



## RESEARCH ARTICLE

# Evaluating effectiveness of public health intervention strategies for mitigating COVID-19 pandemic

Shanghong Xie<sup>1,2</sup>  | Wenbo Wang<sup>3</sup> | Qinxia Wang<sup>2</sup> | Yuanjia Wang<sup>2</sup>  | Donglin Zeng<sup>3</sup>

<sup>1</sup>School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

<sup>2</sup>Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York, USA

<sup>3</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

## Correspondence

Shanghong Xie, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China.

Email: [xiesh@swufe.edu.cn](mailto:xiesh@swufe.edu.cn)

Yuanjia Wang, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA.

Email: [yw2016@cumc.columbia.edu](mailto:yw2016@cumc.columbia.edu)

Donglin Zeng, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

Email: [dzeng@email.unc.edu](mailto:dzeng@email.unc.edu)

## Funding information

Foundation for the National Institutes of Health, Grant/Award Numbers: GM124104, MH117458, NS073671

Coronavirus disease 2019 (COVID-19) pandemic is an unprecedented global public health challenge. In the United States (US), state governments have implemented various non-pharmaceutical interventions (NPIs), such as physical distance closure (lockdown), stay-at-home order, mandatory facial mask in public in response to the rapid spread of COVID-19. To evaluate the effectiveness of these NPIs, we propose a nested case-control design with propensity score weighting under the quasi-experiment framework to estimate the average intervention effect on disease transmission across states. We further develop a method to test for factors that moderate intervention effect to assist precision public health intervention. Our method takes account of the underlying dynamics of disease transmission and balance state-level pre-intervention characteristics. We prove that our estimator provides causal intervention effect under assumptions. We apply this method to analyze US COVID-19 incidence cases to estimate the effects of six interventions. We show that lockdown has the largest effect on reducing transmission and reopening bars significantly increase transmission. States with a higher percentage of non-White population are at greater risk of increased  $R_t$  associated with reopening bars.

## KEYWORDS

COVID-19, difference-in-difference, heterogeneity of treatment effect, infectious disease modeling, non-pharmaceutical interventions, quasi-experiments

## 1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19) pandemic is an unprecedented global health crisis that has brought tremendous challenges to humanity. Countries around the world have introduced mitigation measures and non-pharmaceutical interventions (NPIs) to respond to the crisis before vaccines are widely available. Within the United States (US), there is tremendous heterogeneity in terms of when mitigation strategies were implemented and lifted across states and a varying-degree of combinations of containment, social distancing, and lockdown (ie, physical distance closures including closure of schools and businesses). Decisions for implementing these strategies partially rely on essential

statistics and epidemiological models that characterize the course of COVID-19 outbreak. However, despite numerous disease forecast models proposed in literature,<sup>1</sup> there is a lack of methods to evaluate intervention effects that are robust and generalizable to accurately account for the heterogeneity between populations. There is no study on precision NPIs that tailor interventions to states according to states' characteristics (eg, demographics, percent of high risk populations susceptible to COVID-19 infection) rather than a national strategy that assigns the same NPIs to all states. In addition, states showed heterogeneous effects after implementing NPIs. Some states controlled spread after implementing NPIs while others did not. It is desirable to explore drivers of this heterogeneity. Thus, it is imminent to study average treatment effect and heterogeneity of treatment effect (HTE) to inform health policy on COVID-19 responses.

One essential component of evaluating an NPI is to identify a proper outcome measure. During COVID-19 pandemic, daily cases and deaths are reported in each state in the US. However, it is well known that there are a large number of pre-symptomatic cases accounting for about 40% of transmissions CDC;<sup>2</sup> and there has been a shortage of accurate polymerase chain reaction (PCR) tests especially during the early phase of the pandemic. In addition to lagged reports, the observed cases do not fully reflect how the epidemic evolves in real time, so simply using reported cases or deaths as outcomes may lead to suboptimal decisions. In contrast, mechanistic-based epidemiological models are more likely to capture the true underlying dynamics of COVID-19 epidemic and provide the time-varying effective reproduction number ( $R_t$ ) as an outcome measure. In particular, our earlier work<sup>3</sup> proposed to combine nonparametric statistical curve fitting with infectious disease epidemiological models of the transmission dynamics. This model accounts for pre-symptomatic transmission of COVID-19 and provides estimates of infection rates and reproduction numbers. These quantities, when modeled as time-varying, can effectively capture the underlying dynamics that govern the disease transmission, leading to better prediction performance, and thus are the appropriate measures to be targeted by an intervention. For example, a reproduction number below one indicates that the disease epidemic is shrinking and under control. We use time-varying reproduction number, denoted by  $R_t$  (see Equation (2) in Section 2.1), as the outcome measure of the intervention effect.

To estimate intervention effects on COVID-19 (eg, change of reproduction number before and after an intervention), we consider methods that use natural experiment designs to allow drawing causal inference under assumptions. Since different states implemented interventions at different time points, the effects of mitigation strategies can be treated as quasi-experiments where subjects receive distinct interventions before or after the initiation of the intervention. The longitudinal pre-post intervention designs including regression-discontinuity design<sup>4</sup> and difference in difference (DID) regressions are frequently used in practice to analyze quasi-experiments data.<sup>5,6</sup> Regression-discontinuity design defines a cutoff point to determine which intervention is assigned and estimates intervention effects by comparing observations with values just above and below the cutoff point. DID estimates the intervention effect by examining the interaction term between time and intervention group (ie, treated or untreated group) in a regression model. It allows valid inference assuming that outcome trends are parallel in treated and untreated group and local randomization holds (ie, whether a subject falls immediately before or after the initiation date of an intervention may be considered random, and thus the "intervention assignment" may be considered to be random). When the first assumption does not hold, synthetic control<sup>7</sup> is proposed to weight observations so that pre-intervention average effects are similar between groups.

Several recent works have investigated the intervention effects of COVID-19 mitigation strategies. Process-based infectious disease models are used to simulate counterfactual outcomes under different manipulations of model parameters and assumptions on the intervention effects, such as assuming a 75% reduction in outside household contacts after implementing social distancing of the entire population, and a 50% increase in household contact rates among student families after the closure of schools and universities.<sup>8,9</sup> These models may be useful to simulate disease outcomes under hypothetical scenarios of interventions, but do not estimate intervention effects based on observed data. Auger et al<sup>10</sup> and Rader et al<sup>11</sup> evaluated the associations between the interventions and outcomes (ie, cases, deaths, and  $R_t$ ) by regression models. Davies et al<sup>12</sup> and Flaxman et al<sup>13</sup> assessed the intervention effects by modeling the basic reproduction number  $R_0$  or  $R_t$  as intervention dependent. These approaches included state-level characteristics as covariates in the model, but did not investigate the causal effects. Cho<sup>14</sup> considered synthetic control and DID approach by fitting linear regression with reported cases and deaths as outcomes, but did not take account of the dynamic feature of the disease transmission (eg,  $R_t$ ).

In this article, we propose a novel method to assess the effect of NPIs using estimated  $R_t$  obtained from the reported daily cases from each state in US. Compared to existing literature, our work has several new aspects as follows. First, since

COVID-19 outbreak may occur at different times in each state, calendar time may not be a good measure of the stage of epidemic. To create a meaningful time horizon that reflects each state's epidemic course when comparing intervention effects, we align states by transforming calendar time to time since the first reported case. Second, we use a nested case-control design (eg, treating the implementation of an intervention as an event)<sup>15</sup> and propensity score weighting to estimate intervention effect. Specifically, for each state that has implemented an intervention at a given time point, we define a set of control states as those which have not yet implemented the intervention. Therefore, a state that implements a policy at a later time can serve as control for other states that have acted earlier. This design would allow observations from different time periods in the same state to serve in both treated and untreated groups, so that the longitudinal data from 50 states can be efficiently used. Third, to balance treated and untreated groups, we construct propensity scores using pre-intervention characteristics including state-level social demographic variables (eg, social vulnerability index [SVI], state's average age and race distribution) as well as time-varying characteristics of the epidemic (eg, pre-intervention case rate, hospitalization,  $R_t$ ). We prove that our estimator yields the causal effect of an intervention under assumptions (eg, consistency and no-unobserved confounder). Lastly, we further estimate heterogeneity of treatment effect (HTE) using estimating equations that include important hypothesized moderators such as age, race, and level of poverty. The developed method is applied to analyze US COVID-19 data to estimate the effects of six NPIs. We show that the lockdown during spring of 2020 had the largest effect on reducing  $R_t$  and reopening bars led to significant increase of disease transmission.

## 2 | METHOD FOR EVALUATING INTERVENTION STRATEGIES

### 2.1 | Outcome measure for estimating NPI effects

To estimate the time-varying infection rate or reproduction number as an outcome for assessing NPIs, we adopt a previously developed method, survival-convolution model,<sup>3</sup> over days since the reported first case. This model is inspired by the epidemiological susceptible-exposed-infective-recovered (SEIR) model, but has fewer assumptions and model parameters, and demonstrate adequate prediction performance among an ensemble of models in the CDC forecast task (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>).

To be specific, let  $N_i(t)$  be the number of individuals in the  $i$ th state who are newly infected by COVID-19 at time  $t$ . We assume that the virus transmits from one individual to another at the same rate at a given time  $t$  to investigate the population-level disease transmission. In this population, the duration of an individual remaining infectious in the epidemic is from a homogeneous distribution at any calendar time  $t$  (in days). Let  $S(m)$  be the proportion of persons remaining infectious after  $m$  days of being infected. We assume that individuals will not infect others once they are out of transmission chain due to any possible reasons (eg, prior infection, testing positive and quarantine, or out of infectious period). Since the total number of individuals who are newly infected at time  $(t - m)$  is  $N_i(t - m)$  and assuming the infection rate at  $t$  to be  $a_i(t)$ , then

$$N_i(t + 1) = a_i(t) \sum_{m=0}^{\infty} N_i(t - m)S(m + 1). \quad (1)$$

The details of derivations are provided in the Supporting Information Web Appendix A.

Equation (1) gives a convolution update for the new daily cases using the past days' number of cases. This equation considers three important quantities to characterize COVID-19 transmission: the initial date,  $t_0$ , of the first (likely undetected) case in the epidemic, the survival function of time to out of transmission chain (ie, not infect others),  $S(m)$ , and the infection rate over calendar time,  $a_i(t)$ . Wang et al<sup>3</sup> estimated  $a_i(t)$  as a piece-wise linear function with knots placed at intervention dates and every 2 to 3 weeks, and approximated the survival function  $S(m)$  based on previous literature.<sup>16</sup> Similarly, we computed  $a_i(t)$  as piece-wise linear function, placing knots at the state-specific intervention dates and every 2 weeks between interventions and modeled  $S(m)$  as an exponential distribution. To estimate both  $t_0$  and  $a_i(t)$ , Wang et al<sup>3</sup> proposed to minimize a squared loss between the square-root transformed reported daily new cases and predicted new cases from model (1).

Note that  $a_i(t)$  is time-varying because the infection rate depends on how many close contacts one infected individual may have at day  $t$ , which is affected by NPIs (eg, stay-at-home order, lockdown) and saturation level of the infection in

the whole population. With the number of new infections  $N_i(t)$  estimated from survival-convolution model in (1), the effective reproduction number<sup>17</sup> is defined as

$$R_{it} = \frac{N_i(t)}{\sum_{k=0}^{\infty} w_k N_i(t-k)}, \quad (2)$$

which is the number of secondary infections caused by a primary infected individual in a population at time  $t$  while accounting for the entire incubation period of the primary case. Thus,  $R_{it}$  measures temporal changes of the disease transmission. Here,  $w_k$  is the probability mass function of the distribution of serial intervals for SARS-CoV-2 (a Gamma distribution), which is obtained from Nishiura et al<sup>18</sup> and Scire et al.<sup>19</sup>

## 2.2 | Average intervention effect and assumptions

For the ease of presentation, we focus on a particular intervention (lockdown, for instance) in this section. Our goal is to estimate the overall average effect of the intervention across states. To define the causal estimand, we introduce the following notations to define a time-specific intervention effect. For any time period  $\Delta > 0$  such that the probability of two states implementing the intervention within  $\Delta$  days is zero, we let  $Y_i^{(1)}(t + \Delta; t)$  denote the potential change of the reproduction number between  $t$  and  $(t + \Delta)$ , had the intervention been applied at time  $t$  and had there been no other interventions between time  $t$  and  $t + \Delta$ . Let  $Y_i^{(0)}(t + \Delta; t)$  be the same potential outcome when there was no intervention at time  $t$ . Correspondingly, the time-specific intervention effect is defined as

$$\gamma(\Delta, t) = E[Y_i^{(1)}(t + \Delta; t) - Y_i^{(0)}(t + \Delta; t)].$$

In other words, we consider a hypothetical scenario where at time  $t$ , each state imposes the intervention and the other scenario where there is no such intervention at  $t$  and before. Then  $\gamma(\Delta, t)$  is the expected difference between the change of the reproduction number  $\Delta$  days after time  $t$ . A negative value of  $\gamma(\Delta, t)$  implies that the intervention at time  $t$  can slow down the spread of the virus. However, since very few states impose the intervention on the same day since disease outbreak, estimating  $\gamma(\Delta, t)$  for each  $t$  is not feasible. Instead, we define an average intervention effect (ATE) as the average of  $\gamma(\Delta, t)$  over all possible intervention times, that is,

$$\gamma(\Delta) \equiv \int \gamma(\Delta, t) dF_T(t),$$

where  $F_T(\cdot)$  is the distribution of the intervention time  $T$ . Hence,  $\gamma(\Delta)$  can be viewed as an overall evaluation of the intervention effect over all states. We are interested to estimate  $\gamma(\Delta)$  using empirical data.

For each state  $i$ , we set time zero to be its first reported case and let  $Y_i(t + \Delta; t)$  be the change of reproduction number between  $(t + \Delta)$  and  $t$  (ie,  $R_{it+\Delta} - R_{it}$ ), where the reproduction numbers are estimated as Section 2.1. Let  $X_i$  be the state-specific characteristics including a constant of one. Let  $T_i$  denote the intervention time and let  $T_i = \infty$  if the state has never implemented this intervention. Let  $F_T(t)$  denote the distribution of  $T_i$ , assumed to have a support on  $\mathcal{T}$ . To estimate  $\gamma(\Delta)$  from observed data, we require the following assumptions:

- (a) Suppose no other intervention occurs between  $t$  and  $t + \Delta$ . We assume when  $T_i = t$  (ie, there is an intervention at  $t$ ),  $Y_i^{(1)}(t + \Delta; t) = Y_i(t + \Delta; t)$ .
- (b) Suppose no other intervention occurs between  $t$  and  $t + \Delta$  and the intervention of interest has not been imposed before  $t$ , we assume  $Y_i^{(0)}(t + \Delta; t) = Y_i(t + \Delta; t)$ .
- (c) Assume no unobserved confounders: conditional on  $T_i \geq t$ ,  $T_i = t$  is independent of  $Y_i^{(a)}(t + \Delta; t)$ ,  $a = 0, 1$  given  $X_i$  and  $H_i(t)$ , where  $H_i(t)$  denotes the observed epidemic history by time  $t$ .

Assumptions (a) and (b) are equivalent to the consistency assumption in causal inference. Both (a) and (b) also imply that there are no delayed effects from any other previous interventions prior to time  $t$ . This is plausible since the interventions do not occur frequently and the effects can decline rapidly, as seen by multiple resurges in this pandemic. Furthermore, even though the previous intervention may affect the infection rate at time  $t$ , since the potential outcome

of interest is the change of the infection rate or reproduction number since time  $t$ , the effect on this change can be much smaller. Assumption (c) is the no-unobserved confounder assumption in causal inference literature. If all relevant epidemic history and other information associated with implementing an intervention at time  $t$  are collected as  $H_i(t)$  and  $X_i$ , this assumption holds. In our application, we will explore a list of candidate variables as  $(X_i, H_i(t))$  and identify a subset data-adaptively.

Next, we justify why the assumptions enable us to estimate  $\gamma(\Delta)$ . First, under assumption (c), we have

$$\begin{aligned} \gamma(\Delta, t) = & E \left[ \frac{I(T_i = t)}{P(T_i = t | T_i \geq t, H_i(t), X_i)} \left\{ Y_i^{(1)}(t + \Delta; t) \right\} \right] \\ & - E \left[ \frac{I(T_i > t + \Delta)}{P(T_i > t + \Delta | T_i \geq t, H_i(t), X_i)} \left\{ Y_i^{(0)}(t + \Delta; t) \right\} \right]. \end{aligned} \quad (3)$$

Second, since  $P(T_i > t | T_i \geq t, H_i(t), X_i) = P(T_i > t + \Delta | T_i \geq t, H_i(t), X_i)$  for any  $t$  in the support of  $F_T(t)$ , according to assumptions (a) and (b), the right-hand side is further equal to

$$\begin{aligned} \gamma(\Delta, t) = & E \left[ \frac{I(T_i = t)}{P(T_i = t | T_i \geq t, H_i(t), X_i)} \{ Y_i(t + \Delta; t) \} \right] \\ & - E \left[ \frac{I(T_i > t + \Delta)}{P(T_i > t + \Delta | T_i \geq t, H_i(t), X_i)} \{ Y_i(t + \Delta; t) \} \right]. \end{aligned} \quad (4)$$

Therefore, if we posit a model for the intervention time  $T_i$  given  $H_i(t)$  and  $X_i$ , an inverse probability weighted estimator based on (4) can be used to estimate  $\gamma(\Delta, t)$ . Equation (4) further provides a way to consistently estimate  $\gamma(\Delta)$  by simply averaging the estimated  $\gamma(\Delta, t)$  over all empirical intervention times from all states.

### 2.3 | Inference procedure for the ATE

The main idea for estimation is to create a separate set of control states for “case states” that implemented an intervention at a given time point and then aggregate over case states. To balance pre-intervention differences between states, we will construct propensity scores for case states that intervened at different time points, since eligible control states may differ. Specifically, in the first step, we estimate the propensity scores,  $P(T_i = t | H_i(t), X_i)$  in (4), by fitting a logistic regression model,

$$\text{logit} \{ P(T_i = t | T_i \geq t, H_i(t), X_i) \} = (H_i(t), X_i)^T \beta,$$

where  $X_i$  contains all prognostic variables for the intervention at the baseline such as demographic distributions and SVI index, and  $H_i(t)$  can be the average cases and deaths in the past week(s) before time  $t$ . To estimate  $\beta$ , we solve the following estimating equation

$$\sum_{i=1}^n \int (H_i(t), X_i)^T I(T_i \geq t) \left[ I(T_i = t) - \frac{\exp\{(H_i(t), X_i)^T \beta\}}{1 + \exp\{(H_i(t), X_i)^T \beta\}} \right] d\hat{F}_T(t) = 0,$$

where  $\hat{F}_T(t)$  denotes the empirical distribution of the intervention times. In detail, if we use  $X_{ij}$  to denote  $(H_i(T_j), X_i)$  and  $\delta_{ij} = I(T_i = T_j)$ , we can estimate  $\beta$  by solving

$$\sum_{i=1}^n \sum_{j \in S(i)} X_{ij} \left\{ \delta_{ij} - \frac{\exp\{X_{ij}^T \beta\}}{1 + \exp\{X_{ij}^T \beta\}} \right\} = 0,$$

where  $S(i)$  is a set of state  $i$  and all other eligible control states (eg, states that have not implemented an intervention by  $T_i$ ; similar to a nested case-control design when treating implementation of an intervention as the event). Once we obtain the estimate for  $\beta$ , denoted by  $\hat{\beta}$ , the propensity score for state  $i$  at its intervention time  $t$  is given by

$$\hat{p}_i(t) = \frac{\exp\{(H_i(t), X_i)^T \hat{\beta}\}}{1 + \exp\{(H_i(t), X_i)^T \hat{\beta}\}}.$$

In the second step, using the estimated propensity scores, according to (4) for  $t \in \mathcal{T}$  and by the definition of the ATE  $\gamma(\Delta)$ , we estimate  $\gamma(\Delta)$  explicitly as

$$\hat{\gamma}(\Delta) = \frac{\sum_{i=1}^n \int I(T_i = t) / \hat{p}_i(t) Y_i(t + \Delta) d\hat{F}_T(t)}{\sum_{i=1}^n \int I(T_i = t) / \hat{p}_i(t) d\hat{F}_T(t)} - \frac{\sum_{i=1}^n \int I(T_i > t + \Delta) / (1 - \hat{p}_i(t)) Y_i(t + \Delta) d\hat{F}_T(t)}{\sum_{i=1}^n \int I(T_i > t + \Delta) / (1 - \hat{p}_i(t)) d\hat{F}_T(t)},$$

where for the convenience of notation, we use  $Y(t + \Delta)$  to denote  $Y(t + \Delta; t)$  in subsequent exposition. Removing the denominators in the above expression does not necessarily invalidate the consistency of the estimator, but can lead to substantial efficiency gain as shown in survey sampling literature (eg, using standardized weights may improve efficiency). Specifically, to calculate covariates for state  $i$  at state  $j$ 's intervention day, let  $t_{j(i)}$  be the intervention day for state  $i$ , let  $X_{i,j(i)} = (H_i(t_{j(i)}), X_i)^T$ , and define  $\hat{p}_i = \frac{\exp\{X_{i,j(i)}^T \hat{\beta}\}}{1 + \exp\{X_{i,j(i)}^T \hat{\beta}\}}$ . Then in the second step, we estimate  $\gamma(\Delta)$  by

$$\hat{\gamma}(\Delta) = \frac{\sum_{i=1}^n \sum_{j \in S(i)} d_{ij} \delta_{ij} / \hat{q}_{ij}}{\sum_{i=1}^n \sum_{j \in S(i)} \delta_{ij} / \hat{q}_{ij}} - \frac{\sum_{i=1}^n \sum_{j \in S(i)} d_{ij} (1 - \delta_{ij}) / (1 - \hat{q}_{ij})}{\sum_{i=1}^n \sum_{j \in S(i)} (1 - \delta_{ij}) / (1 - \hat{q}_{ij})},$$

where  $d_{ij}$  is the change in reproduction number (ie,  $Y_i(t_{j(i)} + \Delta)$ , or  $R_{i,t_{j(i)} + \Delta} - R_{i,t_{j(i)}}$ ),  $\delta_{ij}$  is the change in intervention status at time  $j$  for state  $i$ , and

$$\hat{q}_{ij} = \frac{\exp\{X_{ij}^T \hat{\beta}\}}{1 + \exp\{X_{ij}^T \hat{\beta}\}}, \quad i = 1, \dots, n, j \in S(i).$$

Note  $\hat{p}_i = \hat{q}_{i,j(i)}$ .

The following theorem gives the asymptotic distribution for  $\hat{\gamma}(\Delta)$ .

**Theorem 1.** Under assumptions (a) to (c) and assuming that  $(H_i(t), X)$  is linearly independent with positive probability for some  $t$  in  $\mathcal{T}$  and that  $H(t)$  has a bounded total variation in  $\mathcal{T}$ ,  $\sqrt{n}(\hat{\gamma}(\Delta) - \gamma(\Delta))$  converges to a mean-zero normal distribution.

The asymptotic variance in Theorem 1 is given in the proof in the Appendix. A consistent estimator for the variance can be given by a plug-in estimator. Specifically, the proof of Theorem 1 implies that  $\sqrt{n}(\hat{\beta} - \beta)$  is asymptotically normal, where  $\beta$  is the true parameter value in the propensity score model, and the asymptotic variance can be consistently estimated by  $(\sum_{i=1}^n V_i V_i^T) / n$ , where

$$V_i = \left[ n^{-1} \sum_{i=1}^n \sum_{j \in S(i)} X_{ij} X_{ij}^T \hat{q}_{ij} (1 - \hat{q}_{ij}) \right]^{-1} \left\{ \sum_{j \in S(i)} X_{ij} (\delta_{ij} - \hat{q}_{ij}) \right\}.$$

Finally, through the linear expansion given in the proof of Theorem 1, If we let

$$A_i = \sum_{j \in S(i)} d_{ij} \delta_{ij} / \hat{q}_{ij}, \quad B_i = \sum_{j \in S(i)} \delta_{ij} / \hat{q}_{ij}$$

and

$$C_i = \sum_{j \in S(i)} d_{ij} (1 - \delta_{ij}) / (1 - \hat{q}_{ij}), \quad D_i = \sum_{j \in S(i)} (1 - \delta_{ij}) / (1 - \hat{q}_{ij})$$

and  $\bar{A}, \bar{B}, \bar{C}$ , and  $\bar{D}$  be their respective average values, then the asymptotic variance for  $\hat{\gamma}(\Delta)$  can be estimated as  $\hat{\sigma}^2 = n^{-2} \sum_{i=1}^n (U_i - \bar{U})^2$ , where

$$\begin{aligned}
U_i = & \frac{A_i}{n^{-1} \sum_{k=1}^n B_k} - \frac{C_i}{n^{-1} \sum_{k=1}^n D_k} - \frac{n^{-1} \sum_{k=1}^n A_k}{(n^{-1} \sum_{k=1}^n B_k)^2} B_i + \frac{n^{-1} \sum_{k=1}^n C_k}{(n^{-1} \sum_{k=1}^n D_k)^2} D_i \\
& - \left[ \frac{\sum_{k=1}^n \sum_{j \in S(k)} d_{kj} \delta_{kj} (1 - \hat{q}_{kj}) X_{kj}^T / \hat{q}_{kj}}{\sum_{k=1}^n B_k} \right. \\
& \left. - \frac{\sum_{k=1}^n A_k}{(\sum_{k=1}^n B_k)^2} \left( \sum_{k=1}^n \sum_{j \in S(k)} \delta_{kj} (1 - \hat{q}_{kj}) X_{kj}^T / \hat{q}_{kj} \right) \right] V_i \\
& - \left[ \frac{\sum_{k=1}^n \sum_{j \in S(k)} d_{kj} (1 - \delta_{kj}) \hat{q}_{kj} X_{kj}^T / (1 - \hat{q}_{kj})}{\sum_{k=1}^n D_k} \right. \\
& \left. - \frac{\sum_{k=1}^n C_k}{(\sum_{k=1}^n D_k)^2} \left( \sum_{k=1}^n \sum_{j \in S(k)} (1 - \delta_{kj}) \hat{q}_{kj} X_{kj}^T / (1 - \hat{q}_{kj}) \right) \right] V_i \\
& + \frac{1}{\sum_{k=1}^n B_k} \frac{d_{ii}}{q_{ii}} - \frac{\sum_{k=1}^n A_k}{(\sum_{k=1}^n B_k)^2} \frac{1}{q_{ii}} \\
& - \frac{\sum_{k=1}^n (1 - \delta_{ki}) / (1 - q_{ki}) d_{ki}}{\sum_{k=1}^n D_k} + \frac{\sum_{k=1}^n C_k}{(\sum_{k=1}^n D_k)^2} \left( \sum_{k=1}^n (1 - \delta_{ki}) / (1 - q_{ki}) \right).
\end{aligned}$$

Therefore, the 95%-confidence interval for the ATE is  $[\hat{\gamma}(\Delta) - 1.96\sqrt{\hat{\sigma}^2}, \hat{\gamma}(\Delta) + 1.96\sqrt{\hat{\sigma}^2}]$ .

*Remark 1.* Since we may have a small number of states with an NPI when fitting the propensity score, the model can be either saturated or overfitted when the dimension of  $X_i$  and  $H_i(t)$  increases. We perform a screening step to obtain a parsimonious model for estimating the propensity scores.

*Remark 2.* The estimand  $\gamma(\Delta)$  depends on the window size,  $\Delta$ , between the intervention time  $t$  and effect time  $(t + \Delta)$ . We can vary different window sizes so as to obtain the estimated intervention effects over days since the intervention. This can be useful to study how long it might take for an intervention to become effective.

## 2.4 | Estimation of HTE by regression

A similar procedure can be applied to study the effect in a subgroup of states which share similar characteristics of  $Z_i$  and moderation effects of  $Z_i$  (here  $Z_i$  is a subset of  $X_i$ ). To estimate which factors in  $Z_i$  may moderate the intervention effect, we use a regression model by assuming

$$E[Y_i^{(1)}(t + \Delta; t) - Y_i^{(0)}(t + \Delta; t) | Z_i] = \theta^T Z_i.$$

Thus, testing the significance of  $\theta$  identifies significant factors that moderate intervention effect, a.k.a, HTE, which may lead to precision public health policy that targets states with certain characteristics.

Specifically, the estimator for  $\theta$  can be obtained by solving

$$\sum_{i=1}^n Z_i \left[ \int \left\{ Y_i(t + \Delta) \left( \frac{I(T_i = t)}{\hat{p}_i(t)} - \frac{I(T_i > t + \Delta)}{1 - \hat{p}_i(t)} \right) - \theta^T Z_i \right\} I(T_i \geq t) d\hat{F}_T(t) \right] = 0,$$

or equivalently,

$$\sum_{i=1}^n Z_i \left[ \sum_{j \in S(i)} \left\{ d_{ij} \left( \frac{\delta_{ij}}{\hat{q}_{ij}} - \frac{1 - \delta_{ij}}{1 - \hat{q}_{ij}} \right) - \theta^T Z_i \right\} \right] = 0.$$

When  $Z_i = 1$ , the derived estimator is asymptotically equivalent to  $\hat{\gamma}(\Delta)$  studied before. Let  $\hat{\theta}$  denote the estimator. Our next theorem states the asymptotic covariance of  $\hat{\theta}$ .

**Theorem 2.** Under the assumptions in Theorem 1, if we further assume  $E[Z_i Z_i^T]$  is nonsingular, it holds

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &= E \left[ Z Z^T \int I(T \geq t) dF_T(t) \right]^{-1} \\ &\times \left[ Z \left( \int \left\{ Y(t + \Delta) \left( \frac{I(T = t)}{p(t)} - \frac{I(T > t + \Delta)}{1 - p(t)} \right) - \theta^T Z \right\} I(T \geq t) dF_T(t) \right) \right. \\ &\quad \left. - \tilde{E}[\tilde{Z} \int \tilde{Y}(t + \Delta) \left( I(\tilde{T} = t) \frac{1 - p(t)}{p(t)} + I(\tilde{T} > t + \Delta) \frac{p(t)}{1 - p(t)} \right) \right. \\ &\quad \left. \times (\tilde{H}(t), \tilde{X})^T I(T \geq t) dF_T(t)] S_\beta \right] + o_p(1), \end{aligned}$$

where  $\tilde{E}[\cdot]$  denotes the expectation with respect to  $\tilde{Y}$  and  $\tilde{T}$ , and  $S_\beta$  is the influence function for  $\hat{\beta}$  given in the proof of Theorem 1. Consequently,  $\sqrt{n}(\hat{\theta} - \theta)$  converges weakly to a mean-zero normal distribution.

The proof for Theorem 2 uses the same linear expansion argument as in the proof for Theorem 1 so is omitted. As a result of Theorem 2, the variance for  $\hat{\theta}$  can be consistently estimated by the following sandwich estimator,  $\hat{\Psi} = \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}$ , where

$$\Sigma_1 = \sum_{i=1}^n \sum_{j \in S(i)} Z_i Z_i^T$$

and  $\Sigma_2 = \sum_{i=1}^n W_i W_i^T$  with

$$\begin{aligned} W_i &= Z_i \left[ \sum_{j \in S(i)} \left\{ d_{ij} \left( \frac{\delta_{ij}}{\hat{q}_{ij}} - \frac{1 - \delta_{ij}}{1 - \hat{q}_{ij}} \right) - \hat{\theta}^T Z_i \right\} \right] \\ &\quad - \left[ n^{-1} \sum_{k=1}^n Z_k \sum_{j \in S(k)} d_{kj} X_{kj}^T \left( \delta_{kj} \frac{1 - \hat{q}_{kj}}{\hat{q}_{kj}} + (1 - \delta_{kj}) \frac{\hat{q}_{kj}}{1 - \hat{q}_{kj}} \right) \right] V_i. \end{aligned}$$

Therefore, to test whether the  $l$ th component of  $\theta$  is zero at a significance level of  $\alpha$ , we reject the null if  $|\hat{\theta}_l| / \sqrt{\hat{\Psi}_{ll}}$  is larger than the  $(1 - \alpha/2)$ -quantile of the standard normal distribution, where  $\hat{\theta}_l$  is the  $l$ th component of  $\hat{\theta}$  and  $\hat{\Psi}_{ll}$  is the  $l$ th diagonal element of  $\hat{\Psi}$ .

### 3 | SIMULATION STUDIES AND ANALYSIS OF US COVID-19 DATA

We evaluated our method in two simulation settings with sample size of 50 (states) and  $R_t$  decreasing 0.15, 0.2 per day after implementing the intervention. For each simulated dataset, we compared the estimated ATE with the approach of using calendar time as the time scale. In all settings, our method had smaller root mean squared errors (RMSEs) than the approach of using calendar time. We present details of the simulation studies in the Supporting Information Web Appendix B.

We applied our method to analyze US COVID-19 data. Since the first reported case in Washington on January 22, 2020, COVID-19 spread rapidly across US, especially in the northeast. During mid-March to early April, states issued lockdown orders (physical distance closures) after the national emergency was declared on March 13, 2020. Large declines in the number of daily new reported cases and deaths were seen in April and May after lockdown orders. However, a second surge of COVID-19 arrived in June after reopening, primarily in the southern and western states. From November 2020 to early 2021, US has experienced a third surge of COVID-19 while the mass vaccination started to take place.

We consider six state-wide NPIs: lockdown (date defined as the first physical distance closure), stay-at-home order, mandatory facial masks, reopening business, reopening restaurants, and reopening bars. In our analysis, 48 states that have implemented an intervention after their first reported case were included. States issued lockdown orders between March 9 and April 3, 2020; 39 states placed stay-at-home order between March 19 and April 7; and 37 states mandated



facial masks in public between April 8 and November 20. Between April 20 and June 8, 49 states issued reopening business order; 46 states issued reopen restaurant order between April 24 and July 3; and 44 states issued reopen bar order between May 1 and July 3. We aligned states by transforming calendar time to time since the first reported case. Figure 1 aligns states in two different ways: aligning by calendar dates (Figure 1A), and aligning by days since the first reported case (Figure 1B). The two alignments differ, for example, many states implemented lockdown on March 16 but they were at different days since their first reported case. The latter alignment provides more variability between states and more meaningful measure as the stage in the pandemic. Figure 1B shows that stay-at-home order followed quickly after lockdown, and intervention times for other NPIs vary considerably across states. The intervention time of lockdown was between (0, 54) days since the first reported case, stay-at-home was between (6, 65) days, and mandatory facial masks was between (34, 263) days. Reopening economy policies had a wider range of times between states. The gap time between implementing two different interventions also vary across states. We leverage these heterogeneity to match a “case state” with “control states” without interventions.

We fitted survival-convolution models for each state, using the daily incidence cases reported at Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE<sup>20</sup>) from the date of the first observed case as early as January 22, 2020 to February 16, 2021. This model successfully captured the epidemic trends of COVID-19 incidence cases in 50 states (Figure 2). The fitted curves captured surges in large states such as New York, California, Florida, Texas, as well as smaller states including Maine, Wyoming, and the Dakotas. From the estimated new infections, we derive  $R_t$  using Equation (3). We show the estimated  $R_t$  over the epidemic course in the Web Appendix Figure S1.

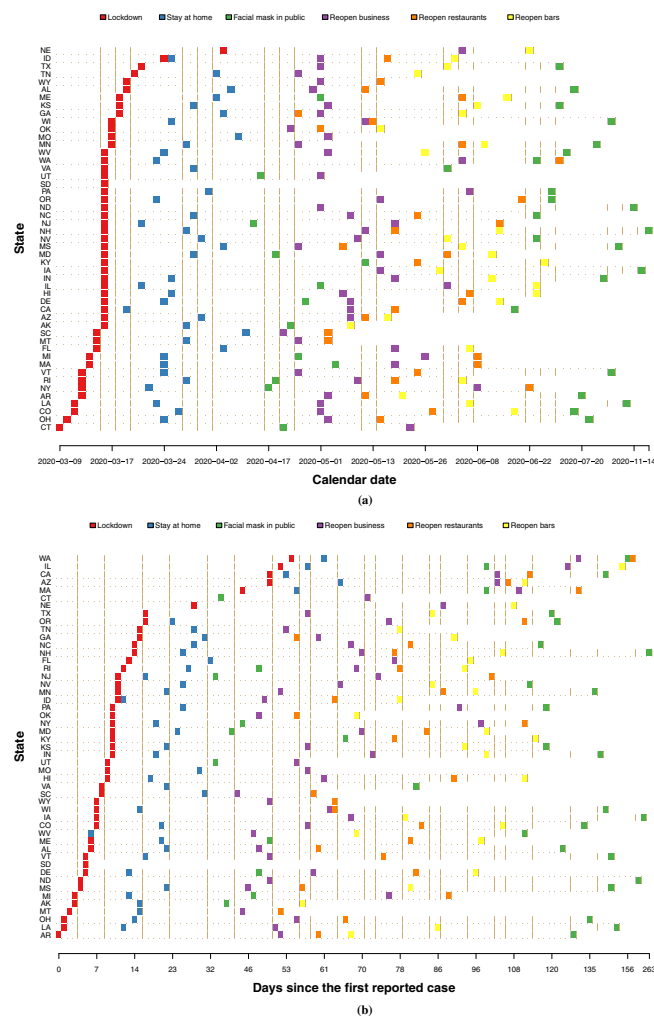
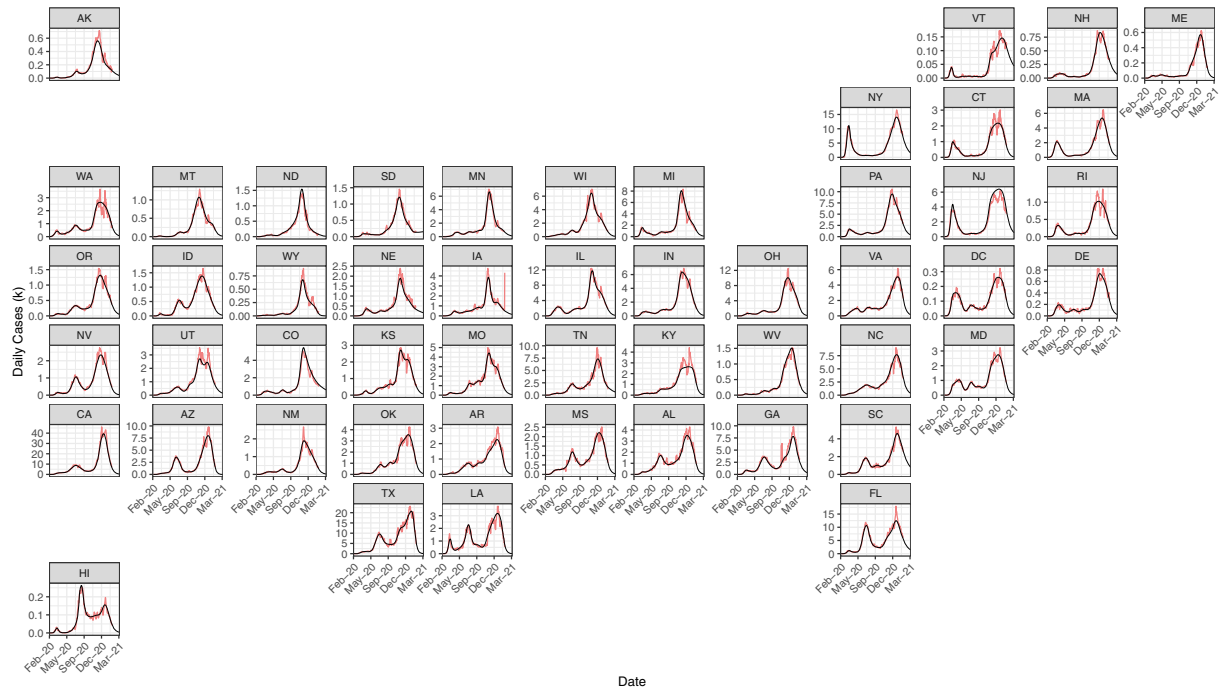


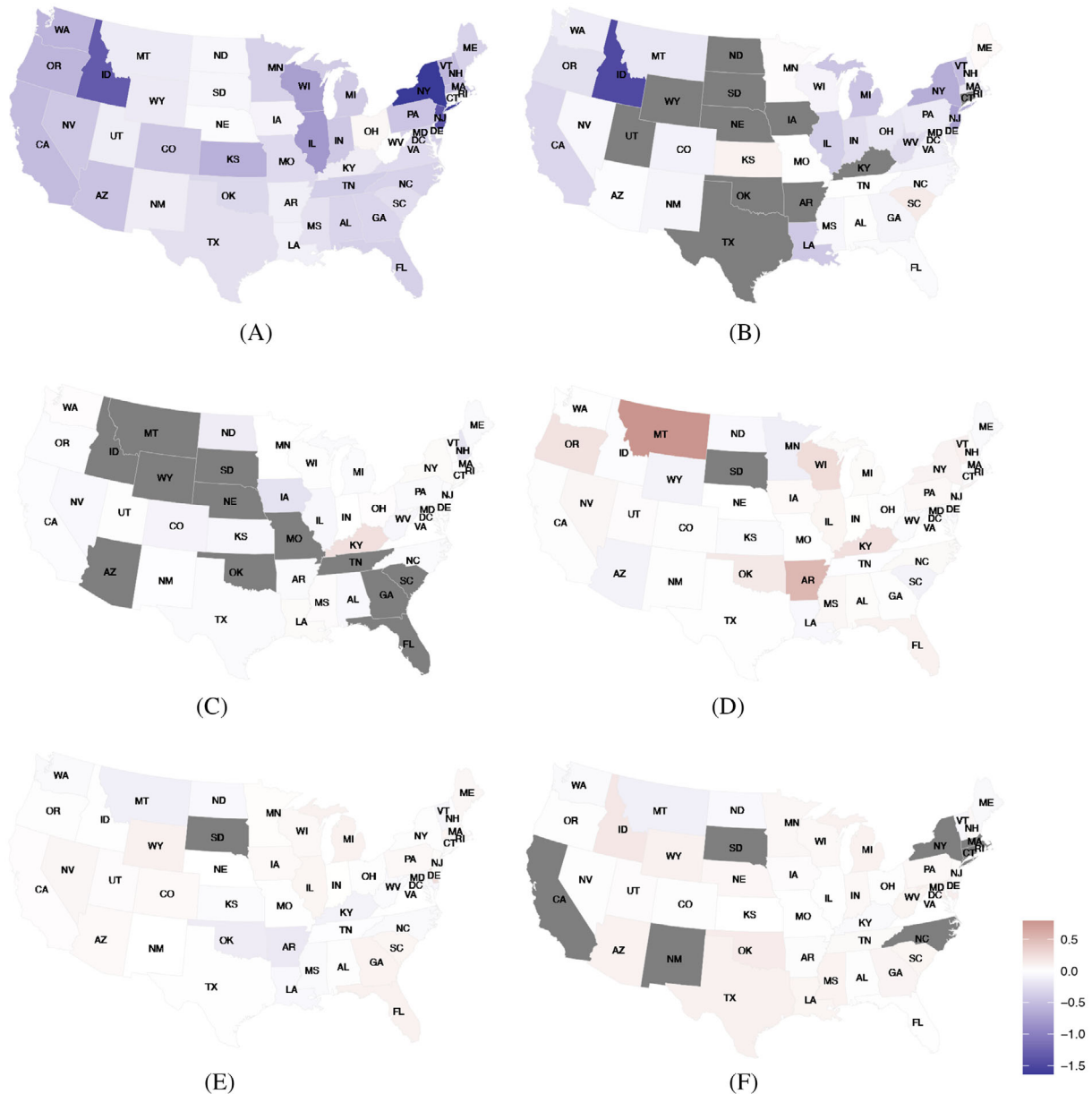
FIGURE 1 Timing of interventions across states: Calendar dates vs days since first reported case. (A) Calendar dates of interventions and (B) intervention time



**FIGURE 2** Observed (7-day moving average; red curve) and fitted (black curve) incidence COVID-19 cases from February 2020 to March 2021 in US states

To visualize observed changes in  $R_t$  after each NPI, we present  $R_t$  differences between 7 days post intervention and 1 day before intervention in Figure 3. A darker cool color indicates a larger decrease in  $R_t$  and a darker warm color indicates a larger increase. The states that did not implement certain NPIs are colored in gray. We see that  $R_t$  in many states in the northeast and west decreased sharply 7 days after lockdown. For most states that had placed stay-at-home orders,  $R_t$  also decreased after the orders. As a comparison, not all states showed a reduction in  $R_t$  after facial mask mandates. Reopening business presents some degree of heterogeneity. Among the three reopening interventions, reopening bars had the largest increase in  $R_t$ . These results show the observed changes in the states that had initiated NPIs, but lacks a control group. We will use the methods developed in Section 2 to formally estimated intervention effects by a DID estimator under the nested case-control design.

Our goal is to formally quantify the impacts of NPIs and separate intervention effect from a natural decrease or increase trend in the absence of intervention using the inversely weighted DID estimator developed in Section 2. We estimated the ATE  $\gamma(\Delta)$ , change in  $R_t$  after  $\Delta$  days of implementing the intervention. In our analysis, we evaluated lockdown's effect up to 6 days, stay-at-home orders up to 11 days, and other interventions up to 14 days. Lockdown and stay-at-home orders had shorter evaluation period because they were enacted at relatively short time interval. A greater  $\Delta$  would not satisfy assumption (b) since other interventions may be introduced during a longer interval. We only regarded the states with intervention time  $T_j > T_i + \Delta$  as eligible "control states" for state  $i$ . State-specific characteristics were included as covariates to construct propensity scores to account for differences between states. Given the associations between state-level characteristics and COVID-19 transmission and NPIs,<sup>10,21,22</sup> the candidate covariates were the demographic characteristics including the percentage of White, the percentage of Latino, the percentage of male, the percentage of age 65 and over, the percentage of male at age 65 and over, CDC SVI variables<sup>23</sup> including the percentage of below poverty, the percentage of unemployed, the percentage of no high school diploma, the percentage of speaking English "less than well," the percentage of housing in structures with 10 or more units, the percentage of mobile homes, the percentage of more people than rooms at household level, the percentage of no vehicle, the percentage of in institutionalized group quarters, the percentage of civilian non-institutionalized population with a disability, the percentage of single parent households with children under 18, and per capita income. The time-varying covariates including average  $R_t$ , average daily new reported cases, average daily new reported deaths, average rate of positive tests, and average percentage of total inpatient beds utilized by patients who have probable or confirmed COVID-19<sup>24</sup> during 1 week prior to the intervention. We standardized the unemployment variable by the state's population of aged 17 to 65, and standardized the other SVI variables except



**FIGURE 3** Difference in  $R_t$  between 7 days post intervention and 1 day before intervention for each NPI in US states. Dark gray color indicates that a state had not implemented an NPI. (A) Lockdown, (B) stay at home, (C) facial mask mandate, (D) reopen business, (E) reopen restaurants, and (F) reopen bars

for per capita income by state's total population. The time-varying covariates were also standardized by state's population and multiplied by 100 000. A different set of propensity scores was constructed for each  $\Delta$  because eligible control states could change. We selected the top 10 covariates based on Spearman rank correlation for each intervention separately and the covariates with a large proportion of missing were excluded.

Web Appendix Tables S3 to S8 show the propensity score estimates of each intervention. The states with higher average pre-intervention  $R_t$ , larger average daily new cases, and larger average daily new deaths, fewer persons who speak English "less than well," higher Latino population, higher institutionalized population, and higher percentage of crowded household were more likely to enact the lockdown order. For stay-at-home order, states with larger average daily new cases and smaller population of no high school diploma were more likely to implement this NPI. The states with larger average daily new cases were more likely to require wearing facial masks, and the states with larger average daily new cases and deaths and fewer mobile homes were less likely to reopen bars.

TABLE 1 Average intervention effects of the six NPIs

Day	Lockdown Estimate (SE)	Stay-at-home Estimate (SE)	Mask mandate Estimate (SE)	Reopen businesses Estimate (SE)	Reopen restaurants Estimate (SE)	Reopen bars Estimate (SE)
$\Delta = 1$	-0.176 (0.022)	-0.006 (0.036)	-0.005 (0.004)	0.022 (0.005)	0.018 (0.004)	0.020 (0.005)
$\Delta = 2$	-0.334 (0.043)	0.027 (0.033)	-0.008 (0.007)	0.033 (0.012)	0.018 (0.007)	0.032 (0.006)
$\Delta = 3$	-0.489 (0.092)	0.027 (0.035)	-0.010 (0.010)	0.036 (0.017)	0.018 (0.012)	0.044 (0.009)
$\Delta = 4$	-0.562 (0.056)	0.010 (0.042)	-0.011 (0.014)	0.041 (0.024)	0.019 (0.018)	0.058 (0.011)
$\Delta = 5$	-0.603 (0.057)	-0.015 (0.048)	-0.014 (0.020)	0.055 (0.027)	0.013 (0.026)	0.071 (0.014)
$\Delta = 6$	-0.759 (0.161)	-0.064 (0.048)	0.001 (0.024)	0.058 (0.036)	0.011 (0.033)	0.082 (0.017)
$\Delta = 7$	-	-0.133 (0.051)	-0.016 (0.030)	0.060 (0.046)	0.006 (0.042)	0.095 (0.020)
$\Delta = 8$	-	-0.113 (0.079)	-0.017 (0.032)	0.035 (0.062)	0.004 (0.049)	0.105 (0.022)
$\Delta = 9$	-	-0.150 (0.080)	0.006 (0.022)	0.023 (0.077)	-0.005 (0.065)	0.120 (0.024)
$\Delta = 10$	-	-0.198 (0.236)	0.009 (0.023)	0.028 (0.084)	-0.027 (0.086)	0.132 (0.026)
$\Delta = 11$	-	-0.233 (0.159)	0.017 (0.026)	0.034 (0.092)	-0.033 (0.096)	0.144 (0.029)
$\Delta = 12$	-	-	0.020 (0.028)	0.049 (0.102)	-0.045 (0.108)	0.154 (0.031)
$\Delta = 13$	-	-	0.022 (0.024)	0.064 (0.110)	-0.047 (0.118)	0.160 (0.032)
$\Delta = 14$	-	-	0.023 (0.026)	0.067 (0.119)	-0.073 (0.140)	0.170 (0.034)

Note: “-” indicates the effect was not applicable at  $\Delta$  day.

The ATEs of the six NPIs are shown in Table 1 and Figure 4. Enacting lockdown significantly decreased  $R_t$  immediately after its implementation, with an average effect of  $-0.759$  (95% CI,  $-1.075$  to  $-0.443$ ) 6 days after. The effect of stay-at-home order reached  $-0.133$  (95% CI,  $-0.233$  to  $-0.033$ ) 7 days post-intervention. Reopening bars significantly increased  $R_t$ . The average effect of reopening bars was an increase of  $0.095$  (95% CI,  $0.056$  to  $0.134$ ) after 7 days and reached  $0.17$  (95% CI,  $0.103$  to  $0.237$ ) after 14 days. The ATE of reopening business was positive but not significant. The ATE of reopening restaurants and mask mandates was not significant.

We further assessed HTE to identify whether any factor moderates the intervention effects of lockdown, stay-at-home, and reopening policies. Our candidate moderators included the percentage of age 65 and over, the percentage of White, the percentage of male, and the percentage below poverty. We did not find any significant moderator. The estimated HTE for race (percentage of White race) was marginally significant for reopening bars (Web Appendix Figure S2 shows the estimated HTE and confidence interval of race on reopening bars).

## 4 | DISCUSSION

In this work, we propose a nested case-control design and propensity score weighting approach to evaluate impact of NPIs on mitigating COVID-19 transmission. Our method aligns states by transforming calendar time to time since the first reported case and allows each state to serve in both treated and control group during different time periods. Our estimator provides causal intervention effect under assumptions and we further identify the factors that moderate intervention effect. Our analysis shows that mobility restricting policies (lockdown and stay-at-home orders) have a large effect on reducing transmission. However, public health officials should be cautious about imposing them due to the social and economic costs. The effect of mask mandate was not significant. However, this result should be interpreted with care because mask mandate may not directly increase the adoption of mask wearing behavior in the public.<sup>11</sup> Using self reported mask wearing data may be more effective in evaluating the effect of masking. Reopening bars had a significant effect on increasing transmission and was more problematic than other reopening other types of business. The evidence can assist public health officials in making decisions.

In our model, we assume NPIs will become effective immediately (ie, within a relatively small period  $\Delta$ ) after being implemented. When NPIs have lagged or delayed effects, methods developed for dynamic treatment regimes

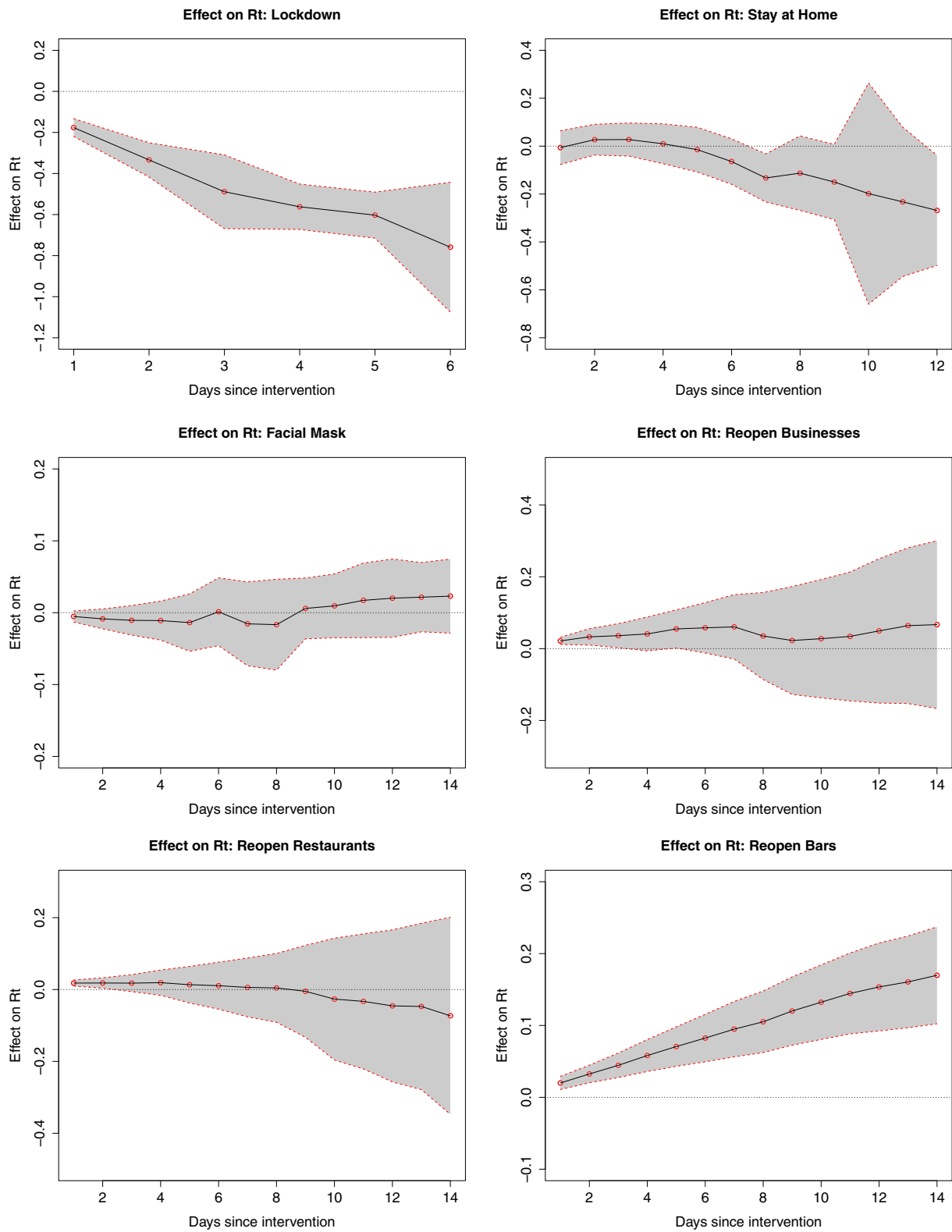


FIGURE 4 Average intervention effects with 95% confidence intervals

may be more appropriate to examine effect of a sequence of interventions. We investigated each intervention separately in this work and did not consider interaction between interventions given the sample size (50 states). To evaluate more detailed intervention packages and interaction between NPIs, county-level data can be useful to increase sample size. Assuming intervention effects to be additive, we can use the estimated treatment effect to determine the optimal sequence of the treatment effects and timing for controlling disease outbreak. We did not account for transmission between asymptomatic individuals due to a lack of reliable antibody testing data, and we assume that the transmissions occur within a state. Our assumptions might be violated if there are interference effects between neighboring states and there might be other potential confounders that are not adjusted for in the propensity score model. As an extension, for county-level analysis we can borrow spatial information from counties that are similar and adjacent to each other to account for the transmission from region to region. Other extensions to our method include using survival analysis to estimate the propensity scores for  $T_i$  or adopting a doubly robust method to improve the IPW DID estimator. There may be other hidden factors predictive of propensity scores. To uncover these factors, we can use additional data sources, such as social and behavioral data captured from Facebook. We can combine information from multiple sources using data linkage techniques. The current framework can be extended to study the effect of vaccination policies and NPIs implemented in universities and health care organizations when more data is available.

## ACKNOWLEDGEMENTS

This research was supported by U.S. NIH grants GM124104, NS073671, and MH117458. Xie was also supported by the Center of Statistical Research, and the Joint Lab of Data Science and Business Intelligence at the Southwestern University of Finance and Economics.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University.

## ORCID

Shanghong Xie  <https://orcid.org/0000-0001-5132-3887>

Yuanjia Wang  <https://orcid.org/0000-0002-1510-3315>

## REFERENCES

1. Ray EL, Wattanachit N, Niemi J, et al. Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. medRxiv; 2020. doi:10.1101/2020.08.19.20177493
2. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. *Ann Intern Med.* 2020;173(5):362-367.
3. Wang Q, Xie S, Wang Y, Zeng D. Survival-convolution models for predicting COVID-19 cases and assessing effects of mitigation strategies. *Front Public Health.* 2020;8:325.
4. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica.* 2001;69(1):201-209.
5. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health.* 2018;39:453-469.
6. Leatherdale ST. Natural experiment methodology for research: a review of how different methods can support real-world research. *Int J Soc Res Methodol.* 2019;22(1):19-35.
7. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's Tobacco Control Program. *J Am Stat Assoc.* 2010;105(490):493-505.
8. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team; 2020.
9. Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci Adv.* 2020;6(49):eabd6370.
10. Auger KA, Shah SS, T R, et al. Association between statewide school closure and COVID-19 incidence and mortality in the US. *JAMA.* 2020;324(9):859-870.
11. Rader B, White LF, Burns MR, et al. Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. *Lancet Digit Health.* 2021;3(3):e148-e157.

12. Davies NG, Kucharski AJ, Eggo RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health*. 2020;5(7):e375-e385.
13. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. 2020;584(7820):257-261.
14. Cho SW. Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: the case of Sweden. *Econom J*. 2020;23(3):323-344.
15. Ernster VL. Nested case-control studies. *Prev Med*. 1994;23(5):587-590.
16. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020;382(13):1199-1207.
17. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*. 2013;178(9):1505-1512.
18. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis*. 2020; 93:284-286.
19. Scire J, Nadeau SA, Vaughan T, et al. Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss Med Wkly*. 2020;150(19-20):w20271.
20. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020; 20(5):533-534.
21. Rader B, Scarpino SV, Nande A, et al. Crowding and the shape of COVID-19 epidemics. *Nat Med*. 2020;26(12):1829-1834.
22. Sy KT, Martinez ME, Rader B, White LF. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. *Am J Epidemiol*. 2020;190(7):1234-1242.
23. CDC. Social vulnerability index; 2020. <https://svi.cdc.gov>
24. HealthData. COVID-19 reported patient impact and hospital capacity by state; 2020. <https://healthdata.gov>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Xie S, Wang W, Wang Q, Wang Y, Zeng D. Evaluating effectiveness of public health intervention strategies for mitigating COVID-19 pandemic. *Statistics in Medicine*. 2022;41(19):3820-3836. doi: 10.1002/sim.9482

## APPENDIX. PROOF OF THEOREM 1

We use  $\mathbf{P}_n$  to denote the empirical measure associated with  $n$  states' observations and use  $\mathbf{P}$  to denote its expectation. First, using the estimating equation for  $\beta$ , we can easily show that

$$\hat{\beta} - \beta = (\mathbf{P}_n - \mathbf{P})S_\beta + o_p(n^{-1/2}), \quad (\text{A1})$$

where

$$S_\beta = \left[ \mathbf{P} \int (H(t), X)^T (H(t), X) I(T \geq t) p(t) (1 - p(t)) dF_T(t) \right]^{-1} \\ \times \int (H(t), X)^T I(T \geq t) \{I(T = t) - p(t)\} dF_T(t),$$

and we note that the matrix inverse exists due to the linear independence assumption. Thus,  $\sqrt{n}(\hat{\beta} - \beta)$  converges to a mean-zero normal distribution with covariance matrix  $E[S_\beta S_\beta^T]$ .

Next, we rewrite  $\hat{\gamma}(\Delta)$  as

$$\hat{\gamma}(\Delta) = \frac{\mathbf{P}_n \int I(T = t) / \hat{p}(t) Y(t + \Delta) d\hat{F}_T(t)}{\mathbf{P}_n \int I(T = t) / \hat{p}(t) d\hat{F}_T(t)} - \frac{\mathbf{P}_n \int I(T > t + \Delta) / (1 - \hat{p}(t)) Y(t + \Delta) d\hat{F}_T(t)}{\mathbf{P}_n \int I(T > t + \Delta) / (1 - \hat{p}(t)) d\hat{F}_T(t)}.$$

Using linear expansion and microscopic arguments, we obtain

$$\begin{aligned} \hat{\gamma}(\Delta) - \gamma(\Delta) &= \frac{(\mathbf{P}_n - \mathbf{P}) \int I(T = t)/\hat{p}(t)Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P}_n \int I(T = t)/\hat{p}(t)d\hat{F}_T(t)} \\ &\quad - \frac{\mathbf{P} \int I(T = t)/\hat{p}(t)Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P}_n \int I(T = t)/\hat{p}(t)d\hat{F}_T(t)\mathbf{P} \int I(T = t)/\hat{p}(t)d\hat{F}_T(t)}(\mathbf{P}_n - \mathbf{P}) \int I(T = t)/\hat{p}(t)d\hat{F}_T(t) \\ &\quad - \frac{(\mathbf{P}_n - \mathbf{P}) \int I(T > t + \Delta)/(1 - \hat{p}(t))Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P}_n \int I(T > t + \Delta)/(1 - \hat{p}(t))d\hat{F}_T(t)} \\ &\quad + \frac{\mathbf{P} \int I(T > t + \Delta)/(1 - \hat{p}(t))Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P}_n \int I(T > t + \Delta)/(1 - \hat{p}(t))d\hat{F}_T(t)\mathbf{P} \int I(T > t + \Delta)/(1 - \hat{p}(t))d\hat{F}_T(t)} \\ &\quad \times (\mathbf{P}_n - \mathbf{P}) \int I(T > t + \Delta)/(1 - \hat{p}(t))d\hat{F}_T(t) \\ &\quad + \frac{\mathbf{P} \int I(T = t)/\hat{p}(t)Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P} \int I(T = t)/\hat{p}(t)d\hat{F}_T(t)} - \frac{\mathbf{P} \int I(T > t + \Delta)/(1 - \hat{p}(t))Y(t + \Delta)d\hat{F}_T(t)}{\mathbf{P} \int I(T > t + \Delta)/(1 - \hat{p}(t))d\hat{F}_T(t)} - \gamma(\Delta). \end{aligned}$$

On the other hand, based on assumptions (a) to (c), using the same argument in Section 2, we know

$$\gamma(\Delta) = \frac{\mathbf{P} \int I(T = t)/p(t)Y(t + \Delta)dF_T(t)}{\mathbf{P} \int I(T = t)/p(t)dF_T(t)} - \frac{\mathbf{P} \int I(T > t + \Delta)/(1 - p(t))Y(t + \Delta)dF_T(t)}{\mathbf{P} \int I(T > t + \Delta)/(1 - p(t))dF_T(t)}.$$

Thus, the last term in the expansion of  $\hat{\gamma}(\Delta) - \gamma(\Delta)$  can be further expanded as a linear functional of  $(\hat{\beta} - \beta)$  and  $(\hat{F} - F)$ , where we further plug in the expansion in (A1) and note that  $(\hat{F} - F_T)(t) = (\mathbf{P}_n - \mathbf{P})I(T \leq t)$ .

Finally, since  $I(T \geq t), \hat{p}(t), \hat{F}_T(t)$  and  $Y(t)$  have bounded total variations so they are  $P$ -Donsker, we conclude

$$\hat{\gamma}(\Delta) - \gamma(\Delta) = (\mathbf{P}_n - \mathbf{P})\Gamma + o_p(n^{-1/2}),$$

where if we define

$$A = \int I(T = t)/p(t)Y(t + \Delta)dF_T(t), \quad B = \int I(T = t)/p(t)dF_T(t)$$

and

$$\begin{aligned} C &= \int I(T > t + \Delta)/(1 - p(t))Y(t + \Delta)dF_T(t), \quad D = \int I(T > t + \Delta)/(1 - p(t))dF_T(t), \\ \Gamma &= \frac{A}{E[B]} - \frac{C}{E[D]} - \frac{E[A]}{E[B]^2} \int \frac{I(T = t)}{p(t)}dF_T(t) + \frac{E[C]}{E[D]^2} \int \frac{I(T > t + \Delta)}{1 - p(t)}dF_T(t) \\ &\quad + \frac{\tilde{E} \left[ \frac{I(\tilde{T}=T)}{p(T)} \tilde{Y}(T + \Delta) \right]}{E[B]} - \frac{E[A]}{E[B]^2} \tilde{E} \left[ \frac{I(\tilde{T} = T)}{p(T)} \right] \\ &\quad - \frac{\tilde{E} \left[ \frac{I(\tilde{T}>T+\Delta)}{1-p(T)} \tilde{Y}(T + \Delta) \right]}{E[D]} + \frac{E[C]}{E[D]^2} \tilde{E} \left[ \frac{I(\tilde{T} > T + \Delta)}{1 - p(T)} \right] \\ &\quad - \left[ \frac{\mathbf{P} \int I(T = t) \frac{1-p(t)}{p(t)} (H(t), X)^T Y(t + \Delta) dF_T(t)}{\mathbf{P} \int \frac{I(T=t)}{p(t)} dF_T(t)} \right. \\ &\quad \left. - \frac{\mathbf{P} \int \frac{I(T=t)}{p(t)} Y(t + \Delta) dF_T(t) \mathbf{P} \int I(T = t) \frac{1-p(t)}{p(t)} (H(t), X)^T dF_T(t)}{(\mathbf{P} \int I(T = t)/p(t)dF_T(t))^2} \right] S_\beta \end{aligned}$$



$$- \left[ \frac{\mathbf{P} \int I(T > t + \Delta) \frac{p(t)}{1-p(t)} (H(t), X)^T Y(t + \Delta) dF_T(t)}{\mathbf{P} \int \frac{I(T > t + \Delta)}{1-p(t)} dF_T(t)} - \frac{\mathbf{P} \int \frac{I(T > t + \Delta)}{1-p(t)} Y(t + \Delta) dF_T(t) \mathbf{P} \int I(T > t + \Delta) \frac{p(t)}{1-p(t)} (H(t), X)^T dF_T(t)}{\left( \mathbf{P} \int \frac{I(T > t + \Delta)}{1-p(t)} dF_T(t) \right)^2} \right] S_\beta.$$

Here,  $\tilde{E}[\cdot]$  denotes the expectation with  $\tilde{T}$  and  $\tilde{Y}$ . Therefore,  $\sqrt{n}(\hat{\gamma}(\Delta) - \gamma(\Delta))$  converges to a mean-zero normal distribution with variance  $E[\Gamma\Gamma^T]$ .