies also aiming to understand the impact of public mask orders on COVID-19 transmission at a population-level.

1.1 Evidence to date

Randomized controlled trials are the gold-standard for providing evidence of effectiveness. To date, however, there have not been any trials of mask wearing and COVID-19 incidence. In words of Greenhalgh et al., to combat this pandemic, "it is time to act without waiting for randomized controlled trial evidence" [23]. Specifically, observational studies, when appropriately analyzed, can also provide strong evidence of policy impact [7].

While differing in methodologic approach, several studies have provided evidence to support public masking. For example, using a previously developed Susceptible-Exposed-Infectious-Recovered (SEIR) model, Eikenberry et al. simulated the impact of mask mandates on the effective transmission rate (the average number of people each infectious individual can infect daily) under different levels of mask efficacy and compliance in two states: New York and Washington [17]. Their simulations suggested that 50% coverage of masks with 50% effectiveness could halve the effective transmission rate, and 80% coverage of of masks with 20% effectiveness could reduce effective transmission rate by one third. These modeling results suggest that while high compliance in combination with highly effective masks (e.g., N95s) is ideal,

universal adoption of less effective masks (e.g., homemade cloth masks) may still be very impactful in mitigating COVID-19.

Using SEIR and agent-based models, Kai et al. conducted simulations from March 23 onward for 500 days, and found that compliance with both social distancing (50%) and masking (80%-90%) must be quite high to markedly reduce COVID-19 transmission [29]. When validating their theoretical simulation results with empirical data (obtained from 38 counties or provinces in Asia, Europe, and North America) on universal masking and case incidence between January 23 to April 10, 2020, they found that universal masking was nearly perfectly correlated with strong reductions peak daily growth rate and successful suppression of daily growth rate (Defined as below 12.5% daily growth). In their study, countries with universal masking by early during the pandemic experienced lower peak daily growth rate than countries that delayed universal masking. Countries with lockdown policies but without universal masking experienced higher daily growth than groups with any universal masking. They suggested delaying implementation past 50 days since the outbreak onset would substantially reduce its impact on subsequent growth rates. In other words, not only does the implementation of universal masking matter, the timing of such policy is also critical.

Using a deterministic SEIR framework and adjusting for covariates such as seasonal trend for pneumonia related deaths and mobility, the IHME COVID-19 Forecasting Team also examined the timing of interventions by projecting COVID-19 cases and mortality across the US from September 2020 through the end February 2021 [26]. Under scenario where social distancing mandates were reinforced only after a threshold of 8 deaths per million was surpassed and without any masking mandate, they projected between 469,578 – 578,347 COVID-19 death could occur by the end of February. However, under the same scenario but with 85% universal mask use, between 60,731 – 170,867 lives could be saved. On a smaller scale, Babino and Mag-

nasco conducted a similar study for Connecticut, Massachusetts, and New York using regularized regression and susceptible-infectious-recovered (SIR) framework [4]. Their model predicted that a public masking policy could potentially prevent 170,000 cases from the time of implementation of stay-at-home orders until their lift.

In addition to simulation studies, a survey conducted by the Delphi group demonstrated a strong state-level correlation between the mask wearing rate and the COVID-19 case rate [16]. While these results are encouraging, correlation is not causation. There are a plethora of reasons why two variables may be correlated, and for effective decision-making we need to explicitly understand how and why a policy impacts the outcome of interest. In particular, failing to appropriately control for common causes of the policy and outcomes can lead to misleading results. For a COVID-19-related demonstration with simulated data, we refer the reader to Balzer and Whitcomb [8]. Furthermore, in another working paper, we have shown that unadjusted estimates of the association between state-level shelter-in-place orders and COVID-19 death rates were substantially inflated, as compared to estimates controlling for the many ways in which states differ from each other.

Formal frameworks for causal inference also require the assumptions needed to interpret estimates as effects (versus associations) to be explicitly stated and critically evaluated. This is in direct contrast with "standard" approaches that overlook these assumptions and interpret results from correlation or modeling studies as providing estimates of causal effects. For these reasons, causal inference frameworks can help provide more meaningful results for decision making, while being transparent about assumptions and limitations.

With respect to mask wearing policies, studies using causal inference methods have produced conflicting results. For example, using the difference-in-difference method, Lyu and Wehby examined the impact of different levels of public masking policies on daily growth rate, defined again as the natural log of total cases in a day minus the

log of total cases in previous day [36]. They compared 15 states and the District of Columbia with a public masking mandate to the remaining 35 states without such a mandate. Comparing to the reference level of daily case rate from 1-5 days prior to masking mandate, they examined five post-intervention periods (corresponding to 1-5, 6-10, 11-15, 16-20, 21+ days after March 25, 2020) and found that the mandate was associated with a decline of 0.9, 1.1, 1.4, 1.7, 2.0 percentage points in state-level daily COVID-19 growth rates. However, their comparison of 20 states with a masking order for only employees but not the general public to 15 states without any masking mandate found no significant association.

On the other hand, using an econometric approach based on linear structural equations, Chernozhukov et al. did find an association between employee only mask mandates and COVID-19 cases and deaths at a national-level in the U.S [14]. Specifically, their analysis suggested that if a national mandate for employee face masking had been issued on March 14, 2020, then the cumulative numbers of cases and fatalities by late April would have would been reduced by 21% and 34%, respectively.

Altogether prior studies consistently show the importance of public mask wearing to help mitigate COVID-19 transmission. Nonetheless, nearly a year after the pandemic onset in the US, public opinion remains divided and policies-makers are still hesitant to issue mandates. The continued variation in policy implementation across states provides an opportunity for further evaluation. Now with more data and possibly a better understanding of COVID-19, our objective is to evaluate the impact of mask mandate at a national-level. To do so, we follow the Causal Roadmap, a formal framework for causal inference with the following steps: defining the scientific question, specifying the causal model based on real-world knowledge, specifying the causal parameters using counterfactuals, linking the observed data to causal model, identifying the statistical estimand, statistical estimation and inference, interpretation and discussion [6]. Given its superior statistical performance and in line with recom-

mendations by Friede for studying COVID-19, we used targeted maximum likelihood estimation (TMLE) for estimation and inference [20, 53].

CHAPTER 2

METHODS

We aim to evaluate the effect of state-level mandates for public masking on the growth rate of COVID-19 cases in the US. To sharply frame this question, we need to be more specific about the exposure and outcome definitions as well as their timing. Indeed, there have been many variations of state-level mask policies, ranging from simple recommendations to requirements for both public indoor and outdoor areas [39]. Throughout, we focus on the strongest policy, assuming that the strongest mandate would lead to the greatest impact, and reserve examination of other mask policies for future work. Therefore, we define our primary exposure A as an indicator that a given state had issued a mandate requiring masks or cloth face coverings in public indoor and outdoor spaces when it was not possible to social distance (i.e., maintain at least 6-feet apart) by the target date (described next).

In the primary analysis, the target date is September 1, 2020, selected to correspond roughly with the start of the school year for many universities and K-12 schools. Given the anticipated shift in behavior for many US residents (i.e., from summertime activities to classroom activities) as well as the increased mobility for college-aged students, we hypothesize that having a masking mandate in place before September 1, 2020 would limit the spread of COVID-19. To further understand the impact of state-level public mask mandates on the COVID-19 epidemic, we conduct a sensitivity analysis with the exposure defined as having issued mandate requiring masks or cloth face coverings in public indoor and outdoor spaces when it was not possible to social distance (i.e., maintain at least 6-feet apart) before shelter-in-place

(a.k.a., stay-at-home) orders were terminated. In this secondary analysis, the "target date" is thereby state-specific.

To account for variability in the arrival of COVID-19 to each state, we avoid raw counts and instead focus our outcome definition on relative changes. Suppose, for example, COVID-19 arrived in State-A one month prior to State-B; comparing their daily or cumulative case counts would be potentially misleading, given that the states were in different stages of the epidemic. Using relative change allows for each state to serve as its own reference. Specifically, we define the outcome Y as the state-specific COVID-19 growth rate, defined as the number of confirmed cases a set number of days after the target date, divided by the number of confirmed cases on the target date.

To capture the immediate impact of the policy, we focus on 14-day delay to account for lags in behavior change and case detection (e.g., the median incubation period of 5.1 days) [33]. To address the sensitivity to the timeframe as well as examine longer term effects, we also consider 7-days, 21-days, 30-days, 45-days, and 60-days after the target date. Thus, we altogether aim to evaluate the state-level effect of having a masking mandate in place prior to September 1, 2020 on the growth rate of COVID-19 cases over the following 7 to 60 days.

2.1 Causal model

To translate our scientific objective into a causal parameter, we specify a structural causal model to express the relationships between key study variables [42, 53]. These models consist of three main components: endogenous variables, exogenous variables, and structural functions. Endogenous variables represent the set of variables that are essential to answering our scientific question. In this cross-sectional study, the endogenous variables consist of the measured confounders W, the exposure A, and the outcome Y.

When selecting the confounders W, we considered state-level variables that a governor might consult when deciding whether to enact a masking mandate and that might also influence COVID-19 case rates. We separated these covariates into broad categories. First, the population demographics of a given state may influence policy and health outcomes. Older persons, racial/ethnic minorities, the economically disadvantaged, and those with preexisting health conditions have been disproportionately burdened by COVID-19 [1, 11]. Secondly, factors related to population density and urbanicity, could influence policy and infectious outcomes. Thirdly, politics inevitably influence both public policy as well as the compliance with those public health policies, with low compliance potentially leading to more transmission. Lastly, prior public health policies and recently observed COVID-19 outcomes are likely to impact both the state government's urgency to act as well as their residents' behavior. As the pandemic progressed, both governments and the general public have become reluctant to continue the more restrictive policies due to the negative social and economic impact. Altogether, the set of potential confounders consisted of the following state-level variables:

1. Population demographics:
    - Age distribution: Over 65 years (%), median age
    - Race/ethnicity: Black or African American (%), Hispanic (%), Asian (%), Mixed Race (%), Caucasian (%)
    - Economic: Median income, households in poverty (%), people in poverty (%)
    - Co-morbidities: Smoker (%), Diabetes(%)
2. Population density and commuting patterns:
    - Population density (people per km^2), urbanicity (%), average household size

- Drove to work (%), worked at home(%), public transportation usage(%), biked to work (%), other commute mode (%), walked to work (%)

3. Political leading

    - Republican (Indicate whether the Republican party won the majority of the presidential vote in a state in 2016)

4. Prior COVID-19 policies and outcomes before the target date

    - Confirmed cases {7, 14, 30}-days earlier
    - Deaths {7, 14}-days earlier
    - Total tests {7, 14}-days earlier
    - Prior policy implementation (Binary): stay-at-home, gathering restrictions, restaurant restrictions, non-essential business closures; other business closures, business masking, school masking
    - Residential mobility change {7, 14}-days earlier (%)

(Data sources are described below.)

As previously discussed, the exposure A is an indicator of having a mandate requiring the use of masks or cloth face coverings in both public indoor and outdoor space when social distancing is not possible by the target date (September 1, 2020 in the primary analysis). Thus, $A = 1$ for states with an "early" masking policy, and $A = 0$ for states with a "delayed" masking policy. Also as previously discussed, the outcome Y is the COVID-19 growth rate, representing the number of confirmed positive cases within the selected days of target date divided by the number of confirmed positive cases on the target date. Formally, it is defined as follows:

$$\frac{\text{\# Cumulative confirmed cases by n days after target date}}{\text{\# Cumulative confirmed cases by target date}} \quad \text{, where } n \in (7, 14, 21, 30, 45, 60) \tag{2.1}$$

As defined in Eq 2.1, Y can only take on value greater than 1 since the value of the numerator is always greater or equal to the denominator. In words, the cumulative confirmed cases by n days after the target date will always exceed the cumulative confirmed cases by the target date.

The second component of the causal model is the exogenous variables, also known as background variables and denoted $U = (U_W, U_A, U_Y)$ with some joint distribution $\mathbb{P}_U$. These represent the set of unmeasured or unknown factors that determine the values that the endogenous variables can take. For example, perceived or actual compliance with previous public health policies may influence a governor's decision on masking mandates as well as previous and subsequent COVID-19 cases. Such a variable represents an unmeasured common cause of the confounders W, the exposure A, and the outcome Y. Another example could be the strength of the state's public health department, which could influence prior COVID-19 outcomes (included in W), whether the state has an early masking policy A, and the subsequent growth rate of COVID-19 cases Y.

The third component of the causal model is the set of structural equations, which define the relationships between endogenous variables and exogenous variables. Specifically, they allow us to express the value of each of the endogenous variable, as a deterministic function of other endogenous variables and the background factors. Since structural equations are not symmetric and have directionality, causality is encoded in the model.

Altogether the structural causal model for our study is given by:

$$
\begin{aligned}
W &= f_W(U_W) \\
A &= f_A(W, U_A) \\
Y &= f_Y(W, A, U_Y)
\end{aligned}
\tag{2.2}
$$

This model encodes the following relationships: the confounders W are some deterministic function f_W of the unmeasured factors U_W; the exposure A is some deterministic function f_A of the confounders W and unmeasured factors U_A, and the outcome Y is some deterministic function f_Y of the confounders W, exposure A and unmeasured factors U_Y. This structural causal model can also be expressed with the following directed acyclic graph shown in Figure 2.1.

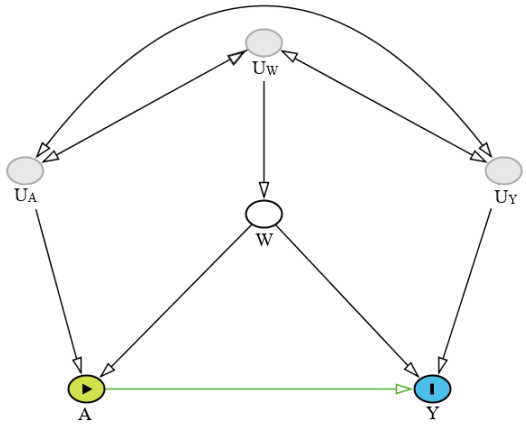


Figure 2.1. Directed acyclic graph representing the masking study.

In either representation, the causal model is specified at the state-level and encodes there is no interference between states. In other words, the outcome for one state is not impacted by the policy of another state. If this assumption is violated (e.g. due to the infectious nature of COVID-19 and travel between states), then our impact estimates are likely to be biased towards the null, while our confidence intervals will be overly precise. In future work, we will relax this assumption by adjusting for the pre-exposure COVID-19 outcomes of neighboring states.

Importantly, we do not specify the functional form of these equations. Equally importantly, we do not specify any independence assumptions on the distribution of unmeasured factors $\mathbb{P}_U$; thus, we allow for unmeasured common causes of the covariates W, the exposure A, and the outcome Y. (When addressing identifiability in Section 2.4, we will revisit these independence assumptions.)

2.2 Counterfactuals & the Causal Parameter

Recall our goal of evaluating the effect of early versus delayed masking mandates on COVID-19 growth rates. To translate this goal into a well-defined causal parameter, we introduce counterfactual outcomes, Y_a, representing the outcome for a state if the exposure-level $A = a$, where $a \in (0, 1)$. Here Y_1 is the relative change in COVID-19 case rate for a state if possibly contrary-to-fact they implemented a masking mandate prior to the target date, while Y_0 is the relative change in COVID-19 case rate for a state if possibly contrary-to-fact they failed to implement a masking mandate prior to the target date. Formally, we can obtain these counterfactual outcomes by deterministically setting the exposure A equal to 1 and 0 in the structural causal model:

$$\begin{aligned} W &= f_W(U_W) \\ A &= a \\ Y &= f_Y(W, a, U_Y) \end{aligned} \tag{2.3}$$

or equivalently on the causal graph shown in Figure 2.2.

With the distribution of counterfactual outcomes, denoted $\mathbb{P}^*$, we can define the target causal parameter. We focus on the causal rate ratio (CRR) in the primary approach and consider the causal rate difference (CRD) in secondary analyses:

$$\text{CRR} = \frac{\mathbb{E}^*(Y_1)}{\mathbb{E}^*(Y_0)} \tag{2.4}$$

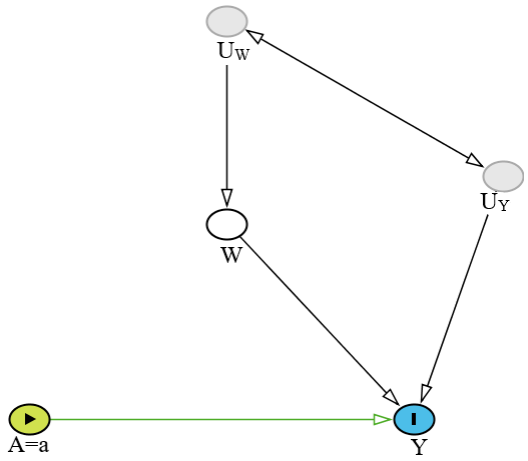


Figure 2.2. Directed acyclic graph under a static intervention to set $A = a$.

$$\text{CRD} = \mathbb{E}^*(Y_1) - \mathbb{E}^*(Y_0) \tag{2.5}$$

These causal parameters correspond to the ratio and difference in the expected COVID-19 growth rate if all states had an early masking policy and expected COVID-19 growth rate if all states had a delayed masking policy.

In the real world, however, we do not get to observe both counterfactual outcomes, since a state either did or did not implement the mask mandate by the target date. Thus, in the next two sections, we define the observed data and their link to the causal model as well as address how these causal parameters can be expressed in terms of the observed data distribution.

2.3 Linking the Causal Model to the Observed Data

For the empirical analysis, we obtain state-level covariates from the Bureau of Transportation Statistics [9], Iowa Community Indicators Program [28], Kaiser Family Foundation [30], MIT Election Data and Science Lab [38], Google Mobility Report [22], and the US Census [49, 50, 51]. Data on COVID-19 policies were collected from the GitHub repository maintained by the COVID-19 State Policy Team at the University of Washington, with verification as needed from the Kaiser Family Foundation,

CNN, and state government websites [21, 31, 32]. For each policy, the issued, enacted, expired, and ended date were collected. From these, we created binary indicators to represent whether a policy was in place by a given date. Given variability in the strictness, masking mandates were categorized into 3 levels: limited to specific public settings, more broadly required indoors and in enclosed spaces, and generally required for both indoor and outdoor public spaces where 6-feet distance cannot be maintained. For this study, we define $A = 1$ for a state with a timely implementation (before September 1, 2020) of strictest policy (general requirement), and define $A = 0$ for a state without timely implementation of the strictest policy or with timely implementation of a weaker policy. Time series data on COVID-19 outcomes were collected from the COVID Tracking Project [48].

From these data sources, we construct the empirical data set. For a given state, the observed data O consist of the measured confounders W, the masking mandate indicator A, and the growth rate outcome Y. We denote the distribution of the observed data as $O \sim \mathbb{P}$. We assume these data arose by sampling repeatedly from some data generating process compatible with the causal model, given in Section 2.1. In other words, we assume the causal model provides a description of the data generating process under observed conditions as well as under specific interventions. This provides a link between the counterfactual outcomes and observed outcomes and in other causal frameworks is called the "consistency" assumption [6, 53].

As shown in Equation 2.3, the observed outcome Y will have the same value as the counterfactual outcome Y_a when the observed exposure $A = a$. For our study, we assume that the observed COVID-19 case growth rate for a state with a delayed mandate would equal to the COVID-19 case growth rate if possibly-contrary-to-fact they had delayed. Similarly, we assume that the observed COVID-19 case growth rate for a state with an early mandate would equal to the counterfactual COVID-19 case growth rate to if they had acted early. By explicitly considering and defining

the level of the mask policy of interest, we helped to ensure the exposure variable is well-defined and improved the plausibility of this assumption.

2.4 Identifiability of causal parameter

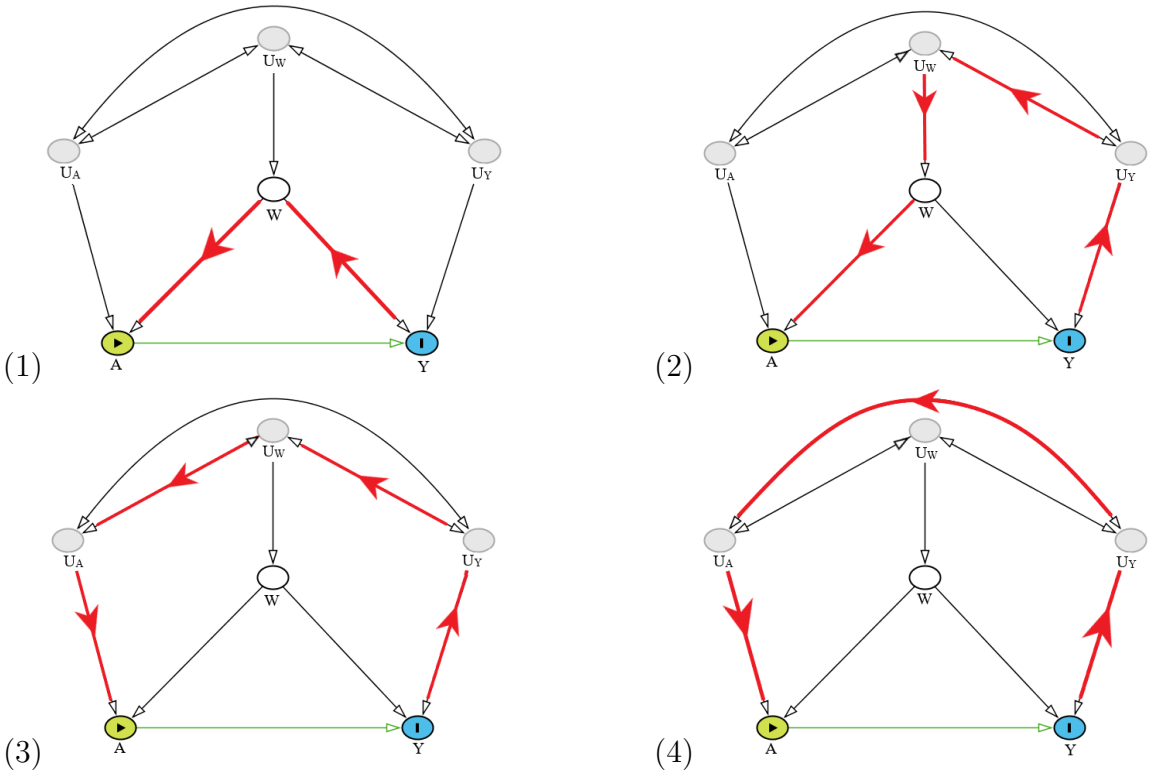
We now assess identification: what are the conditions needed to express the causal parameter as some function of the observed data distribution $O \sim \mathbb{P}$? The commonly cited assumptions of temporality and no interference (a.k.a. stability) have already encoded our in the causal model [6, 53]. Likewise, we established consistency between the observed outcomes and the counterfactual outcomes in the previous step. Nonetheless, to express the expected counterfactual outcome $\mathbb{E}^*(Y_a)$ as function of the observed data distribution $\Psi(\mathbb{P})$, we additionally need there to be no unmeasured common causes of the exposure A and the outcome Y. In other words, we need there to be no unmeasured confounders of the relationship of interest. This assumption is also called "conditional exchangeablity" and the "ignorability" [6, 53].

We can assess the plausibility of this assumption with our causal graph and the Back-door Criterion of Pearl [42]. Back-door paths are a sequence of nodes from the outcome Y into the exposure A. These paths are important as they represent possible relationships that can confound or obscure the causal effect of interest. In order to satisfy the criterion, there must exist a set of observed variables that can be conditioned on (i.e. included in our adjustment set) so that all back-door path are blocked.

As shown in Figure 2.3, in our study, there are four back-door paths to examine: (1) $Y \rightarrow W \rightarrow A$; (2) $Y \rightarrow U_Y \rightarrow U_W \rightarrow W \rightarrow A$; (3) $Y \rightarrow U_Y \rightarrow U_W \rightarrow U_A \rightarrow A$; and (4) $Y \rightarrow U_Y \rightarrow U_A \rightarrow A$.

The first two paths can be blocked by adjusting for measured confounders W. However, the third and fourth paths cannot be blocked, because $U = (U_W, U_A, U_Y)$ represents the set of unmeasured common causes, and we cannot statistically adjust

Figure 2.3. Possible Back-door Paths from Y to A



for unmeasured variables. Therefore, we conclude that the causal effect of interest is not identifiable in this observational study.

Nonetheless, we can still discuss the scenarios in which the causal effect would be identifiable. To do so, recall $\mathbb{P}_U$ represents the joint distribution of the unmeasured factors influencing the covariates U_W, exposure U_A, and outcome U_Y, respectively. If the following independence assumptions on the distribution $\mathbb{P}_U$ held in our original causal model (Section 2.1), the causal effect would be identifiable:

- $U_A \perp\!\!\!\perp U_Y$ AND
- $U_W \perp\!\!\!\perp U_A$ OR $U_W \perp\!\!\!\perp U_Y$

The first condition specifies that the background variables affecting exposure must be independent from background variables affecting the outcome. This is equivalent to stating there are no unmeasured common causes of the exposure and outcome. The

second condition specifies that the background variables affecting the covariates must be independent from either background variables affecting the exposure or affecting the outcome. In trials where the exposure is completely randomized or randomized within covariate strata, we have $U_A \perp\!\!\!\perp U_Y$ and $U_A \perp\!\!\!\perp U_W$ by design and the observed association between the exposure and outcome can be attributable to the causal effect of interest.

However, in our observational study, we cannot rule out the possibility of unmeasured confounding, as reflected in the original causal model (Section 2.1). As previously discussed, perceived or actual compliance with public health policies and the strength of the state's public health department are examples of unmeasured sources of correlation between the exposure and outcome. Therefore, even after adjusting for the measured confounders W, we cannot be sure that observed association between public masking mandate and COVID-19 cases growth rate are due solely to our causal effect of interest.

Nevertheless, we can still specify a well-defined statistical estimand $\Psi(\mathbb{P})$, which would equal our wished-for causal effect if the above assumptions held. If the set of covariates W were to capture all the common causes of the exposure A and the outcome Y, then we could express the causal rate ratio (Eq 2.4) and the causal rate difference (Eq. 2.5) via the G-computation formula [46]

$$\text{RR} = \frac{\mathbb{E}[\mathbb{E}(Y|A=1,W)]}{\mathbb{E}[\mathbb{E}(Y|A=0,W)]} \tag{2.6}$$

$$\text{RD} = \mathbb{E}[\mathbb{E}(Y|A=1,W)] - \mathbb{E}[\mathbb{E}(Y|A=0,W)] \tag{2.7}$$

where in all cases the outer expectation is over the covariate distribution. In words, $\mathbb{E}[\mathbb{E}(Y|A=1,W)]$ is the expected COVID-19 growth rate, given early implementation and the confounders W, and then standardized with respect to the confounder distribution in the population. $\mathbb{E}[\mathbb{E}(Y|A=0,W)]$ is the expected COVID-19 growth rate, given delayed implementation and the confounders W, and then standardized

with respect to the confounder distribution in the population. A rate ratio less than 1 represents that masking mandate is associated with reductions in COVID-19 growth rate, whereas a rate ratio greater than 1 represents that the masking mandate is associated with increases in COVID-19 growth rates. Likewise, for the rate difference, a point estimate less than 0 means that the mandate was associated with reductions in the growth rate, whereas the opposite is true for an estimate greater than 0.

For the statistical estimands corresponding the rate ratio and rate difference (Eq. 2.6 and Eq. 2.7) to be well-defined, we also need an additional condition on data support, sometimes called the "positivity assumption", "overlap", and the "experimental treatment assignment assumption" [44]:

$$min_{a \in \mathcal{A}} \, \mathbb{P}(A = a | W = w) > 0, \text{ for all possible } w \tag{2.8}$$

where $\mathcal{A} = \{0, 1\}$ represents the set of exposure of interest. This condition ensures that there is a greater than zero probability of each level of exposure for all possible values of adjustment variables.

In our study, the positivity assumption would be violated if the intervention is impossible for states with specific covariate values. This could occur, for example, if all Republican states delayed the masking mandate or if all states with the highest COVID-19 case rates before September 1, 2020 implemented the policy. For the confounders considered, we have no reason to believe the positivity assumption was theoretically violated.

2.5 Statistical estimation & Inference

After reviewing the assumptions needed for identifiability, we recognize that the assumption of no unmeasured confounding is not satisfied in this observational study on masking. Nonetheless, we proceed with estimation and inference based on the

G-computation formula (Eq. 2.6 and Eq. 2.7), which comes as close as possible to the wished-for causal effect given the limitations in the data. To do so, we need to define an estimator which is a function whose input is the observed data $O \sim \mathbb{P}$ and whose output is a real value in the parameter space. Two common estimators are parametric G-computation (a.k.a. the simple substitution estimator) and inverse probability of treatment weighting (IPTW) estimator [6, 53]. However, these single-robust estimators have drawbacks. First, to obtain a consistent point estimate, the simple substitution estimator relies on consistently estimating the conditional mean outcome $\mathbb{E}(Y|A,W)$, and IPTW relies on consistently estimating the propensity score $\mathbb{P}(A=1|W)$. Additionally, implementation of these estimators has traditionally relied on very strong parametric assumptions. Due to their limitations, we employ Targeted Maximum Likelihood Estimation (TMLE), a doubly robust estimator which naturally incorporates machine learning while maintaining the basis for valid statistical inference [6, 53]. In the following, we provide a brief overview of TMLE and then describe our specific implementation for the masking study.

2.5.1 TMLE

Recall the observed data consist of the measured confounders, the exposure indicator, and the outcome: $O = (W, A, Y)$. We can factor the observed data distribution according to temporal ordering: $\mathbb{P}(Y|A,W) * \mathbb{P}(A|W) * \mathbb{P}(W)$. Unlike the single-robust estimators, TMLE requires estimation of all three parts of this distribution. To achieve a consistent point estimate with TMLE, however, we only need that the conditional mean outcome $\mathbb{E}(Y|A,W)$ or the propensity score $\mathbb{P}(A=1|W)$ is consistently estimated; hence TMLE is doubley-robust. Furthermore, if both are consistently estimated at reasonable rates, then TMLE is considered efficient with lowest asymptotic variance among other estimators. Finally, we can utilize machine learn-

ing algorithm to avoid strong modeling assumptions, while obtaining valid statistical inference (i.e. 95% confidence intervals with nominal coverage).

For simplicity, we focus the remainder of this section on using TMLE to estimate the G-computation formula for a single level of the exposure $\psi(a) = \mathbb{E}[\mathbb{E}(Y|A = a, W)]$, while noting we can repeat this process for both levels of the exposure ($a \in \{0, 1\}$) and then take the relevant contrasts to estimate the rate ratio (Eq. 2.6) or rate difference (Eq. 2.7). With a single intervention variable, such as considered in the masking study, we can implement TMLE as follows: [52]

1. Estimate conditional expectation of the outcome, given the exposure and adjustment variables: $\mathbb{E}(Y|A, W)$
2. Estimate the propensity score: $\mathbb{P}(A = 1|W)$
3. Construct a "clever covariate": $\hat{H}(A, W) = \frac{I(A=a)}{\hat{\mathbb{P}}(A=a|W)}$
4. Update the initial estimator
5. Obtain a point estimate by substituting the updated estimates into the parameter mapping and obtain a variance estimate with the estimated influence curve

For our masking study, we first estimate the expected COVID-19 growth rate, given the observed masking mandate and state-level confounders, denoted $\mathbb{E}(Y|A, W)$. To do so, we used Super Learner, an ensemble machine learning method [52]. Super Learner utilizes cross-validation to determine the optimal combination of predictions from a pre-specified collection of algorithms. In more detail, the predictions from various algorithms are assigned different weights based on out-of-sample, performance estimates (i.e. cross-validated risk estimates). By using ensemble machine learning, we aim to avoid unnecessary parametric assumptions and thereby respect the non-parametric statistical model. After fitting Super Learner for the expected outcome,

we then generate estimations for all states, given the exposure-level of interest and measured confounders: $\hat{\mathbb{E}}(Y|A = a, W_i)$ for $i = \{1, \ldots, 50\}$. Here, $\hat{\mathbb{E}}(Y|A = a, W_i)$ represents our estimate of the expected COVID-19 growth rate for state i, given a masking mandate of interest $A = a$ and its covariates W.

In the second step, we again utilize Super Learner to estimate the probability of a state having the masking mandate of interest, given the measured confounders: $\hat{\mathbb{P}}(A = a|W_i)$ for $i = \{1, \ldots, 50\}$. Next, we construct a clever covariate

$$\hat{H}(A_i, W_i) = \frac{\mathbb{I}(A_i = a)}{\hat{\mathbb{P}}(A = a|W_i)} \tag{2.9}$$

which will be used to reduce bias within our initial estimator of the expected outcome $\hat{\mathbb{E}}(Y|A, W)$. The clever covariate is, thus, an indicator of having the exposure of interest ($A = a$), divided by the estimated probability of having that exposure given the measured confounders.[1] The clever covariate, thus, captures how likely each state is to receive the exposure of interest. States with an unlikely exposure-confounder combination have a small denominator and thus a large clever covariate.

With the clever covariate, we can define the updated estimator, denoted with $\hat{\mathbb{E}}^*(Y|A, W)$, as follows for unbounded continuous outcomes and for bounded outcomes, respectively:

$$\hat{\mathbb{E}}^*(Y|A, W) = \hat{\mathbb{E}}(Y|A, W) + \hat{\epsilon} \times \hat{H}(A, W) \tag{2.11}$$

---

[1]We note that if we were only interested in the contrasts on the absolute scale (i.e the risk difference in Eq. 2.7), then we could use the following clever covariate:

$$\hat{H}(A_i, W_i) = \frac{\mathbb{I}(A_i = 1)}{\hat{\mathbb{P}}(A = 1|W_i)} - \frac{\mathbb{I}(A_i = 0)}{\hat{\mathbb{P}}(A = 0|W_i)} \tag{2.10}$$

For early acting states (i.e. those with $A = 1$), the clever covariate is $\frac{1}{\hat{\mathbb{P}}(A=1|W_i)}$ as the second term evaluates to 0. Conversely, for all delayed states (i.e. those with $A = 0$), the first term disappears and the clever covariate becomes $\frac{-1}{\hat{\mathbb{P}}(A=0)|W_i}$.

$$logit[\hat{\mathbb{E}}^*(Y|A,W)] = logit[\hat{\mathbb{E}}(Y|A,W)] + \hat{\epsilon} \times \hat{H}(A,W) \tag{2.12}$$

where $logit[\cdot] = log[\cdot/(1-\cdot)]$. In both cases, $\hat{\epsilon}$ is the fluctuation parameter and captures targeted deviations between our initial estimator $\hat{\mathbb{E}}(Y|A,W)$ and the true conditional expectation $\mathbb{E}(Y|A,W)$. Informally, $\hat{\epsilon}$ is small when there are minimal deviations between the initial estimator and truth and is large when there is substantial residual confounding. Deviations between the estimator and the truth for rare exposure-confounder combinations, as captured by the clever covariate, are given larger weights when updating. More formally, this updating is done to obtain the optimal bias-variance trade-off for the statistical estimand (i.e., the G-computation formula) and to solve the relevant component of the efficient influence function [53].

Updating on the logistic scale (i.e. using Eq. 2.12) is recommended in practice and is applicable to nearly all outcomes in public health and medicine; for further discussion, we refer the reader to [24]. We, therefore, focus the remainder of our presentation on updating on the logistic scale. To estimate ϵ, we can run maximum likelihood logistic regression of the observed outcome Y on the clever covariate $\hat{H}(A,W)$, fixing the initial estimate $\hat{\mathbb{E}}(Y|A,W)$ as an offset and suppressing the intercept. Alternatively, we can also estimate ϵ by running an intercept-only weighted logistic regression of the outcome Y on the initial estimates (as offset) and with weights as the clever covariate. The latter implementation can provide robustness with near or actual positivity violations and is implemented in the `ltmle` package [34]. By substituting in the estimated $\hat{\epsilon}$, we now have the updated estimator of the expected outcome and can generate targeted predictions under the exposure of interest ($A = a$) for all observations $i = \{1, \ldots, n\}$:

$$\hat{\mathbb{E}}^*(Y|a,W_i) = logit^{-1}\Big\{ logit[\hat{\mathbb{E}}(Y|a,W_i)] + \hat{\epsilon} \times \hat{H}(A_i,W_i) \Big\} \tag{2.13}$$

In words, $\hat{\mathbb{E}}^*(Y|a,W_i)$ is the targeted prediction of the COVID-19 growth rate for state i, given the mandate implementation of interest and the confounders.

Finally, we obtain an estimate $\psi(a)$ by averaging our targeted predictions across the sample:

$$\hat{\psi}(a) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}^*(Y|A=a, W_i)$$

Under certain regularity conditions [53], TMLE is asymptotically linear, meaning that we can use the Central Limit Theorem for inference. We obtain a variance estimate with the sample variance of the estimated influence curve divided by sample size:

$$\hat{\sigma}(a)^2 = \frac{Var[\hat{IC}(a)]}{n}$$

where the estimated influence curve for observation i is given by

$$\hat{IC}(a)_i = \hat{H}(A_i, W_i) \times \left[Y - \hat{\mathbb{E}}^*(Y|A=a, W_i)\right] + \hat{\mathbb{E}}^*(Y|A=a, W_i) - \hat{\psi}(a) \quad (2.14)$$

We can then conduct tests of the null hypothesis as well as construct Wald-Type 95% confidence intervals as $\hat{\psi}(a) \pm 1.96 * \hat{\sigma}(a)$. (For sample size <30 independent units, we can use the Student's t-distribution with $n-2$ degrees of freedom in place of the standard normal.)

If we repeat the above process for each level of the exposure of interest $a \in \{0, 1\}$, we can then obtain point estimates on the relative or absolute scale as follows and inference via the Delta Method:

$$\begin{aligned} \hat{RR} &= \frac{\hat{\psi}(1)}{\hat{\psi}(0)} = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}^*(Y|A=1, W_i)}{\frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}^*(Y|A=0, W_i)} \\ \hat{RD} &= \hat{\psi}(1) - \hat{\psi}(0) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}^*(Y|A=1, W_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{\mathbb{E}}^*(Y|A=0, W_i) \end{aligned} \quad (2.15)$$

We note there are a variety of additional implementations, and we can avoid repeating the process by using a two-dimensional clever covariate. However, exposition of all the possible implementations is beyond the scope of this project.

2.5.2 Implementation to Examine the State-level Impact of Masking

Recall for the primary analysis, our exposure of interest is whether a state has implemented public masking mandate prior to the target date of September 1, 2020, and our outcome is the growth rate of COVID-19 cases over the following 7, 14, 21, 30, 45, and 60 days. By varying the follow-up period, we hope to account for the incubation period and potential lags in behavior change as well as evaluate the potential for long-term impact. For estimation of the rate ratio (Eq. 2.6) and rate difference (Eq. 2.7), we implemente TMLE as described in the previous section.

Specifically, we use the `ltmle` package [34] with Super Learner for initial estimation of the expected outcome $\mathbb{E}(Y|A,W)$ and the propensity score $\mathbb{P}(A=1|W)$. For both, we use 10-fold cross-validation and included the following candidate prediction algorithms in our Super Learner library: the empirical mean, Generalized Additive Model (GAM), Recursive Partitioning Model (Rpart), and Extreme Gradient Boosting (XGBoost)[12]. In the primary analyses, we reduce the potential confounder set to 21 variables based on univariate correlations with the outcome $p < 0.05$. Such an approach helps minimize the risk of adjustment for instrumental variables and practical violations of the positivity assumption.

In secondary analyses, we implement the unadjusted estimator as the simple ratio or difference in exposure-specific mean outcomes:

$$\begin{aligned} \hat{RR}_{\text{unadj}} &= \frac{\hat{\mathbb{E}}(Y=1|A=1)}{\hat{\mathbb{E}}(Y=1|A=0)} \\ \hat{RR}_{\text{unadj}} &= \hat{\mathbb{E}}(Y=1|A=1) – \hat{\mathbb{E}}(Y=1|A=0) \end{aligned} \tag{2.16}$$

The unadjusted estimator can be considered a special case of TMLE where the adjustment set $W = \{\}$.

2.6 Sensitivity Analyses

In addition to the primary and secondary analyses, we conduct sensitivity analyses to examine the robustness of our findings. All sensitivity analyses are conducted using the previously described procedures with alterations in variable definition and covariate adjustment sets. First, we include the complete set of potential confounders without screening and implement TMLE as in primary analyses. Second, we adjust the target date from September 1st to when states lifted or eased their initial shelter-in-place order. For states that never had a shelter-in-place order, their target date is assigned to the earliest date shelter-in-place was lifted amongst the other states with the order. This definition allows for flexibility in timing of implementation which could be influenced by states' differential epidemic curve.

CHAPTER 3

RESULTS

In this study, 25 states are considered early implementors of statewide masking mandate while the other 25 are considered delayed. Among the delayed, 7 states never implemented any statewide masking mandate, 12 states had a less strict implementation than our definition of masking mandate, and 6 states implemented the desired mandate after the target date.

Summaries of the covariate sets overall and by exposure group are provided in Table 3.1. There are some notable difference between the groups. In the early implementation states, we observe a higher percentage of Black and Hispanic residents, higher population density, fewer Republican voters, more COVID-19 tests, cases, and deaths leading up September 1, 2020. Furthermore, states with early implementation of the masking mandate were more likely to have implemented a shelter-in-place order and school masking orders.

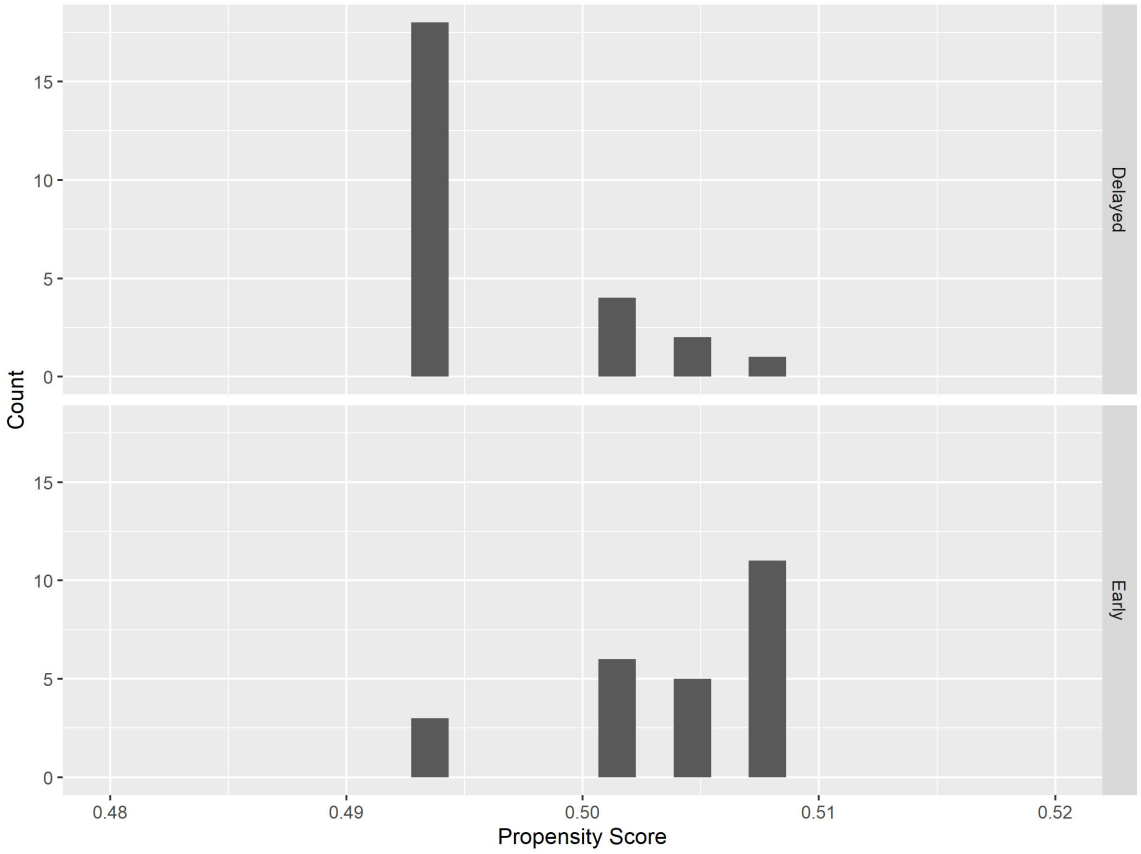
Figure 3.1 shows the estimated propensity score distribution $\hat{\mathbb{P}}(A = 1|W)$ in the primary analyses, which screened out potential confounders not associated in univariate analyses with outcome. We can see that for both states with early and delayed implementation, the estimated propensity score is bounded away from zero. Assuming Super Learner is providing a good estimate of the exposure mechanism, this suggests that the positivity assumption is not violated in practice.

Table 3.1. Summary of baseline characteristics overall and by exposure group. All metrics are given in mean (25% quantile, 75% quantile) unless noted. Number of states that had Republican majority votes, had Shelter-in-place, Gathering Restrictions, or School Masking are given in counts (percentage).

	All (N=50)	Early (N=25)	Delayed (N=25)
Population Demographics			
Black or African American (%)	7 (3.2, 14.2)	9.9 (5.7, 14)	4.2 (2, 15.3)
Hispanic (%)	9.4 (5.1, 13.8)	11.8 (5.1, 17.1)	7 (4.3, 10.6)
Mixed Race (%)	2.2 (1.9, 2.5)	2.1 (1.9, 2.5)	2.2 (2, 2.5)
Caucasian (%)	71.8 (59.2, 79.7)	68.5 (55.6, 75.9)	78.3 (63.1, 82)
Smoker (%)	17.1 (15, 19.3)	17 (14.1, 19.3)	17.2 (15.6, 19.3)
Population Density and Urbanicity			
Population Density (people per km^2)	41.2 (17.7, 84.7)	67.9 (24.6, 160.7)	26.8 (9.6, 62.3)
Urbanicity in 2010 (%)	73.8 (65.1, 87)	81 (73.2, 88)	66.4 (64, 75.1)
Public Transportation Usage (%)	1.4 (0.8, 3.4)	1.8 (0.9, 5.8)	1.2 (0.8, 2)
Political leading			
Republican	30 (60%)	10 (40%)	20 (80%)
COVID-19 related data			
Confirmed Positive 30 days before	51022.5 (19576.8, 105478)	60658 (27812, 118458)	45492 (6854, 91782)
Confirmed Positive 14 days before	64350 (23603.2, 120281.2)	73207 (35167, 125579)	52951 (8765, 108282)
Confirmed Positive 7 days before	68560 (24749, 124125.5)	79206 (38401, 130035)	55800 (10211, 114635)
Deaths 14 days before	1414 (399, 4309.8)	3165 (723, 7499)	989 (164, 1936)
Deaths 7 days before	1534 (426.5, 4346.2)	3241 (750, 7605)	1052 (187, 2037)
Total Tests 14 days before	871097 (397771, 1682918.5)	1291302 (632652, 1946517)	775000 (310009, 1385832)
Total Tests 7 days before	913548 (430918.8, 1804195.2)	1380050 (673704, 2157864)	830831 (333831, 1495014)
Implemented Shelter-in-place	43 (86%)	24 (96%)	19 (76%)
Implemented Gathering Restrictions	49 (98%)	25 (100%)	24 (96%)
Implemented School Masking	20 (40%)	14 (56%)	6 (24%)
Mobility Change[1]14 days before(%)	8.5 (7, 10)	9 (7, 11)	7 (6, 10)
Mobility Change 7 days before(%)	9 (7, 11)	9 (8, 11)	8 (6, 10)

[1] Mobility Change is calculated based on mobility trends for residential places. Baseline value for specific days of the week were obtained in a 5-week period between Jan 3 to Feb 6, 2020.[22]

Figure 3.1. Estimated Propensity Score Distribution $\hat{\mathbb{P}}(A = 1|W)$



3.1 Main Results

As discussed in Chapter 2, our primary analysis relies on TMLE and Super Learner with screening of the potential confounders based on univariate associations with the outcome and our secondary analysis is unadjusted. Table 3.2 provides estimated growth rate ratio estimated from the three approaches. Focusing on the primary approach, which adjusts for key differences between states, we see that after one week, the growth rates begin to diverge with early masking implementation versus delayed implementations. Specifically, the expected growth rate is 1.05 (95% CI: 1.04-1.06) with early implementation and 1.07 (95% CI: 1.05-1.08) with delayed implementation for a rate ratio of 0.98 (95% CI: 0.97-0.99). Within two weeks of September 1, 2020,

the rate ratio reduces to 0.96 (95% CI: 0.95-0.98) and continues to decline over time (Figure 3.2). After 45 and 60 days, the estimated rate ratios are 0.89 (95% CI: 0.85-0.93) and 0.84 (95% CI: 0.78-0.91), respectively. The two-month outcome, thus represents a 16% reduction in COVID-19 growth rates with state-level early masking orders.

In unadjusted secondary analyses, the estimated associations were considerably larger, especially at longer follow-up times (Table 3.2). At 7, 14, 45, and 60 days, the rate ratios were 0.97 (95% CI: 0.95-0.99), 0.94 (95% CI: 0.91-0.97), 0.79 (95% CI: 0.7-0.88), and 0.72 (95% CI: 0.60-0.85), respectively. The two-month association, thus, corresponds to a 18% reduction in the COVID-19 case growth rate with state-level masking mandates. Exaggerated associations are expected due to failure to adjust for measured differences between states.

In all approaches, the upper bounds on the 95% confidence intervals exclude the null value of one, suggesting both short term and longer term impacts of masking mandates on COVID-19 case growth rates. From Figure 3.2, a clear trend of increasing associations (i.e. decreasing rate ratios) can be observed with increasing follow-up. For rate differences, we observe similar results (Table A.2 and Figure A.1).

Figure 3.2. Estimated Growth Rate Ratios

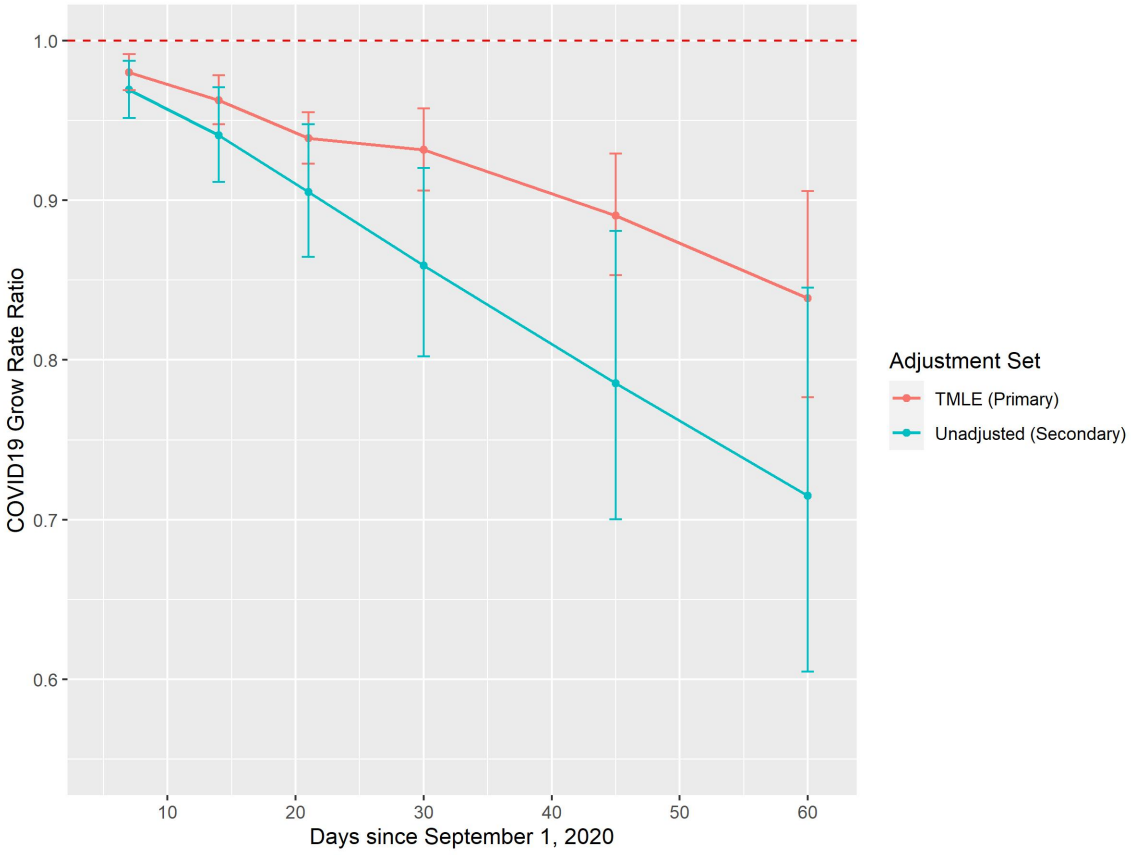


Table 3.2. Growth Rate Ratio

Adjustment	Outcome at	Txt(95% CI)	Con(95% CI)	RR(95% CI)	RR p-value
Primary Analysis	7 days	1.05 (1.04, 1.06)	1.07 (1.05, 1.08)	0.98 (0.97, 0.99)	9.4e-04
Primary Analysis	14 days	1.1 (1.08, 1.11)	1.14 (1.11, 1.16)	0.96 (0.95, 0.98)	1.8e-05
Primary Analysis	21 days	1.15 (1.13, 1.17)	1.22 (1.19, 1.26)	0.94 (0.92, 0.96)	2.1e-09
Primary Analysis	30 days	1.24 (1.2, 1.28)	1.33 (1.27, 1.39)	0.93 (0.91, 0.96)	4.6e-06
Primary Analysis	45 days	1.42 (1.35, 1.49)	1.59 (1.48, 1.71)	0.89 (0.85, 0.93)	1.6e-06
Primary Analysis	60 days	1.67 (1.56, 1.77)	1.99 (1.77, 2.21)	0.84 (0.78, 0.91)	3.2e-05
Secondary Analysis	7 days	1.04 (1.03, 1.05)	1.07 (1.06, 1.09)	0.97 (0.95, 0.99)	1.3e-03
Secondary Analysis	14 days	1.08 (1.06, 1.1)	1.15 (1.12, 1.18)	0.94 (0.91, 0.97)	2.9e-04
Secondary Analysis	21 days	1.13 (1.1, 1.16)	1.25 (1.2, 1.29)	0.91 (0.86, 0.95)	6.7e-05
Secondary Analysis	30 days	1.19 (1.15, 1.23)	1.39 (1.3, 1.47)	0.86 (0.8, 0.92)	5.2e-05
Secondary Analysis	45 days	1.33 (1.27, 1.4)	1.7 (1.52, 1.87)	0.79 (0.7, 0.88)	1.0e-04
Secondary Analysis	60 days	1.54 (1.44, 1.65)	2.16 (1.83, 2.49)	0.72 (0.6, 0.85)	2.0e-04

3.2 Additional Sensitivity Analyses

Our main sensitivity analysis uses TMLE with Super Learner without screening. Adjusting for all potential confounders, this analysis yielded very similar results to the primary analysis. From Table A.1, at 7, 14, 45, and 60 days, the rate ratios were 0.97 (95% CI: 0.96-0.99), 0.95 (95% CI: 0.92-0.97), 0.85 (95% 0.79-0.91), and 0.80 (95% CI: 0.71-0.9), respectively.

In sensitivity analyses where we change the target date to the date that the state-level shelter-in-place order was first lifted, we observe similar patterns as the primary analysis. However, the estimated associations are overall more conservative. From Table A.4, with the primary approach, the estimated rate ratios are 0.98 (95% CI: 0.98-0.99), 0.98 (95% CI: 0.97-0.99), 0.91 (95% CI: 0.87-095), and 0.88 (95% CI: 0.84-0.93) for 7, 14, 45, and 60 days, respectively. Thus, after two months of lifting shelter-in-place, early masking mandates were associated with 12% decrease in state-level, COVID-19 case growth rates.

With no adjustment, we observe exaggerated estimates once again: 0.96 (95% CI: 0.94-0.98), 0.93 (95% CI: 0.90-0.97), 0.80 (95% CI: 0.73-0.89), and 0.76(95% CI: 0.65-0.89) for 7, 14, 45, and 60 days, respectively. The unadjusted relative reduction in COVID-19 case growth rates at two-month is estimated to be 24%. Figure A.2 shows a decreasing trend line for all adjustment sets. Additional information is provided in the Supplemental Materials.

CHAPTER 4

DISCUSSION

In this study, our objective was to evaluate the effect of public masking mandate on COVID-19 case growth rates in U.S. states. Throughout the pandemic, thus far, the wide variation in COVID-19 related policies has provided us a natural experiment to evaluate impact. Utilizing a causal inference framework [6], we estimated the association of having a state-level, public masking mandate in place prior to September 1, 2020, the approximate start of the school year, on the subsequent growth of COVID-19, defined as the total case count a set number of days after September 1, divided by the total case count on September 1. We critically examined all assumptions needed justify a causal claim and avoided restrictive modeling assumptions by using non-parametric estimators.

We found that early implementation of public masking mandate was significantly associated with reductions in state-level COVID-19 case growth rates. We evaluated the outcomes for 7 to 60 days after the target date and observed that association increased meaningfully over time. After one week, the reduction in expected growth rates with early versus delayed masking mandate was already observable. By the second week, there is a 4% reduction (RR: 0.96; 95%CI: 0.95-0.98) and by 2 months, the reduction increased to 16% (RR: 0.84; 95%CI: 0.78-0.91). Sensitivity analyses (e.g.: varying the analytic approach, the scale of inference, or the target date) yielded similar results. The largest association was observed with unadjusted analyses, suggesting that early implementation of a state-level masking mandate was associated with 28% reduction in the expected growth rate (RR: 0.72; 95% CI: 0.60-0.85).

Our results were consistent with prior studies, reviewed in the Introduction. When compared with results generated from traditional modeling approaches, our associations were more muted. Larger associations from other approaches may be due to differences in the outcome definition, the unit of analysis (counties versus state), adjustment sets, or violations of the positivity assumption. While these alternative approaches provide informative and meaningful results, they tend to presume the assumptions needed for valid causal and statistical inference hold. Failure to critically evaluate the needed assumptions may yield misleading conclusions and thereby misguidance to policy makers and public health officials.

Our estimates were similar to other studies using "causal" approaches, despite varying definitions of the outcome or unit of analysis. Our results suggested a 4% absolute reduction in expected state-level growth rates after 14 days, while Lyu and Wehby estimated 1.4% absolute reduction in expected county-level growth rates for over 11-15 days [36]. Although Chernozhukov et al. used cumulative scale for the outcome, our estimates were once again similar. Furthermore, all studies suggested that the association increased over time [14].

Altogether, our study builds on growing evidence that public masking mandates are associated with reductions of COVID-19 spread. Prior "causal" studies have largely focused on county-level analyses and employed difference-in-differences method, which requires strong assumptions that might be difficult to hold in practice [2]. To the best of our knowledge, this is the first work employing the Causal Roadmap and TMLE to evaluate impact at the state-level. As a result, this paper fills in the gap between modeling studies and other existing causal inference studies.

Nonetheless, there are several limitations to this work. First, as nature of an observation study, our findings may be subject to bias due to unmeasured confounding. Secondly, while deaths related to COVID-19 are commonly studied outcomes, our analyses were restricted to growth rates of confirmed cases. There is more un-

certainty associated with confirmed cases, including test accuracy, accessibility, and time lag. However, our study focused on the period from September 1- November 1, 2020 when testing was widely available. We also note that recent evidence suggests that starting September 1, 2020, every 100-cases would be expected to result in approximately 1.6 deaths 22-days later [35]. Therefore, we expect the state-level mask mandate, studied here, also to impact growth in COVID-19 death rates. Along the same lines, we focused on implementation of the strictest public masking mandate. It would be interesting to see if and how the impact varies by type of masking mandate. Lastly, issuing a public masking mandate does not guarantee perfect masking compliance. As modeling studies have suggested, compliance can play an essential role in determining the effect of the masking mandate. Our work evaluated the state-level impact of the mandate, without formally accounting for compliance. These limitations encourage us to extend this work to other directions. In the future, we hope to examine the effect of less restrictive masking mandate as well as with consideration of mask compliance (possibly through instrumental variable analysis). As the pandemic continues, we must carefully consider the importance of masking mandate and the downside of ignoring a relatively cost-effective intervention. From the primary analysis, this study suggests that in both the short and longer term, state-level public masking mandates were associated with lower COVID-19 growth rates. The potential benefit of a masking mandate in long term must not be overlooked. State-level governments should implement universal masking policies as soon as possible and encourage compliance through public health messaging and education. A timely masking mandate may help prevent exponential growth, seen during outbreaks and surges. Although promising vaccines have been developed and distribution has started, there is still an extended period before herd immunity can be achieved. Before then, non-pharmaceutical interventions, such as public masking, are essential. Given the minimal impact on the economy and ease of implementation as compli-

ment to other strategies, masking policies are promising public health strategies to mitigate further spread of COVID-19.

APPENDIX

SUPPLEMENTAL MATERIALS

Table A.1. Growth Rate Ratio (TMLE without screening) - Sensitivity Analysis

Sensitivity Analysis	7 days	1.04 (1.03, 1.05)	1.07 (1.06, 1.09)	0.97 (0.96, 0.99)	8.8e-04
Sensitivity Analysis	14 days	1.09 (1.07, 1.1)	1.15 (1.12, 1.17)	0.95 (0.92, 0.97)	8.3e-05
Sensitivity Analysis	21 days	1.14 (1.11, 1.16)	1.23 (1.19, 1.27)	0.92 (0.9, 0.95)	8.1e-06
Sensitivity Analysis	30 days	1.22 (1.18, 1.25)	1.35 (1.29, 1.42)	0.9 (0.86, 0.94)	9.0e-06
Sensitivity Analysis	45 days	1.38 (1.32, 1.44)	1.62 (1.48, 1.76)	0.85 (0.79, 0.91)	3.3e-05
Sensitivity Analysis	60 days	1.61 (1.51, 1.72)	2.03 (1.76, 2.29)	0.8 (0.71, 0.9)	3.0e-04

Figure A.1. Growth Rate Difference

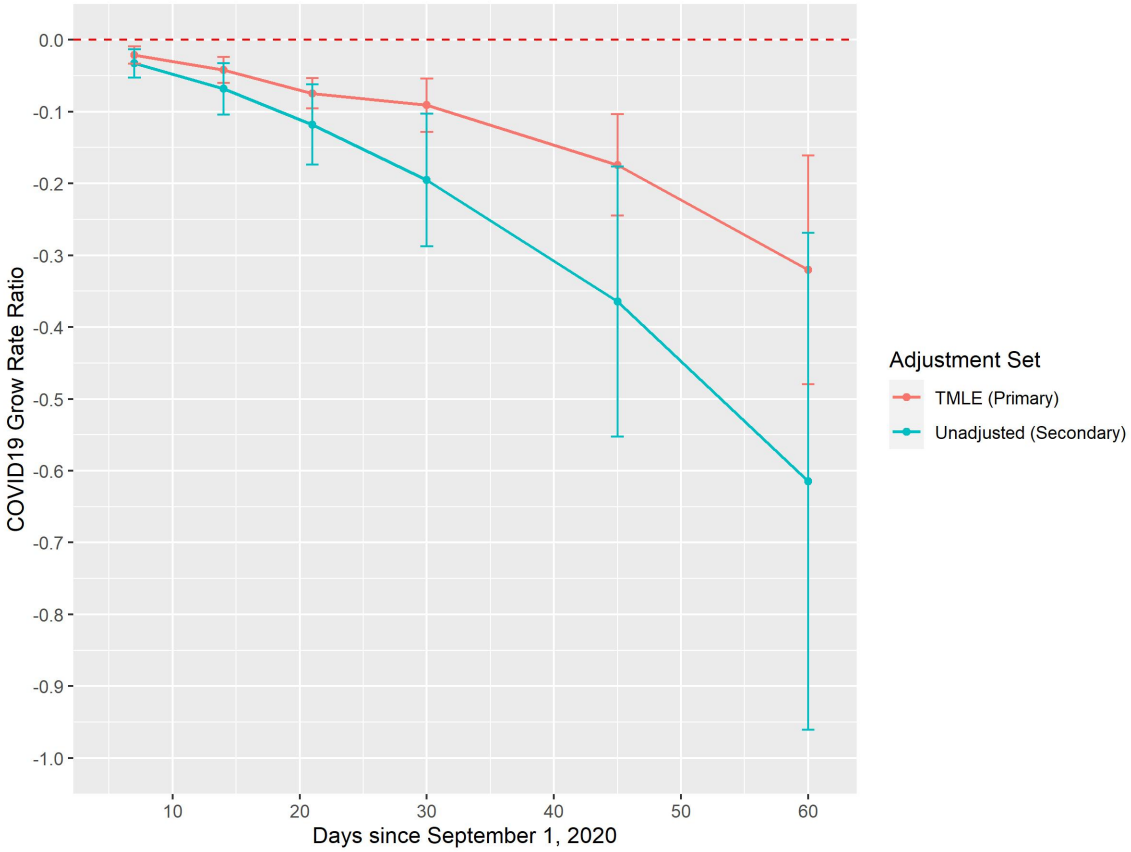


Table A.2. Growth Rate Difference

Adjustment	Outcome at	Txt(95% CI)	Con(95% CI)	RD(95% CI)	RD p-value
Primary Analysis	7 days	1.05 (1.04, 1.06)	1.07 (1.05, 1.08)	-0.02 (-0.03, -0.01)	9.9e-04
Primary Analysis	14 days	1.1 (1.08, 1.11)	1.14 (1.11, 1.16)	-0.04 (-0.06, -0.02)	2.3e-05
Primary Analysis	21 days	1.15 (1.13, 1.17)	1.22 (1.19, 1.26)	-0.07 (-0.1, -0.05)	4.6e-09
Primary Analysis	30 days	1.24 (1.2, 1.28)	1.33 (1.27, 1.39)	-0.09 (-0.13, -0.05)	1.0e-05
Primary Analysis	45 days	1.42 (1.35, 1.49)	1.59 (1.48, 1.71)	-0.17 (-0.24, -0.1)	8.5e-06
Primary Analysis	60 days	1.67 (1.56, 1.77)	1.99 (1.77, 2.21)	-0.32 (-0.48, -0.16)	1.9e-04
Secondary Analysis	7 days	1.04 (1.03, 1.05)	1.07 (1.06, 1.09)	-0.03 (-0.05, -0.01)	1.4e-03
Secondary Analysis	14 days	1.08 (1.06, 1.1)	1.15 (1.12, 1.18)	-0.07 (-0.1, -0.03)	3.4e-04
Secondary Analysis	21 days	1.13 (1.1, 1.16)	1.25 (1.2, 1.29)	-0.12 (-0.17, -0.06)	9.3e-05
Secondary Analysis	30 days	1.19 (1.15, 1.23)	1.39 (1.3, 1.47)	-0.2 (-0.29, -0.1)	9.5e-05
Secondary Analysis	45 days	1.33 (1.27, 1.4)	1.7 (1.52, 1.87)	-0.36 (-0.55, -0.18)	3.0e-04
Secondary Analysis	60 days	1.54 (1.44, 1.65)	2.16 (1.83, 2.49)	-0.61 (-0.96, -0.27)	8.2e-04
Sensitivity Analysis	7 days	1.04 (1.03, 1.05)	1.07 (1.06, 1.09)	-0.03 (-0.05, -0.01)	9.3e-04
Sensitivity Analysis	14 days	1.09 (1.07, 1.1)	1.15 (1.12, 1.17)	-0.06 (-0.09, -0.03)	1.0e-04
Sensitivity Analysis	21 days	1.14 (1.11, 1.16)	1.23 (1.19, 1.27)	-0.09 (-0.13, -0.05)	1.2e-05
Sensitivity Analysis	30 days	1.22 (1.18, 1.25)	1.35 (1.29, 1.42)	-0.13 (-0.19, -0.08)	1.9e-05
Sensitivity Analysis	45 days	1.38 (1.32, 1.44)	1.62 (1.48, 1.76)	-0.24 (-0.36, -0.13)	1.1e-04
Sensitivity Analysis	60 days	1.61 (1.51, 1.72)	2.03 (1.76, 2.29)	-0.41 (-0.65, -0.18)	1.1e-03

Table A.3. Background characteristics summary (Alternative Target Date) overall and by intervention. Number of states that had Republican majority votes, had Shelter-in-place, Gathering Restrictions, or School Masking are given in counts (percentage).

	All (N=50)	Early (N=8)	Delayed (N=42)
Population Demographics			
Black or African American (%)	7 (3.2, 14.2)	8.4 (4.7, 14.1)	6.7 (3.2, 14.8)
Hispanic (%)	9.4 (5.1, 13.8)	15.8 (11.2, 17.6)	7.4 (4.5, 12.3)
Mixed Race (%)	2.2 (1.9, 2.5)	2 (2, 2.2)	2.2 (1.9, 2.7)
Caucasian (%)	71.8 (59.2, 79.7)	64.6 (59.9, 71.7)	74.4 (59.2, 80.9)
Smoker (%)	17.1 (15, 19.3)	15.2 (14, 17.1)	17.3 (15.6, 20.6)
Population Density and Urbanicity			
Population Density (people per km^2)	41.2 (17.7, 84.7)	174.5 (71.1, 298.7)	35.6 (17.5, 68.6)
Urbanicity in 2010 (%)	73.8 (65.1, 87)	88 (81.8, 89)	72.8 (64.2, 83.3)
Public Transportation Usage (%)	1.4 (0.8, 3.4)	3.3 (1.6, 9.6)	1.2 (0.8, 2.6)
Political leading			
Republican (%)	30 (60%)	0 (0%)	30 (71.4%)
COVID-19 related data			
Confirmed Positive 30 days before	51022.5 (19576.8, 105478)	35413 (18003.8, 134649.5)	53893.5 (22433.2, 92720.2)
Confirmed Positive 14 days before	64350 (23603.2, 120281.2)	37417 (19667.2, 145422.5)	67739 (28597.2, 110251.5)
Confirmed Positive 7 days before	68560 (24749, 124125.5)	38287.5 (20269.5, 151036.8)	71995.5 (30564.2, 117055.2)
Death 14 days before	1414 (399, 4309.8)	2740 (690.5, 8204.5)	1414 (376, 3528.8)
Death 7 days before	1534 (426.5, 4346.2)	2751 (713.2, 8334.5)	1534 (404.2, 3590.5)
Total Test 14 days before	871097 (397771, 1682918.5)	837892 (403138.8, 2320038.8)	871097 (397771, 1640971.2)
Total Test 7 days before	913548 (430918.8, 1804195.2)	905666.5 (440268.8, 2564014.8)	913548 (430918.8, 1764688.5)
Implemented SAH	43 (86%)	8 (100%)	35 (83.3%)
Implemented Gather Restriction	49 (98%)	8 (100%)	41 (97.6%)
Implemented School Masking	20 (40%)	5 (62.5%)	15 (35.7%)
Mobility change 14 days before(%)	8.5 (7, 10)	9.5 (9, 11)	8 (6.2, 10)
Mobility change 7 days before(%)	9 (7, 11)	10 (9, 11)	8 (7, 10.8)

Figure A.2. Sensitivity Analysis (Alternative Target Date)- Growth Rate Ratio

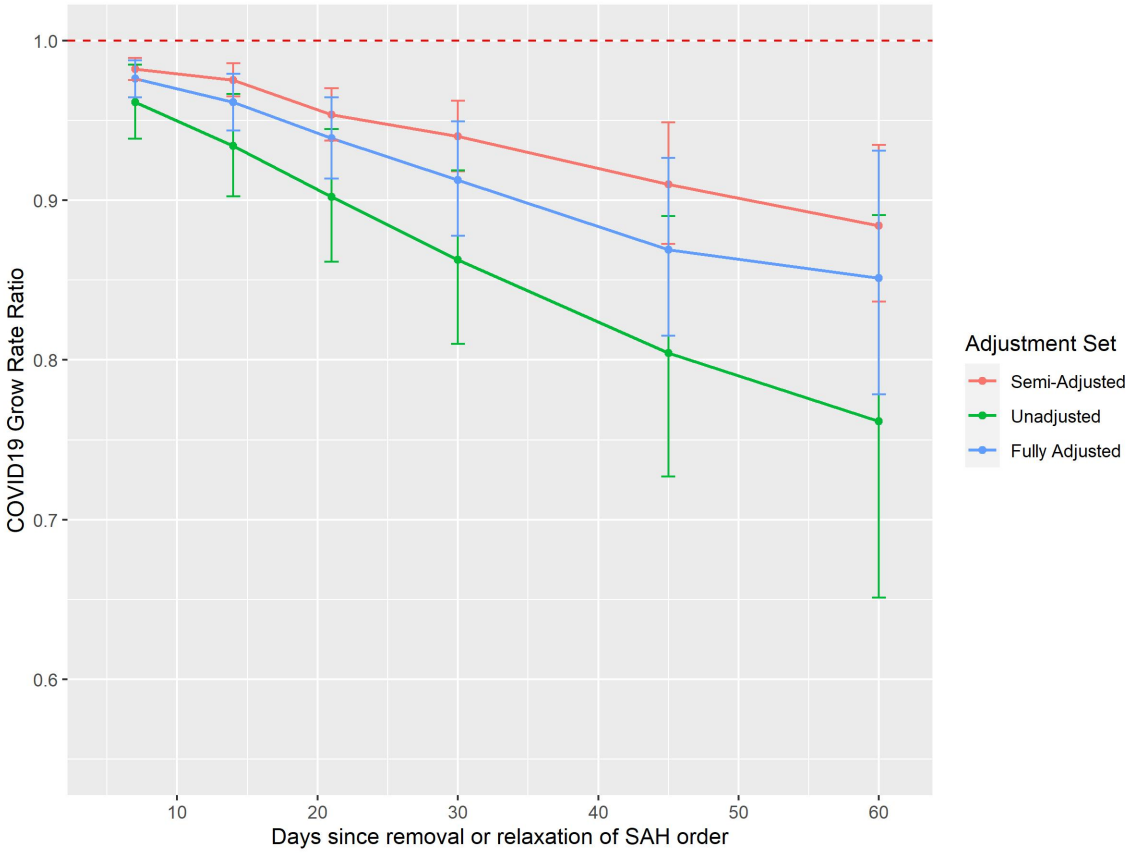


Table A.4. Sensitivity Analysis (Target Date of SAH Relaxation) - Growth Rate Ratio

Adjustment	Outcome at	Txt(95% CI)	Con(95% CI)	RR(95% CI)	RR p-value
Semi-Adjusted	7 days	1.04 (1.03, 1.05)	1.06 (1.05, 1.07)	0.98 (0.98, 0.99)	6.7e-06
Semi-Adjusted	14 days	1.09 (1.08, 1.11)	1.12 (1.1, 1.14)	0.98 (0.97, 0.99)	2.1e-05
Semi-Adjusted	21 days	1.14 (1.12, 1.15)	1.19 (1.16, 1.22)	0.95 (0.94, 0.97)	1.2e-06
Semi-Adjusted	30 days	1.22 (1.19, 1.25)	1.3 (1.24, 1.35)	0.94 (0.92, 0.96)	3.2e-06
Semi-Adjusted	45 days	1.39 (1.34, 1.44)	1.53 (1.42, 1.64)	0.91 (0.87, 0.95)	4.1e-05
Semi-Adjusted	60 days	1.66 (1.56, 1.75)	1.87 (1.68, 2.07)	0.88 (0.84, 0.93)	4.9e-05
Unadjusted	7 days	1.02 (1, 1.05)	1.06 (1.05, 1.07)	0.96 (0.94, 0.98)	1.9e-03
Unadjusted	14 days	1.05 (1.02, 1.08)	1.13 (1.11, 1.15)	0.93 (0.9, 0.97)	2.2e-04
Unadjusted	21 days	1.09 (1.05, 1.13)	1.21 (1.17, 1.24)	0.9 (0.86, 0.94)	4.3e-05
Unadjusted	30 days	1.14 (1.09, 1.19)	1.32 (1.26, 1.38)	0.86 (0.81, 0.92)	2.2e-05
Unadjusted	45 days	1.26 (1.18, 1.34)	1.56 (1.44, 1.69)	0.8 (0.73, 0.89)	7.7e-05
Unadjusted	60 days	1.47 (1.31, 1.62)	1.92 (1.7, 2.15)	0.76 (0.65, 0.89)	1.0e-03
Fully-Adjusted	7 days	1.04 (1.03, 1.05)	1.06 (1.05, 1.07)	0.98 (0.96, 0.99)	1.6e-04
Fully-Adjusted	14 days	1.08 (1.07, 1.1)	1.13 (1.11, 1.15)	0.96 (0.94, 0.98)	8.5e-05
Fully-Adjusted	21 days	1.13 (1.11, 1.14)	1.2 (1.17, 1.23)	0.94 (0.91, 0.96)	2.2e-05
Fully-Adjusted	30 days	1.19 (1.17, 1.22)	1.31 (1.25, 1.36)	0.91 (0.88, 0.95)	2.5e-05
Fully-Adjusted	45 days	1.35 (1.31, 1.39)	1.55 (1.43, 1.66)	0.87 (0.82, 0.93)	5.9e-05
Fully-Adjusted	60 days	1.61 (1.53, 1.69)	1.89 (1.69, 2.1)	0.85 (0.78, 0.93)	7.1e-04

Figure A.3. Sensitivity Analysis (Alternative Target Date) - Growth Rate Difference

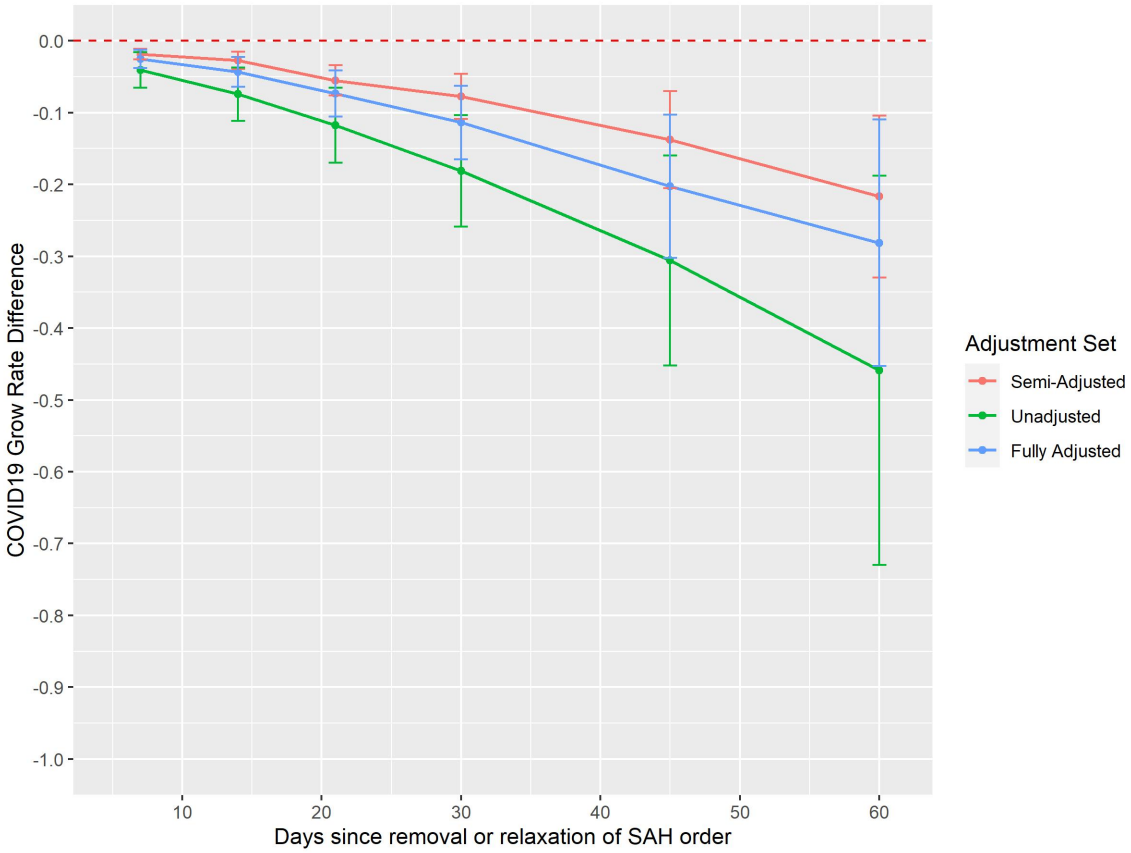


Table A.5. Sensitivity Analysis (Target Date of SAH Relaxation) - Relative Rate Difference

Adjustment	Outcome at	Txt(95% CI)	Con(95% CI)	RD(95% CI)	RD p-value
Semi-Adjusted	7 days	1.04 (1.03, 1.05)	1.06 (1.05, 1.07)	-0.02 (-0.03, -0.01)	7.9e-06
Semi-Adjusted	14 days	1.09 (1.08, 1.11)	1.12 (1.1, 1.14)	-0.03 (-0.04, -0.02)	2.9e-05
Semi-Adjusted	21 days	1.14 (1.12, 1.15)	1.19 (1.16, 1.22)	-0.06 (-0.08, -0.03)	2.6e-06
Semi-Adjusted	30 days	1.22 (1.19, 1.25)	1.3 (1.24, 1.35)	-0.08 (-0.11, -0.05)	1.0e-05
Semi-Adjusted	45 days	1.39 (1.34, 1.44)	1.53 (1.42, 1.64)	-0.14 (-0.21, -0.07)	1.6e-04
Semi-Adjusted	60 days	1.66 (1.56, 1.75)	1.87 (1.68, 2.07)	-0.22 (-0.33, -0.1)	3.3e-04
Unadjusted	7 days	1.02 (1, 1.05)	1.06 (1.05, 1.07)	-0.04 (-0.07, -0.02)	1.7e-03
Unadjusted	14 days	1.05 (1.02, 1.08)	1.13 (1.11, 1.15)	-0.07 (-0.11, -0.04)	1.9e-04
Unadjusted	21 days	1.09 (1.05, 1.13)	1.21 (1.17, 1.24)	-0.12 (-0.17, -0.07)	3.8e-05
Unadjusted	30 days	1.14 (1.09, 1.19)	1.32 (1.26, 1.38)	-0.18 (-0.26, -0.1)	2.4e-05
Unadjusted	45 days	1.26 (1.18, 1.34)	1.56 (1.44, 1.69)	-0.31 (-0.45, -0.16)	1.1e-04
Unadjusted	60 days	1.47 (1.31, 1.62)	1.92 (1.7, 2.15)	-0.46 (-0.73, -0.19)	1.3e-03
Fully-Adjusted	7 days	1.04 (1.03, 1.05)	1.06 (1.05, 1.07)	-0.03 (-0.04, -0.01)	1.7e-04
Fully-Adjusted	14 days	1.08 (1.07, 1.1)	1.13 (1.11, 1.15)	-0.04 (-0.06, -0.02)	1.0e-04
Fully-Adjusted	21 days	1.13 (1.11, 1.14)	1.2 (1.17, 1.23)	-0.07 (-0.11, -0.04)	3.2e-05
Fully-Adjusted	30 days	1.19 (1.17, 1.22)	1.31 (1.25, 1.36)	-0.11 (-0.17, -0.06)	4.8e-05
Fully-Adjusted	45 days	1.35 (1.31, 1.39)	1.55 (1.43, 1.66)	-0.2 (-0.3, -0.1)	1.7e-04
Fully-Adjusted	60 days	1.61 (1.53, 1.69)	1.89 (1.69, 2.1)	-0.28 (-0.45, -0.11)	1.8e-03

BIBLIOGRAPHY

[1] Abedi, Vida, Olulana, Oluwaseyi, Avula, Venkatesh, Chaudhary, Durgesh, Khan, Ayesha, Shahjouei, Shima, Li, Jiang, and Zand, Ramin. Racial, economic, and health inequality and covid-19 infection in the united states. *Journal of racial and ethnic health disparities* (Sep 1, 2020), 1–11.

[2] Angrist, Joshua D., and Pischke, Jörn-Steffen. *Mostly Harmless Econometrics*. Princeton University Press, Princeton, NJ, USA, 2008.

[3] Asadi, Sima, Cappa, Christopher D., Barreda, Santiago, Wexler, Anthony S., Bouvier, Nicole M., and Ristenpart, William D. Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities. *Scientific Reports 10*, 1 (-09-24 2020), 1–13.

[4] Babino, Andres, and Magnasco, Marcelo O. Masks and covid-19: a causal framework for imputing value to public-health interventions.

[5] Bai, Nina. Still confused about masks? here's the science behind how face masks prevent coronavirus, June 26, 2020.

[6] Balzer, Laura, Petersen, Maya, and van der Laan, Mark. *Tutorial for Causal Inference*. Handbook of Big Data. Chapman and Hall/CRC, 2016, pp. 377–402.

[7] Balzer, Laura B., and Dominici, Francesca. Randomization versus real-world evidence. *The New England journal of medicine 383*, 4 (Jul 23, 2020), e21.

[8] Balzer, Laura B., and Whitcomb, Brian W. Coronavirus deaths in san francisco vs. new york: What causes such big differences in cities' tolls?, June 2, 2020.

[9] Bureau of Transportation Statistics, . Commute mode statistics.

[10] Centers for Disease Control and Prevention. CDC COVID Data Tracker, Nov 2020.

[11] Centers for Disease Control and Prevention. Health equity considerations and racial and ethnic minority groups, -02-11 2020.

[12] Chen, Tianqi, and Guestrin, Carlos. Xgboost. KDD '16, ACM, pp. 785–794.

[13] Cheng, Kar Keung, Lam, Tai Hing, and Leung, Chi Chiu. Wearing face masks in the community during the covid-19 pandemic: altruism and solidarity. *The Lancet (British edition)* (Apr 16, 2020).

[14] Chernozhukov, Victor, Kasahara, Hiroyuki, and Schrimpf, Paul. Causal impact of masks, policies, behavior on early covid-19 pandemic in the u.s. *Journal of Econometrics 220*, 1 (January 1, 2021), 23–62.

[15] Courtemanche, Charles, Garuccio, Joseph, Le, Anh, Pinkston, Joshua, and Yelowitz, Aaron. Strong social distancing measures in the united states reduced the covid-19 growth rate. *Health Affairs 39*, 7 (May 14, 2020), 1237–1246.

[16] Delphi Group, Alex Reinhart. New and improved covid symptom survey tracks testing and mask-wearing, -10-12 2020.

[17] Eikenberry, Steffen E., Mancuso, Marina, Iboi, Enahoro, Phan, Tin, Eikenberry, Keenan, Kuang, Yang, Kostelich, Eric, and Gumel, Abba B. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. *Infectious Disease Modelling 5* (2020), 293–308.

[18] Fisher, Kiva A., Barile, John P., Guerin, Rebecca J., Esschert, Kayla L. Vanden, Jeffers, Alexiss, Tian, Lin H., Garcia-Williams, Amanda, Gurbaxani, Brian, Thompson, William W., and Prue, Christine E. Factors associated with cloth face covering use among adults during the covid-19 pandemic - united states, april and may 2020. *MMWR. Morbidity and mortality weekly report 69*, 28 (Jul 17, 2020), 933–937.

[19] Friedman, Uri. Face masks are in, -04-02 2020.

[20] Friedrich, Sarah, and Friede, Tim. Causal inference methods for small non-randomized studies: Methods and an application in covid-19. *Contemporary Clinical Trials 99* (-12 2020), 106213.

[21] Fullman, Nancy, Bang-Jensen, Bree, Reinke, Grace, Magistro, Beatrice, Castellano, Rachel, Erickson, Megan, Amano, Kenya, Wilkerson, John, and Adolph, Christopher. State-level social distancing policies in response to covid-19 in the us.

[22] Google. Google covid-19 community mobility reports, Mar 29, 2020.

[23] Greenhalgh, Trisha, Schmid, Manuel B., Czypionka, Thomas, Bassler, Dirk, and Gruer, Laurence. Face masks for the public during the covid-19 crisis. *BMJ 369* (Apr 9, 2020), m1435.

[24] Gruber, Susan, and van der Laan, Mark J. A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *The international journal of biostatistics 6*, 1 (Aug 1, 2010), Article 26.

[25] Howard, Jeremy, Huang, Austin, Li, Zhiyuan, Tufekci, Zeynep, Zdimal, Vladimir, van der Westhuizen, Helene-Mari, von Delft, Arne, Price, Amy, Fridman, Lex, Tang, Lei-Han, Tang, Viola, Watson, Gregory L., Bax, Christina E., Shaikh, Reshama, Questier, Frederik, Hernandez, Danny, Chu, Larry F., Ramirez, Christina M., and Rimoin, Anne W. An evidence review of face masks against covid-19. *Proceedings of the National Academy of Sciences - PNAS 118*, 4 (Jan 26, 2021), 1.

[26] IHME COVID-19 Forecasting Team., Reiner, R.C., Barber, R.M. et al. Modeling covid-19 scenarios for the united states. *Nature medicine 27*, 1 (Jan 1, 2021), 94–105.

[27] Institute for Health Metrics and Evaluation. Covid-19 projections, March 25, 2020.

[28] Iowa Community Indicators Program. Urban percentage of the population for states, historical.

[29] Kai, De, Goldstein, Guy-Philippe, Morgunov, Alexey, Nangalia, Vishal, and Rotkirch, Anna. Universal masking is urgent in the covid-19 pandemic: Seir and agent based models, empirical validation, policy recommendations.

[30] Kaiser Family Foundation. State health facts custom report.

[31] Kates, Jennifer, Michaud, Josh, and Tolbert, Jennifer. Stay-at-home orders to fight covid-19 in the united states: The risks of a scattershot approach, Apr 5, 2020.

[32] Kim, Allen, Andrew, Scottie, and Froio, James. These are the states requiring people to wear masks when out in public, August 17, 2020.

[33] Lauer, Stephen A., Grantz, Kyra H., Bi, Qifang, Jones, Forrest K., Zheng, Qulu, Meredith, Hannah R., Azman, Andrew S., Reich, Nicholas G., and Lessler, Justin. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine* (-3-10 2020).

[34] Lendle, Samuel D., Schwab, Joshua, Petersen, Maya L., and van der Laan, Mark J. ltmle: An r package implementing targeted minimum loss-based estimation for longitudinal data. *Journal of statistical software 81*, 1 (2017), 1–21.

[35] Leonhardt, David. Charting the virus, Dec 21, 2020.

[36] Lyu, Wei, and Wehby, George L. Community use of face masks and covid-19: Evidence from a natural experiment of state mandates in the us. *Health Affairs 39*, 8 (June 16, 2020), 1419–1425.

[37] Massachusetts Department of Public Health. Mask up ma!, 11/02/ 2020.

[38] MIT Election Data and, Science Lab. U.s. president 1976–2020, 2017.

[39] New York Times. See which states and cities have told residents to stay at home, Mar 23, 2020.

[40] Office of Governor Charlie Baker. Covid-19 order 55, Nov 2, 2020.

[41] Office of Governor Charlie Baker. Covid-19 order no.31, May 1, 2020.

[42] Pearl, Judea. Causal inference in statistics: An overview. *Statistics surveys 3* (2009), 96–146.

[43] Peeples, Lynne. Face masks: what the data say. *Nature 586*, 7828 (-10-06 2020), 186–189.

[44] Petersen, Maya L., Porter, Kristin E., Gruber, Susan, Wang, Yue, and van der Laan, Mark J. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research 21*, 1 (Feb 2012), 31–54.

[45] Petersen, Maya L., and van der Laan, Mark J. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass.) 25*, 3 (-05 2014), 418–426.

[46] Robins, James. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling 7*, 9 (1986), 1393–1512.

[47] Stevens, Harry. The post's virus simulator: Why outbreaks like coronavirus spread exponentially, and how to 'flatten the curve.', Mar 18, 2020.

[48] The COVID Tracking Project, and The Atlantic Monthly Group. The COVID Tracking Project.

[49] United States Census Bureau. Acs highlights population housing basics - state centroids.

[50] United States Census Bureau. Average household size and population density - state.

[51] United States Census Bureau. Population and poverty - states.

[52] van der Laan, Mark J, Polley, Eric C., and Hubbard, Alan E. Super learner. *Statistical Applications in Genetics and Molecular Biology 6*, 1 (Sep 16, 2007), 1309.

[53] van der Laan, Mark J, and Rose, Sherri. *Targeted Learning*. Springer New York, New York, NY, 2011.

[54] Vardavas, Raffaele, Strong, Aaron, Bouey, Jennifer, Welburn, Jonathan William, de Lima, Pedro Nascimento, Baker, Lawrence, Zhu, Keren, Priest, Michelle, Hu, Lynn, and Ringel, Jeanne S. The health and economic impacts of nonpharmaceutical interventions to address covid-19: A decision support tool for state and local policymakers.

[55] Wong, Brittany. Why east asians were wearing masks long before covid-19, -09-18T23:47:11Z 2020.