Received: 5 May 2022 Accepted: 23 March 2023

DOI: 10.1111/biom.13862

# Identifying and estimating effects of sustained interventions under parallel trends assumptions

**Audrey Renson**[1] | **Michael G. Hudgens**[2] | **Alexander P. Keil**[3] | **Paul N. Zivich**[3] | **Allison E. Aiello**[4]

[1]Department of Population Health, New York University Grossman School of Medicine, New York, New York, USA

[2]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA

[3]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[4]Columbia Aging Center and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA

**Correspondence**
Audrey Renson, Department of Population Health, New York University Grossman School of Medicine, New York, New York, USA.
Email: audrey.o.renson@gmail.com

**Funding information**
Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Numbers: P2C HD050924, T32-HD091058-02; National Institute of Allergy and Infectious Diseases, Grant/Award Numbers: R01 AI085073, T32-AI007001

**Abstract**

Many research questions in public health and medicine concern sustained interventions in populations defined by substantive priorities. Existing methods to answer such questions typically require a measured covariate set sufficient to control confounding, which can be questionable in observational studies. Differences-in-differences rely instead on the parallel trends assumption, allowing for some types of time-invariant unmeasured confounding. However, most existing difference-in-differences implementations are limited to point treatments in restricted subpopulations. We derive identification results for population effects of sustained treatments under parallel trends assumptions. In particular, in settings where all individuals begin follow-up with exposure status consistent with the treatment plan of interest but may deviate at later times, a version of Robins' g-formula identifies the intervention-specific mean under stable unit treatment value assumption, positivity, and parallel trends. We develop consistent asymptotically normal estimators based on inverse-probability weighting, outcome regression, and a double robust estimator based on targeted maximum likelihood. Simulation studies confirm theoretical results and support the use of the proposed estimators at realistic sample sizes. As an example, the methods are used to estimate the effect of a hypothetical federal stay-at-home order on all-cause mortality during the COVID-19 pandemic in spring 2020 in the United States.

**KEYWORDS**
causal inference, difference-in-differences, g-formula, observational study, unmeasured confounding

## 1 | INTRODUCTION

Many epidemiologic and other empirical studies concern the effects of sustained treatment strategies on population average outcomes over time. A sustained treatment or intervention is one that sets values of a time-varying exposure via a predetermined plan or algorithm. For example, clinical studies are often concerned with optimal dosing plans for therapeutic drugs, and policy investigations are often concerned with policies that determine exposure distributions repeatedly over time for the population residing in a jurisdiction. To be concrete, suppose a binary exposure $A_t \in \{0, 1\}$ is measured at three time points ($t = 0, 1, 2$); then one example of sustained intervention would be the plan to set $A_0 = 0$, $A_1 = 1$, and $A_2 = 0$.

Existing approaches to estimating effects of sustained interventions in well-defined populations include g-computation (Robins, 1986), inverse probability of treatment weighted (IPTW) marginal structural models (Robins, 2000; Robins et al., 2000), g-estimation of structural nested models (Robins, 1989), and double robust methods such as augmented IPTW (Bang & Robins, 2005) and targeted maximum likelihood (van der Laan & Gruber, 2012). Importantly, these approaches base causal identification on a sequential version of exchangeability (Robins, 1986), also known as sequential ignorability or no unmeasured confounders (Robins, 2000). Sequential exchangeability posits that the potential outcomes are independent of treatment, given the history of some set of measured (possibly time-varying) covariates and treatment; this assumption is unverifiable and can be implausible in many settings. For example, individuals may select medical treatments based on unmeasured risk factors, and public policies are decided in highly complex political contexts that may influence health.

In contrast, difference-in-differences (DID) methods typically base identification on parallel trends assumptions rather than sequential exchangeability (Ashenfelter & Card, 1985; Roth et al., 2022). Parallel trends assumptions posit that time trends in average potential outcomes are independent of the observed treatment (Ashenfelter & Card, 1985; Marcus & Sant'Anna, 2021). DID methods typically focus on the average treatment effect in the treated (ATT) for a treatment occurring at a single time point, although recently extensions have considered certain types of sustained treatment regimes (for a review, see Roth et al., 2022). In particular, Callaway and Sant'Anna (2021), de Chaisemartin and D'Haultfoeuille (2020), and de Chaisemartin and D'Haultfoeuille (2021b) consider effects of time-varying treatments conditional on each observed treatment path. Relevant to the present work, these recent DID developments have included doubly robust estimators (Callaway & Sant'Anna, 2021; Sant'Anna & Zhao, 2020). In the majority of recent time-varying extensions to DID, group-time-specific ATTs (or similarly conditional parameters) are estimated, and typically averaged in some way to obtain an overall result. The parameters targeted by these averages may be challenging to interpret because the "treated" group changes over time, so that an average represents a dynamic population rather than a well-defined population (as in a randomized trial). Therefore, the resulting parameters are not causal effects in the sense that they are not contrasts between potential outcomes under different interventions for the same population (Maldonado & Greenland, 2002).

Departing from treatment-conditional parameters, de Chaisemartin and D'Haultfoeuille (2021a) consider interventions fixing a time-varying exposure to its baseline status, focusing on unconditional cost-effectiveness ratios and outcome regression estimators. Similarly, in this paper, we consider inference about marginal effects of general sustained treatment strategies under parallel trends assumptions. The proposed estimators build on the IPTW, g-computation, and doubly-robust targeted maximum likelihood estimation (TMLE) approaches developed in the context of sequential exchangeability, thus providing important links between the biostatistics literature on time-varying treatments (Bang & Robins, 2005; Robins, 1986, 2000; van der Laan & Gruber, 2012) and the econometric literature on DID (Ashenfelter & Card, 1985; Callaway & Sant'Anna, 2021). Independently and concurrently with the present work, Shahn et al. (2022) developed g-estimation of structural nested models for general sustained treatment regimes under parallel trends, with results that imply identification for the intervention-specific means considered here. While it is possible (but complex) to estimate the latter quantity using the g-estimation approach of Shahn et al. (2022), the main strength of g-estimation is in exploring effect heterogeneity by time-varying covariates. The proposed methods thus complement the existing DID literature by focusing on different causal parameters and the g-estimation methods of Shahn et al. (2022) by providing alternative estimation strategies.

## 2 | PRELIMINARIES

### 2.1 | Data

Suppose data $O_{it} = \{W_{it}, A_{it}, Y_{it}\}$ are observed on $i = 1, 2, \dots, n$ individuals (or units) at time points $t = 0, 1, \dots, \tau$, where $W_{it}$ are (possibly vector-valued) covariates; $A_{it}$ are discrete, possibly multivariate treatments realized after $W_{it}$; and $Y_{it}$ are outcomes realized after $A_{it}$, all measured without error. Denote history of a variable with overbars, for example, $\overline{A}_{it} = (A_{i0}, A_{i1}, \dots, A_{it})$, with $\overline{A}_i \equiv \overline{A}_{i\tau}$ and $A_{ik} = \{\emptyset\}$ for $k < 0$ by convention. Upper case is used throughout to refer to random variables, lower case refers to specific realizations, and scripts refer to the support. The $i$ subscript is omitted unless needed to resolve ambiguity. Throughout, it is assumed that $\overline{O}_i \equiv \{\overline{W}_i, \overline{A}_i, \overline{Y}_i\}$ $(i = 1, 2, \dots, n)$ represent independent and identically distributed (iid) draws from a relevant target population.

Assume the data come from a *staggered discontinuation design*, defined as follows. Suppose the target estimand is $\mathbb{E}\{Y_t(\overline{a}^*)\}$ where $\overline{a}^* = (a_0^*, a_1^*, \dots, a_\tau^*)$ denotes the treatment strategy or intervention plan of interest, and $Y_t(\overline{a}^*)$

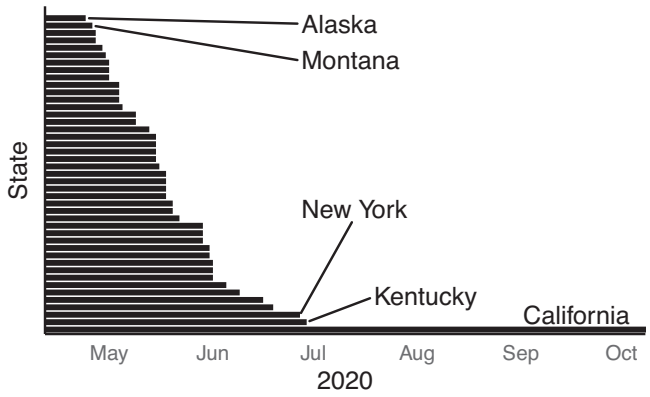


**FIGURE 1** Dates of state-issued stay-at-home orders in U.S. states during the COVID-19 pandemic in 2020.

denotes a potential outcome; that is, the value $Y_t$ would take under the intervention setting $\overline{A} = \overline{a}^*$. The approach in this paper requires that in the observed data distribution, $\Pr(A_0 = a_0^*) = 1$, or in other words that $A_{i0} = a_0^*$ for all $i$. Such a scenario is called a staggered discontinuation design with respect to the treatment plan $\overline{a}^*$ of interest, because units begin follow-up under the treatment plan but may discontinue at later points in a staggered way. Note that we do not require monotonic treatment assignment, in contrast to recent DID papers (e.g., Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021).

## 2.2 | Motivating example

Consider the question, "what effects on all-cause mortality would a U.S. federal stay-at-home order have had in spring 2020 during the COVID-19 pandemic?" Let $Y_{it}$ be a binary indicator that individual $i$ died during week $t$, $t = 0, 1, \ldots, 11$, measured as weeks since April 6, 2020. Let $A_{it}$ be a binary indicator that the state in which individual $i$ was living during week $t$ was under a state-level stay-at-home or shelter-in-place order. Suppose that it is of interest to estimate $\mathbb{E}\{Y_t(\overline{1})\} - \mathbb{E}(Y_t)$, the difference in U.S. mortality rates under a hypothetical federal stay-at-home order versus under the observed treatment trajectory (i.e., the "natural course"). As of April 6, 43/50 U.S. states were under stay-at-home orders, which were discontinued at times ranging from late April to late June, with the exception of California that continued through December (Figure 1). Thus, the observed treatment trajectories give rise to a staggered discontinuation design with respect to the treatment plan $\overline{a}^* = \overline{1}$ setting everyone to remain under stay-at-home order in those 43 states. The methods developed below can be used to draw inference about what would have happened had such a policy been implemented.

# 3 | IDENTIFICATION

In this section, we consider identification, given data $\overline{O}$, of the quantity $\mu_t \equiv \mathbb{E}\{Y_t(\overline{a}^*)\}$, the mean outcome at time $t$, under the intervention to set all individuals to $\overline{A}_i = \overline{a}^*$. Throughout, it is assumed that interest lies in only one intervention $\overline{a}^*$. Note that $\mu_t$ depends on $\overline{a}^*$, which is left implicit for notational simplicity. Consider the following assumptions:

**Assumption 1.** (Stable unit treatment value assumption [SUTVA]): If $\overline{A}_{it} = \overline{a}_t^*$, then $Y_{it} = Y_{it}(\overline{a}_t^*)$ for $t \in \{0, 1, \ldots \tau\}$.

**Assumption 2.** (Positivity): If $f(\overline{w}_t|\overline{A}_{t-1} = \overline{a}_{t-1}^*) > 0$, then $f(\overline{a}_t^*|\overline{W}_t = \overline{w}_t, \overline{A}_{t-1} = \overline{a}_{t-1}^*) > 0$, for $\overline{w}_t \in \overline{\mathcal{W}}_t; t \in \{1, 2, \ldots, \tau\}$.

Here and throughout, $f(x|\cdot)$ refers to a conditional density if $X$ is continuous, and a conditional probability mass function if $X$ is discrete. Assumption 2 requires that units whose treatment history up to time $t-1$ is consistent with the regime in question ($\overline{a}^*$) have positive probability of remaining under treatment plan $\overline{a}^*$ at time $t$. Positivity can sometimes be a verifiable assumption. In particular, if both $\overline{W}$ and $\overline{A}$ are low-dimensional and discrete, then among units with $\overline{A}_{t-1} = \overline{a}_{t-1}^*$, if one observes units who remain under $\overline{a}^*$ at time $t$ in every stratum of $\overline{W}_t$, this implies positivity in the population with probability 1 (but not necessarily the reverse).

**Assumption 3.** (Parallel trends): For $t \in \{1, 2, \ldots, \tau\}, k \leq t$ :

$$\begin{aligned}\mathbb{E}\{Y_t(\overline{a}^*) \; &- Y_{t-1}(\overline{a}^*)|\overline{W}_k, \overline{A}_{k-1} = \overline{a}_{k-1}^*\} \\ &= \mathbb{E}\{Y_t(\overline{a}^*) - Y_{t-1}(\overline{a}^*)|\overline{W}_k, \overline{A}_k = \overline{a}_k^*\}.\end{aligned}$$

In words, Assumption 3 states that among individuals whose treatment status is consistent with the intervention $\overline{a}^*$ up to time $k-1$, had (counter to fact) all individuals followed the intervention through time $t$, trends would have been parallel for those who do and do not follow the intervention at time $k$ but have equal covariate histories. Assumption 3 is identical to the time-varying conditional parallel trends assumption of Shahn et al. (2022). Assumption 3 is similar to those adopted in event study and staggered adoption DID methods, with the key distinction that typically only a "never-treated" regime ($\overline{a}^* = \overline{0}$) for a binary or ordinal treatment is considered. Aside from considering more general regimes $\overline{a}^*$, Assumption 3 is identical to the Assumption 5 in Callaway and Sant'Anna (2021) except that the latter only conditions on baseline covariates. Similarly, Assumption 12 in de Chaisemartin

and D'Haultfoeuille (2021a) only conditions on contemporaneous values of time-varying covariates (rather than the full history) and presumes a parametric model relating these covariates to the outcome trends. Thus, the proposed assumption generalizes the standard parallel trends assumptions by allowing for conditioning on the whole history of time-varying covariates and by considering any *static* regime $\overline{a}^*$. A static regime is one whose values are fixed and do not depend on covariates (Section 7.2 considers extensions of the proposed methods to dynamic regimes that can depend on the history of covariates).

Parallel trends are unverifiable, though closely related conditions can often be checked (Roth, 2019). Note that $W_{it}$ may include prior outcomes $Y_{im}$ for $m < t$. However, if $Y_{i,t-1}$ is included in $W_{it}$, then the parallel trends assumption is equivalent to sequential exchangeability, in which case existing causal inference methods for observational data with a longitudinal exposure can be used (e.g., Robins, 1986, 2000; van der Laan & Gruber, 2012).

An important consideration is whether a given parallel trends assumption places restrictions on treatment effect heterogeneity. Just as the standard parallel trends assumptions in the DID literature place restrictions only on untreated potential outcomes, Assumption 3 places restrictions only on potential outcomes corresponding to the regime $\overline{a}^*$, and therefore, does not imply any restrictions on treatment effect heterogeneity. However, additionally, assuming parallel trends for any other regimen $\overline{a} \neq \overline{a}^*$ would imply such restrictions (Callaway et al., 2021; Shahn et al., 2022).

The following lemma presents the main identification results in this paper, which show that Assumptions 1–3 are sufficient to equate $\mu_t$ to a function of the observed data distribution.

**Lemma 1.** *(Parallel trends g-formula) Define the functional (i.e., statistical parameter)* $\psi_t \equiv \mathbb{E}(Y_0) + \sum_{k=1}^{t} \int \mathbb{E}(Y_k - Y_{k-1} | \overline{W}_k = \overline{w}_k, \overline{A}_k = \overline{a}_k^*) \prod_{m=0}^{k} dF(w_m | \overline{w}_{m-1}, \overline{a}_{m-1}^*)$. *Under a staggered discontinuation design and if Assumptions 1–3 hold, then* $\psi_t = \mu_t$.

Here and throughout, $F(\cdot|\cdot)$ refers to a conditional cumulative distribution function. Lemma 1 states that the target causal quantity $\mu_t$ is identified by the parameter $\psi_t$. The parameter $\psi_t$ is referred to as the *parallel trends g-formula* because it represents a modification of the usual g-formula (the dependence of $\psi_t$ on $\overline{a}^*$ is also left implicit). A formal proof of Lemma 1 by induction is presented in Web Appendix A. Here we give a less formal explanation to build intuition. We have:

$$\begin{aligned}
\mu_t &\equiv \mathbb{E}\{Y_t(\overline{a}^*)\} = \mathbb{E}\{Y_0(\overline{a}^*)\} + \sum_{k=1}^{t} \mathbb{E}\{Y_k(\overline{a}^*) - Y_{k-1}(\overline{a}^*)\} \\
&= \mathbb{E}\{Y_0(\overline{a}^*)\} + \sum_{k=1}^{t} \mathbb{E}[\mathbb{E}\{Y_k(\overline{a}^*) - Y_{k-1}(\overline{a}^*) | \overline{W}_1\}] \\
&= \mathbb{E}\{Y_0(\overline{a}^*)\} + \sum_{k=1}^{t} \mathbb{E}[\mathbb{E}\{Y_k(\overline{a}^*) - Y_{k-1}(\overline{a}^*) | A_1 = a_1^*, \overline{W}_1\}],
\end{aligned}$$

where the first equality follows by adding and subtracting constants, the second by iterated expectation, and the third by Assumption 3. Repeatedly applying iterated expectation and Assumption 3, we have:

$$\begin{aligned}
\mu_t &= \mathbb{E}\{Y_0(\overline{a}^*)\} \\
&\quad + \sum_{k=1}^{t} \mathbb{E}\{\mathbb{E}(\cdots \mathbb{E}[\mathbb{E}\{Y_k(\overline{a}^*) - Y_{k-1}(\overline{a}^*) | \\
&\quad \overline{A}_k = \overline{a}_k^*, \overline{W}_k\} | \overline{A}_{k-1} = \overline{a}_{k-1}^*, \overline{W}_{k-1}] \cdots | A_1 = a_1^*, \overline{W}_1)\} \\
&= \mathbb{E}(Y_0) + \sum_{k=1}^{t} \mathbb{E}(\mathbb{E}[\cdots \mathbb{E}\{\mathbb{E}(Y_k - Y_{k-1} | \\
&\quad \overline{A}_k = \overline{a}_k^*, \overline{W}_k) | \overline{A}_{k-1} = \overline{a}_{k-1}^*, \overline{W}_{k-1}\} \cdots | A_1 = a_1^*, \overline{W}_1]) \\
&\quad = \psi_t,
\end{aligned}$$

where the second equality follows from Assumption 1 and the last equality from iterated expectation.

## 4 | ESTIMATORS

This section presents estimators for the statistical parameter $\psi_t$, which equals the target quantity $\mu_t$ under the above stated assumptions. The estimators in this section utilize existing estimators of $\mu_t$ under sequential exchangeability rather than parallel trends, all of which are consistent and asymptotically normal (CAN) estimators of the g-formula by virtue of being solutions to unbiased estimating equations (Stefanski & Boos, 2002). Since $\psi_t$ is a continuous function of several g-formulas, the same function applied to estimators of those g-formulas is a CAN estimator for $\psi_t$. The remainder of this section formalizes this logic and gives examples of specific estimators that function in this capacity. The estimators presented in this section are provided in an R package (see the Supporting Information).

## 4.1 | General form

Here we derive a general form of a CAN estimator for the target statistical parameter, $\psi_t$. First, define:

$$\phi_{j,k} = \int \mathbb{E}(Y_j|\overline{A}_k = \overline{a}_k^*, \overline{W}_k = \overline{w}_k) \times \prod_{m=0}^{k} dF(w_m|\overline{W}_{m-1} = \overline{w}_{m-1}, \overline{A}_{m-1} = \overline{a}_{m-1}^*). \quad (1)$$

Equation (1) is the g-formula, developed in the context of identifying parameters like $\mu_t$ under sequential exchangeability. Here, sequential exchangeability is not assumed, and therefore, $\phi_{j,k}$ is not interpretable as a causal parameter; instead, existing estimators of the statistical parameter $\phi_{j,k}$ are used to assemble estimators of $\psi_t$ (which equals the causal parameter $\mu_t$ under Assumptions 1–3) by noting that $\psi_t = \phi_{0,0} + \sum_{k=1}^{t}(\phi_{k,k} - \phi_{k-1,k})$. Next, suppose that there is an estimator $\widehat{\phi}_{jk}$ of $\phi_{jk}$ that is the solution to an unbiased estimating function $d_{\phi_{jk}}(O;\phi_{jk})$; that is, $0 = \mathbb{E}\{d_{\phi_{jk}}(O;\widehat{\phi}_{jk})\}$. Let $\boldsymbol{\phi} = (\phi_{0,0}, \phi_{0,1}, \ldots, \phi_{t-1,t}, \phi_{t,t})$. Then simply define

$$d_{\psi_t}(O;\boldsymbol{\phi},\psi_t) = \phi_{0,0} + \sum_{k=1}^{t}\left(\phi_{k,k} - \phi_{k-1,k}\right) - \psi_t. \quad (2)$$

Clearly, $\mathbb{E}\{d_{\psi_t}(O;\boldsymbol{\phi},\psi_t)\} = 0$, indicating that an estimator $(\widehat{\boldsymbol{\phi}},\widehat{\psi}_t)$ that jointly solves $0 = \sum_i d_{\psi_t}(O_i;\widehat{\boldsymbol{\phi}},\widehat{\psi}_t)$ will yield a CAN estimator $\widehat{\psi}_t$ for $\psi_t$.

The following subsections show how different options for $d_{\phi_{jk}}(O;\phi_{jk})$ (including IPTW, g-computation, and TMLE) can be constructed and stacked with (2) to form estimators of $\psi_t$ that inherit desirable properties (e.g., consistency, asymptotic normality, double robustness).

## 4.2 | Inverse probability of treatment weighted estimator

Define stabilized inverse probability of treatment weights (IPTWs) as $\pi_k(\overline{a};\overline{W}) = \prod_{m=1}^{k} f(a_m|\overline{a}_{m-1}) / \prod_{m=1}^{k} f(a_m|\overline{a}_{m-1},\overline{W}_m)$. Robins (2000) (Lemma 1.1) showed that $\phi_{j,k} = c_{jk}(\overline{a}_k^*)$, where $c_{jk}(\cdot)$ is the unique function such that $\mathbb{E}[q_{jk}(\overline{A}_k)\{Y_j - c_{jk}(\overline{A}_k)\}\pi_k(\overline{A};\overline{W})] = 0$ for all functions $q_{jk}(\overline{A}_k)$ where the expectation exists. The equation

$\mathbb{E}[q_{jk}(\overline{A}_k)\{Y_j - c_{jk}(\overline{A}_k)\}\pi_k(\overline{A}_k;\overline{W})] = 0$ defines a regression of $Y_j$ on $\overline{A}_k$ weighted by $\pi_k(\overline{A};\overline{W})$. In the context of sequential exchangeability, estimators based on this weighted regression formulation are called IPTW-marginal structural model estimators and given a causal interpretation (Robins et al., 1992; Robins, 2000). However, in the context of this paper, sequential exchangeability is not assumed, and therefore, the weighted regression equation does not have a causal interpretation on its own. Instead, the above result together with results from Section 4.1 implies that an IPTW estimator of $\psi_t$ can be formed by using a linear combination of IPTW marginal structural model estimators for $\phi_{j,k}$.

For simplicity of presentation, assume that for $k = 0, 1, \ldots, t$, $f(a_k|\overline{a}_{k-1})$ and $f(a_k|\overline{a}_{k-1},\overline{w}_k)$ are known up to a finite-dimensional parameter. That is, define $g_{0,k}(\overline{a},\emptyset) \equiv \prod_{s=0}^{k} f(a_s|\overline{a}_{s-1})$ and $g_{0,k}(\overline{a},\overline{w}) \equiv \prod_{s=0}^{k} f(a_s|\overline{a}_{s-1},\overline{w}_s)$, and say we are willing to assume that $g_{0,k}(\overline{a},\emptyset)$ is uniquely determined by the parametric model $g_k(\overline{a},\emptyset;\alpha_0)$, and similarly that $g_{0,k}(\overline{a},\overline{w}) = g_k(\overline{a},\overline{w};\alpha_1)$, where $\alpha_0$ and $\alpha_1$ are finite-dimensional parameter vectors. Say we have an estimator that solves an unbiased estimating equation for $(\alpha_0,\alpha_1)$. For example, $g_k(\overline{a},\emptyset;\alpha_0)$ and $g_k(\overline{a},\overline{w};\alpha_1)$ may consist of generalized linear models with parameters estimated by maximum likelihood. Specify $c_{jk}(\overline{A}_k)$ as some appropriate functional form for the expected value of $Y_j$ conditional on $\overline{A}_k$ in the weighted data distribution, such as $c_{jk}(\overline{A}_k) = \gamma_{0jk} + \gamma_{1jk}I(\overline{A}_k = \overline{a}_k^*)$ (i.e., leaving the model unrestricted when $\overline{A}_k = \overline{a}_k^*$). Then an estimator $\widehat{\phi}_{j,k}^{IPTW}$ that solves $0 = c_{jk}(\overline{a}_k^*) - \phi_{j,k}$ is CAN for $\phi_{j,k}$ if $c_{jk}(\overline{A}_k), g_k(\overline{A},\emptyset;\alpha_0)$ and $g_k(\overline{A},\overline{w};\alpha_1)$ are correctly specified. Finally, stack the score equations for $g_k(\overline{a},\emptyset;\alpha_0)$ and $g_k(\overline{a},\overline{w};\alpha_1)$, along with $0 = c_{jk}(\overline{a}_k^*) - \phi_{j,k}$ $(k = 0, 1, \ldots, t; j = k-1, k)$ and equation (2) to yield an estimator for $\psi_t$, say $\widehat{\psi}_t^{IPTW}$.

In other words, the IPTW estimator for the target parameter of interest is $\widehat{\psi}_t^{IPTW} = \widehat{\phi}_{0,0}^{IPTW} + \sum_{k=1}^{t}(\widehat{\phi}_{k,k}^{IPTW} - \widehat{\phi}_{k-1,k}^{IPTW})$, where $\widehat{\phi}_{j,k}^{IPTW}$ are estimators of each appropriate g-formula parameter based on an IPTW model. Note that under our assumption set, $\widehat{\phi}_{j,k}^{IPTW}$ are not estimators of causal quantities in and of themselves, but simply functions of the observed data distribution that may be assembled appropriately to form the causal estimator $\widehat{\psi}_t^{IPTW}$. Clearly, $\widehat{\psi}_t^{IPTW}$ solves an estimating equation that is unbiased if $g_k(\overline{A},\emptyset;\alpha_0), g_k(\overline{A},\overline{w};\alpha_1)$, and $c_{jk}(\overline{A}_k)$ are all correctly specified, implying that $\widehat{\psi}_t^{IPTW}$ is CAN for $\psi_t$ under the same conditions. However, IPTW estimators are known to be inefficient and $\widehat{\psi}_t^{IPTW}$ may similarly inherit this property. The following subsections present estimators that may improve on efficiency relative to IPTW.

## 4.3 | Iterated conditional expectation estimator

Bang and Robins (2005) describe an estimator of $\phi_{j,k}$ based on the following iterated conditional expectation (ICE) representation:

$$\phi_{j,k} = \mathbb{E}(\mathbb{E}[\cdots\mathbb{E}\{\mathbb{E}(Y_j|\overline{A}_k = \overline{a}^*_k, \overline{W}_k)| \overline{A}_{k-1} = \overline{a}^*_{k-1}, \overline{W}_{k-1}\}\cdots|\overline{A}_1 = \overline{a}^*_1, \overline{W}_1]|A_0 = a^*_0, W_0), \quad (3)$$

which can equivalently be written as $\phi_{j,k} = \mathbb{E}\{Q_0^{j,k,0}(\overline{a}^*)\}$ where, for $m = 0, 1, \ldots, k$, $Q_0^{j,k,m}(\overline{a}^*) = \mathbb{E}\{Q_0^{j,k,m+1}(\overline{a}^*)|\overline{A}_m = \overline{a}^*_m, \overline{W}_m\}$ and $Q_0^{j,k,k+1}(\overline{a}^*) = Y_j$.

An estimator of $\psi_t$ can then be formulated based on this representation. For simplicity, say we are willing to assume that $Q_0^{j,k,m}(\overline{a}^*)$ are known up to a finite-dimensional parameter for $m = 0, 1, \ldots, k$. That is, assume $Q_0^{j,k,m}(\overline{a}^*) = Q^{j,k,m}(\overline{a}^*; \beta_m)$, where $\beta_m$ for $m = 0, 1, \ldots, k$ are finite-dimensional parameters. For example, $Q^{j,k,m}(\overline{a}^*; \beta_m)$ may be a generalized linear model with parameters $\beta_m$. Say we are in possession of an unbiased estimating function $d_{j,k,m}\{O, Q^{j,k,m+1}(\overline{a}^*; \beta_{m+1}); \beta_m\}$ for $\beta_m$. For example, if maximum likelihood is used, then $d_{j,k,m}\{O, Q^{j,k,m+1}(\overline{a}^*; \beta_{m+1}); \beta_m\}$ is the vector of first derivatives of the model log-likelihood with respect to $\beta_m$. Note that including $Q^{j,k,m+1}(\overline{a}^*; \beta_{m+1})$ as an argument to the estimating function makes explicit the nested nature of the iterated expectations being modeled. The ICE estimator of $\phi_{j,k}$ is then defined (Bang & Robins, 2005) as the solution $\widehat{\phi}^{ICE}_{j,k}$ to $\mathbf{0} = \sum_{i=1}^n d_{j,k}(O_i; \phi_{j,k})$ where

$$d_{j,k}(O; \phi_{j,k}) = \begin{pmatrix} d_{j,k,k}(O; \beta_k) \\ d_{j,k,k-1}\{O, Q^{j,k,k}(\overline{a}^*; \beta_k); \beta_{k-1}\} \\ \vdots \\ d_{j,k,0}\{O, Q^{j,k,1}(\overline{a}^*; \beta_1); \beta_0\} \\ Q^{j,k,0}(\overline{a}^*; \beta_0) - \phi_{j,k} \end{pmatrix}.$$

Then, simply stack $d_{j,k}(O; \phi_{j,k})$ with (2) to yield an estimator $\widehat{\psi}^{ICE}_t$ for $\psi_t$. In other words, the ICE estimator of the target parameter is $\widehat{\psi}^{ICE}_t = \widehat{\phi}^{ICE}_{0,0} + \sum_{k=1}^t (\widehat{\phi}^{ICE}_{k,k} - \widehat{\phi}^{ICE}_{k-1,k})$, where each $\widehat{\phi}^{ICE}_{j,k}$ is an estimator of the corresponding g-formula parameter based on ICE g-computation. Clearly, $\widehat{\psi}^{ICE}_t$ solves an unbiased estimating equation whenever all the iterated outcome models $\{Q^{j,k,m}(\overline{a}^*; \beta_m) : k = 0, 1, \ldots, t; j = k, k-1; m = 0, 1, \ldots, k+1\}$ are correctly specified. Estimators of $\phi_{j,k}$ based on outcome regression generally have smaller asymptotic variance that IPTW estimators, and $\widehat{\psi}^{ICE}_t$ may inherit this property.

## 4.4 | Doubly robust targeted maximum likelihood estimator

IPTW estimators are only guaranteed to be CAN if the treatment models are correctly specified, and ICE estimators are only guaranteed to be CAN if all the outcome models are correctly specified. Doubly robust estimators are CAN if either the outcome or treatment models are correct (but not necessarily both), which is an advantage because one is rarely certain that models are correctly specified.

Doubly robust estimators of $\phi_{j,k}$ generally consist of augmenting the ICE algorithm by including predicted values from the treatment models used to construct IPTWs in some way. Such estimators are called semiparametric efficient if they solve the estimating equation corresponding to the following efficient influence curve (Tran et al., 2019; van der Laan & Gruber, 2012):

$$\sum_{m=0}^k \frac{I(\overline{A}_m = \overline{a}^*_m)}{g_{0,m}(\overline{a}^*, \overline{w})}\{Q_0^{j,k,m+1}(\overline{a}^*) - Q_0^{j,k,m}(\overline{a}^*)\} + Q_0^{j,k,0}(\overline{a}^*) - \phi_{j,k} \quad (4)$$

with $g_{0,m}(\overline{a}^*, \overline{w})$ and $Q_0^{j,k,m}$ defined as in previous sections. Many estimators correspond to this efficient influence curve, meaning that they all have the smallest asymptotic variance of any regular asymptotically linear estimator in this class (Bang & Robins, 2005; van der Laan & Gruber, 2012). We present one such example of a TMLE that may outperform others in finite samples (Tran et al., 2019).

First consider the TMLE of $\phi_{j,k}$. For simplicity, assume that outcome models $\{Q_0^{j,k,m}(\overline{a}^*) : m = 0, 1, \ldots, k\}$ and treatment models $\{g_{0,m}(\overline{a}^*) : m = 0, 1, \ldots, k\}$ are known up to a finite-dimensional parameter. That is, assume $g_{0,m}(\overline{a}^*) = g_m(\overline{a}^*; \alpha_m)$ and $Q_0^{j,k,m}(\overline{a}^*) = Q^{j,k,m}(\overline{a}^*; \beta_m)$, where $\alpha_m$ and $\beta_m$ are finite-dimensional parameters, $m = 0, 1, \ldots, k$. Then proceed as follows:

1. For $m = 0, 1, \ldots, k$, estimate $\alpha_m$, for example, using maximum likelihood. Denote estimators of $\alpha_m$ as $\widehat{\alpha}_m$ and corresponding estimators of $g_m(\overline{a}^*; \alpha_m)$ as $g_m(\overline{a}^*; \widehat{\alpha}_m)$.
2. For $m = k$, estimate $\beta_m$, for example, using maximum likelihood, denoting this estimator $\widehat{\beta}_m$. Calculate $Q_i^{j,k,m}(\overline{a}^*; \widehat{\beta}_m)$ for each unit $i$ and denote this estimator $\widehat{Q}_i^{j,k,m}(\overline{a}^*)$. Note that these are model predictions that implicitly depend on the data, and so vary across units $i$.
3. Also, for $m = k$, update the initial fit $\widehat{Q}_i^{j,k,m}(\overline{a}^*)$ by fitting a new model, defined as $h\{Q_i^{j,k,m,*}(\overline{a}^*)\} = h\{\widehat{Q}_i^{j,k,m}(\overline{a}^*)\} + \epsilon_{j,k,m}$, where $h(\cdot)$ is an appropriate link function, $\epsilon_{j,k,m}$ is an intercept, and $Q_i^{j,k,m,*}(\overline{a}^*)$ are conditional expectations under the updated model. Note that the response variable in this model is $Q_i^{j,k,k+1}(\overline{a}^*) = Y_j$. The logit link is recommended to ensure that the estimator respects bounds implied by the data (if $Y_j$ is not bounded by (0,1), it will need to be appropriately transformed for the logit function to be defined) (van der Laan & Gruber, 2012). Estimators $\widehat{Q}_i^{j,k,m,*}(\overline{a}^*)$ for the updated fit are found by

**TABLE 1** Simulation results.

| | $n = 1000$ | | | $n = 10,000$ | | | $n = 100,000$ | | |
|---|---|---|---|---|---|---|---|---|---|
| estimator | variance[1]$/n$ | bias[2] | p[3] | variance[1]$/n$ | bias[2] | p[3] | variance[1]$/n$ | bias[2] | p[3] |
| ice_qfal | 4.3 | 1.15 | 0.68 | 4.6 | 1.10 | 0.81 | 4.5 | 1.16 | 0.76 |
| ice_true | 4.2 | −0.04 | 0.82 | 4.5 | −0.02 | 0.23 | 4.2 | 0.02 | 0.45 |
| iptw_gfal | 4.5 | 1.17 | 0.86 | 4.7 | 1.17 | 0.49 | 4.6 | 1.25 | 0.69 |
| iptw_true | 6.0 | −0.17 | 0.37 | 6.3 | −0.08 | 0.87 | 7.4 | −0.01 | 0.57 |
| tmle_bfal | 4.4 | 1.19 | 0.51 | 4.6 | 1.15 | 0.65 | 4.5 | 1.22 | 0.62 |
| tmle_gfal | 4.2 | −0.08 | 0.86 | 4.5 | −0.04 | 0.28 | 4.3 | 0.02 | 0.21 |
| tmle_qfal | 5.9 | −0.09 | 0.15 | 6.4 | −0.07 | 0.78 | 7.6 | −0.02 | 0.41 |
| tmle_true | 5.4 | −0.03 | 0.82 | 5.6 | 0.01 | 0.11 | 5.7 | 0.01 | 0.69 |

0[1]Empirical variance of estimates over 1000 simulated datasets. [2]Multiplied by 100. [3] P-value for Lilliefors test against the null hypothesis of normality.
0Abbreviations: ice=iterated conditional expectation, iptw=inverse probability of treatment weighted, tmle=targeted maximum likelihood, qfal=outcome models misspecified, gfal=treatment models misspecified, bfal=both sets of models misspecified, true=all models correctly specified.

maximizing an appropriate weighted likelihood with weights $I(\bar{A}_{im} = \bar{a}^*_m)/g_m(\bar{a}^*; \hat{\alpha}_m)$.

4. Repeat steps 2–3, estimating $Q^{j,k,m}(\bar{a}^*; \beta_m)$ and $Q^{j,k,m,*}(\bar{a}^*)$ for $m = k-1, k-2, \dots, 0$.
5. The TMLE for $\phi_{j,k}$ is then defined as $\hat{\phi}^{TMLE}_{j,k} = n^{-1} \sum_{i=1}^{n} \hat{Q}^{j,k,0,*}_i(\bar{a}^*)$.

Then, the TMLE for $\psi_t$ is defined as $\hat{\psi}^{TMLE}_t = \hat{\phi}^{TMLE}_{0,0} + \sum_{k=1}^{t}(\hat{\phi}^{TMLE}_{k,k} - \hat{\phi}^{TMLE}_{k-1,k})$. Since $\hat{\phi}^{TMLE}_{j,k}$ solves the estimating equation corresponding to the efficient influence curve (4), it will be CAN for $\phi_{j,k}$ so long as either (i) the set of outcome models $\{Q^{j,k,m}(\bar{a}^*; \beta_m) : m = 0, 1, \dots, k\}$ are correctly specified, or (ii) the set of treatment models $\{g_m(\bar{a}^*; \alpha_m) : m = 0, 1, \dots, k\}$ is correctly specified, but it is not necessary that both be correct. Therefore, if one of these two conditions holds for all $k = 0, 1, \dots, t$ and $j = k-1, k$, then $\hat{\psi}^{TMLE}_t$ will be CAN for $\psi_t$. The double robustness property carries through to $\hat{\psi}^{TMLE}_t$ by virtue of the fact that the estimating equation in (2) is unbiased if the estimating equations for all the $\phi_{j,k}$ are unbiased, which is the case for $\hat{\phi}^{TMLE}_{j,k}$ under conditions (i) or (ii) above.

## 5 | SIMULATION STUDY

A simulation study was conducted to evaluate the finite sample performance of the IPTW, ICE, and TMLE estimators described in Section 4 when Assumptions 1–3 hold and all models were correctly specified. The TMLE estimator was also evaluated under misspecification of either the treatment or outcome model. Implementation details and code for the simulation are provided in the Supporting Information.

Table 1 shows estimates of the bias, variance, and $p$-values from a Lilliefors test for normality, based on 1000 simulations each sample sizes $n$ =1000, 10,000, and 100,000 for each estimator of $\mu_5$. The results suggest that all stated theoretical properties hold approximately in simulated data. First, when all models are correctly specified, all estimators appear approximately unbiased with decreasing variance as the sample size increases. When outcome models and treatment models are misspecified, ICE and IPTW estimators appear biased, respectively. TMLE appears consistent when either the treatment or outcome models are correctly specified, but not when both are misspecified, supporting the double robustness property. Lastly, all estimators appear normally distributed for all sample sizes considered, based on Lilliefors tests.

## 6 | COVID-19 APPLICATION

### 6.1 | Data

This section presents an analysis of the motivating example, introduced in Section 2.2. Code and data are provided in the Supporting Information. State-level weekly mortality data come from the Centers for Disease Control and Prevention's National Death Index, and weekly counts of COVID-19 cases from the COVID-19 Data Repository at the Center for Systems Science and Engineering at Johns Hopkins University. Data on state-level stay-at-home orders come from the COVID-19 U.S. State Policy database. Though the outcome variable of interest is an individual-level indicator of death in week $t$, this variable is not directly observed; instead, the observed data represent counts of deaths occurring in each state. Let $s = 1, 2, \dots, 43$ be a state index, and let $Y_{ist}$ be an indicator of mortality during week $t$ for the $i$th individual ($i = 1, \dots, n_s$) living in state $s$, where $n_s$ denotes the population size in state $s$, and $n = \sum_{s=1}^{43} n_s \approx 309$ million. The observed outcome

variable is $Y_{st} = \sum_{i=1}^{n_s} Y_{ist}$, the state-level weekly sum of individual-level mortality counts, along with population counts $n_s$ (drawn from the 2010 Census). The observed treatment variable $A_{st}$ is an indicator of state $s$ being under stay-at-home order in week $t$. Finally, let $W_{st}$ be the change in confirmed COVID-19 cases reported per 100k population in the previous 4 weeks (i.e., the difference from week $t-4$ to $t$) in state $s$. Thus, in this example, the parallel trends assumption is conditional on the local state of the pandemic, which may be plausible for pandemic-related policies (Callaway & Li, 2021).

## 6.2 | Estimator implementation

### 6.2.1 | IPTW

For the treatment models, the following parametric models pooled over $k = 1, \dots, 11$ were assumed:

$$f(A_{sk}|\overline{A}_{s,k-1};\alpha_0) =$$
$$\text{Bernoulli}\{\text{logit}^{-1}(\alpha_{00} + \alpha_{01}\omega(k) + \alpha_{02}A_{s,k-1})\}$$
$$f(A_{sk}|\overline{A}_{s,k-1},\overline{W}_{sk};\alpha_1) =$$
$$\text{Bernoulli}\{\text{logit}^{-1}(\alpha_{10} + \alpha_{11}\omega(k) + \alpha_{12}A_{s,k-1} + \alpha_{13}\log W_{sk})\},$$

where $\omega(k)$ is a natural cubic spline basis with 3 degrees of freedom for time $k$. The outcome model $c_{jk}(\overline{A}) = \gamma_{0jk} + \gamma_{1jk}I(\overline{A}_k = \overline{1})$, $k = 1, \dots, 11$, $j = k, k-1$ was specified, which allows the outcome to depend on the full exposure history. The parameters $\alpha_0 = (\alpha_{00}, \alpha_{01}, \alpha_{02})$ and $\alpha_1 = (\alpha_{10}, \dots, \alpha_{13})$ were estimated using maximum likelihood, weighted by $1/n_s$ to account for differing population sizes across states. Then, $\gamma_{0jk}, \gamma_{1jk}, k = 1, \dots 11, j = k, k-1$ were estimated by maximizing the state-level binomial likelihood weighted by inverse probability of treatment weights $\pi_k(\overline{A};\overline{W},\widehat{\alpha}) = \prod_{m=1}^{k} f(A_m|\overline{A}_{m-1};\widehat{\alpha}_0)/\prod_{m=1}^{k} f(A_m|\overline{A}_{m-1},\overline{W}_m;\widehat{\alpha}_1)$, where $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ denote maximum likelihood estimators of $\alpha_0$ and $\alpha_1$. Then estimators $\widehat{\psi}_t^{IPTW}, t = 0, \dots, 11$ were calculated as $\widehat{\phi}_{0,0}^{IPTW} + \sum_{k=1}^{t}(\widehat{\phi}_{k,k}^{IPTW} - \widehat{\phi}_{k-1,k}^{IPTW})$, where $\widehat{\phi}_{j,k}^{IPTW} = \widehat{\gamma}_{0jk} + \widehat{\gamma}_{1jk}$ and $\widehat{\gamma}_{0jk}, \widehat{\gamma}_{1jk}$ denote the weighted maximum likelihood estimators.

### 6.2.2 | ICE

For ICE estimators, the following parametric outcome regression models pooled over $k = 1, \dots, 11$ were assumed:

$$Q^{j,k,m}(\overline{a}^*;\beta_m) = \text{logit}^{-1}\{\beta_{0jm} + \beta_{1jm}\omega(k) + \beta_{2jm}\omega(k)a_m^* + \beta_{3jm}\log W_m\}$$

for $j = k, k-1$ and $m = k, k-1, \dots, 0$, where again $\omega(k)$ refers to a natural cubic spline basis with 3 degrees of freedom. Note that due to the monotonic treatment pattern, the interaction between time and treatment allows the outcome to depend on the full exposure history. The parameters $\beta_m$ were estimated by maximizing a binomial quasi-likelihood, with estimators denoted as $\widehat{\beta}_m$. To account for varying state population sizes, state contributions to the quasi-likelihood were weighted by $1/n_s$. Finally, ICE estimators $\widehat{\psi}_t^{ICE}, t = 0, \dots, 11$ were calculated as $\widehat{\phi}_{0,0}^{ICE} + \sum_{k=1}^{t}(\widehat{\phi}_{k,k}^{ICE} - \widehat{\phi}_{k-1,k}^{ICE})$, where $\widehat{\phi}_{j,k}^{ICE} = \sum_{r=1}^{43} Q_r^{j,k,0}(\overline{a}^*;\widehat{\beta}_m)/43$ (as there are 43 states included in the analysis).

### 6.2.3 | TMLE

For TMLE, the same treatment models as specified for IPTW were used, along with the same outcome models as specified for ICE. Specifically, when estimating $\phi_{jk}$, for the $m$th ICE step ($m = k, k-1, \dots, 0$), the TMLE updating step was performed by maximizing another weighted quasi-binomial likelihood with response variable $Q_s^{j,k,m+1}(\overline{a}^*;\beta_{m+1})$ with an intercept and offset $Q_s^{j,k,m}(\overline{a}^*;\widehat{\beta}_m)$, weighted by $I(\overline{A}_k = \overline{1})/g_k(\overline{A},\widehat{\alpha}_k)$. Predictions $\widehat{Q}_s^{j,k,m,*}(\overline{a}^*)$ from this model were then passed to the $(m-1)$th ICE step, and the process was repeated for $m = k, k-1, \dots, 0$. Finally, $\widehat{\psi}_t^{TMLE}, t = 0, \dots, 11$ were calculated as $\widehat{\phi}_{0,0}^{TMLE} + \sum_{k=1}^{t}(\widehat{\phi}_{k,k}^{TMLE} - \widehat{\phi}_{k-1,k}^{TMLE})$, where $\widehat{\phi}_{j,k}^{TMLE} = \sum_{r=1}^{43} \widehat{Q}_r^{j,k,0,*}(\overline{a}^*)/43$.

### 6.2.4 | Bootstrap standard errors and confidence intervals

Standard errors were estimated using a nonparametric bootstrap. Specifically, for $B$ bootstrap replicates ($b = 1, 2, \dots, B$), a resampled outcome variable $Y_{st}^b = \sum_{i=1}^{n_s} Y_{ist}^b$ ($t = 0, \dots, 12$) was drawn from a multinomial distribution with $n_s$ trials and probabilities $n_s^{-1}(Y_{s0}, Y_{s1}, \dots, Y_{s,12})$, where $Y_{s,12}$ denotes the number of individuals who survived beyond $t = 11$ in state $s$. IPTW, ICE, and TMLE estimators $\widehat{\psi}_t^{IPTW,b}, \widehat{\psi}_t^{ICE,b}, \widehat{\psi}_t^{TMLE,b}$ were calculated on each replicate ($b = 1, \dots, B$). Then, Wald 95% confidence intervals were computed using the standard deviation of bootstrap estimates.

## 6.3 | Results

Figure 2 shows results in the form of estimated U.S. weekly mortality rates per 100,000 person weeks over the study

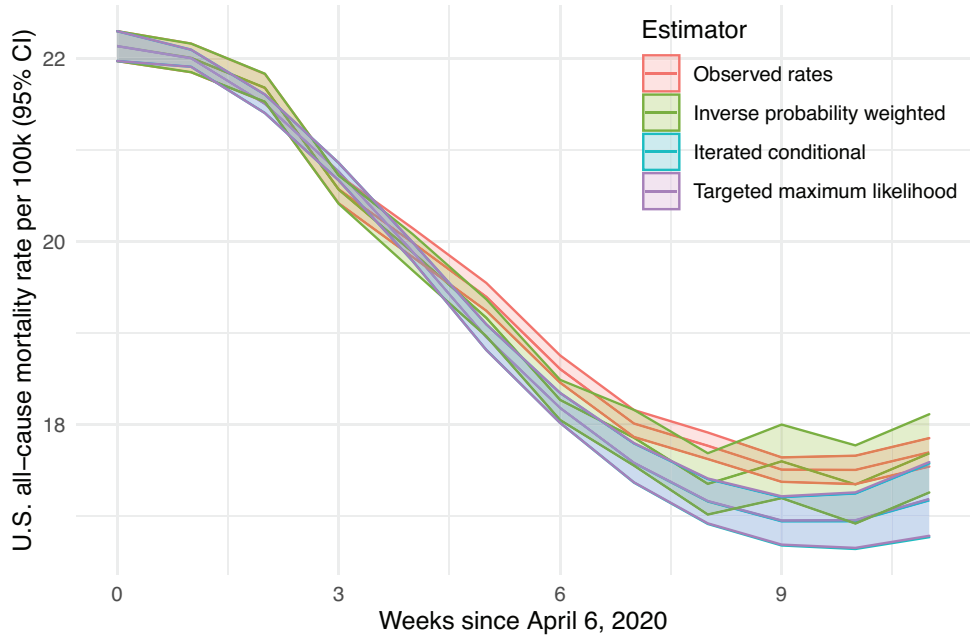


**FIGURE 2** Estimated U.S. weekly mortality rates—observed (red) and estimated under hypothetical treatment setting all states to remain under stay-at-home order using IPTW (green), ICE (blue), and TMLE (purple). Note that TMLE and ICE estimates and 95% CIs are nearly identical. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

period under the natural course (red) and under the hypothetical sustained treatment of setting $A_t = 1$ for all $t$, that is, under a scenario where all 43 included states maintained stay-at-home orders through June 2020. This figure appears in color in the electronic version of this article, and any mention of color refers to that version. The three estimators largely agree in their predictions that all-cause mortality rates would have been moderately lower throughout most of the study period, had stay-at-home orders remained in place. Translating the counterfactual mortality rate estimates to lives saved, if all causal and modeling assumptions hold, based on TMLE, stay-at-home orders remaining in place from April through June 2020 would have saved appoximately 11,100 (95% CI: 6,800, 15,500) lives in those 43 states during the same time period. Results based on ICE were similar (point estimate: 11,300, 95% CI: 6900, 15,600), whereas IPTW gave a smaller point estimate and somewhat wider CI (point estimate: 4100, 95% CI: −500, 8700).

## 7 | EXTENSIONS

### 7.1 | Violations of parallel trends

In some applications, the parallel trends assumption (Assumption 3) may be questionable, and investigators may be interested in how inferences are altered by plausible deviations from parallel trends. A sensitivity analysis can be conducted as follows. Let

$$\begin{aligned}\Delta(\overline{w}_k, t) &= \mathbb{E}\{Y_t(\overline{a}^*) - Y_{t-1}(\overline{a}^*) | \overline{W}_k = \overline{w}_k, \overline{A}_{k-1} = \overline{a}^*_{k-1}\} \\ &\quad - \mathbb{E}\{Y_t(\overline{a}^*) - Y_{t-1}(\overline{a}^*) | \overline{W}_k = \overline{w}_k, \overline{A}_k = \overline{a}^*_k\},\end{aligned}$$

where $\Delta(\overline{w}_k, t)$ quantifies a deviation from parallel trends, which may depend on both the covariates $\overline{w}_k$ and time $t$. Then, consider the following statistical parameter:

$$\begin{aligned}\overline{\Delta}_k &= \sum_{m=0}^{k-1} \int \mathbb{E}\{\Delta(\overline{W}_{m+1}, k) | \overline{A}_m = \overline{a}^*_m, \overline{W}_m = \overline{w}_m\} \\ &\quad \times \prod_{s=0}^{m} dF(w_s | \overline{w}_{s-1}, \overline{a}^*_{s-1}).\end{aligned}$$

If Assumptions 1 and 2 hold, then

$$\mu_t = \psi_t^{\Delta} \equiv \psi_t + \sum_{k=1}^{t} \overline{\Delta}_k. \tag{5}$$

A proof of (5) is given in Web Appendix D. If a particular value is assumed known for $\Delta(\overline{w}_k, t)$, then estimation can proceed by noting that, like $\psi_t$, $\overline{\Delta}_k$ is also a linear combination of g-formulas, where in this case, the outcome variable is $\Delta(\overline{W}_{m+1}, k)$. Thus, one can form estimators $\widehat{\Delta}_k$ of $\overline{\Delta}_k$ using IPTW, ICE, or TMLE, and define the estimator $\widehat{\psi}_t^{\Delta} = \widehat{\psi}_t + \sum_{k=1}^{t} \widehat{\Delta}_k$, where $\widehat{\psi}_t$ is one of the estimators of $\psi_t$

described in Section 4. By the same arguments as in Section 4, $\widehat{\psi}_t^{\Delta}$ will be CAN for $\psi_t$ under Assumptions 1 and 2, and correct model specification (if using TMLE to estimate $\overline{\Delta}_k$, either the treatment or outcomes models can be misspecified, but not both). In practice, $\Delta(\overline{w}_k, t)$ will typically not be known, and thus, estimates may be computed over a range of plausible values of $\Delta(\overline{w}_k, t)$. Differences in trends between subgroups of units before discontinuation occurs may be helpful in determining plausible values of $\Delta(\overline{w}_k, t)$ (Roth & Rambachan, 2019).

## 7.2 | Dynamic regimes

In addition to the static regimes considered above, the proposed approach can accommodate regimes where treatment decisions may depend on the history of covariates and/or treatments. Let $\overline{g} = \{g_0(w_0), g_1(\overline{w}_1), \dots, g_\tau(\overline{w}_\tau)\}$ denote a dynamic regime, where $g_k(\overline{w}_k)$ returns the treatment value $a_k$ that would be assigned given covariate history $\overline{w}_k$. Note that $g_k(\cdot)$ may also depend on treatment history, which we suppress for notational simplicity. Likewise, let $Y_k(\overline{g})$ be a potential outcome under treatment regime $\overline{g}$. Suppose interest is in the estimand $\mu_t^g = \mathbb{E}\{Y_t(\overline{g})\}$. Then, consider the following modifications to Assumptions 1–3.

**Assumption 4.** (SUTVA for dynamic regimes): If $\overline{A}_{it} = \overline{g}_t(\overline{W}_{it})$, then $Y_{it} = Y_{it}(\overline{g}_t)$ for $t \in \{0, 1, \dots \tau\}$.

**Assumption 5.** (Positivity for dynamic regimes): If $f\{\overline{w}_t|\overline{A}_{t-1} = \overline{g}_{t-1}(\overline{W}_{t-1})\} > 0$, then $f\{g_t(\overline{w}_t)|\overline{W}_t = \overline{w}_t, \overline{A}_{t-1} = \overline{g}_{t-1}(\overline{w}_{t-1})\} > 0$, for $\overline{w}_t \in \overline{\mathcal{W}}_t; t \in \{1, 2, \dots, \tau\}$.

**Assumption 6.** (Parallel trends for dynamic regimes): For $t \in \{1, 2, \dots, \tau\}, k \leq t$ :

$$\begin{aligned}\mathbb{E}\{Y_t(\overline{g}) \; &- Y_{t-1}(\overline{g})|\overline{W}_k, \overline{A}_{k-1} = \overline{g}_{k-1}(\overline{W}_{k-1})\} \\ &= \mathbb{E}\{Y_t(\overline{g}) - Y_{t-1}(\overline{g})|\overline{W}_k, \overline{A}_k = \overline{g}_k(\overline{W}_k)\}.\end{aligned}$$

**Lemma 2.** *(Parallel trends g-formula, dynamic regimes) Define the functional (i.e., statistical parameter)*

$$\begin{aligned}\psi_t^g \equiv \mathbb{E}(Y_0) \; &+ \sum_{k=1}^{t} \int \mathbb{E}\{Y_k - Y_{k-1}|\overline{W}_k = \overline{w}_k, \\ \overline{A}_k \; &= \overline{g}_k(\overline{w}_k)\} \prod_{m=0}^{k} dF\{w_m|\overline{w}_{m-1}, \overline{g}_{m-1}(\overline{w}_{m-1})\}.\end{aligned}$$

*Under a staggered discontinuation design and if Assumptions 4–6 hold, then $\psi_t^g = \mu_t^g$.*

The proof of Lemma 2 follows from results in Web Appendix A. Thus, the IPTW, ICE, and TMLE estimators described can be used, with $\phi_{j,k}$ appropriately redefined.

# 8 | DISCUSSION

This paper considers a new approach to identifying effects of sustained intervention strategies based on an assumption set that includes parallel trends. This assumption is popular in DIDs because it allows for some degree of unmeasured confounding (Zeldow & Hatfield, 2021). Recently, parallel trends assumptions have been leveraged to target sustained treatment estimands, mainly considering certain types of treatment regimes (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfoeuille, 2020, 2021b, 2021a). Relative to previous work, the main contribution of this paper is a framework for estimating marginal intervention-specific means for general treatment regimes (including dynamic regimes) under parallel trends, thus connecting disparate causal inference literatures from biostatistics (Bang & Robins, 2005; Robins, 1986, 2000; van der Laan & Gruber, 2012) and econometrics (Ashenfelter & Card, 1985; Callaway & Sant'Anna, 2021).

The proposed methods bear a close relationship to recently developed econometric methods for DID. Specifically, existing estimators for time-varying ATT parameters in event study and staggered adoption designs could be adapted to target the estimands and more general regimes considered here. Given any regime $\overline{a}^*$, one could relabel individuals consistent with $\overline{a}^*$ as "untreated" and those who deviated as "treated," estimate group-time ATTs considered by Callaway and Sant'Anna (2021), and then take an appropriate linear combination to arrive at an estimate of $\mathbb{E}[Y_t(\overline{a}^*)]$. In the absence of time-varying covariates, the estimators of Callaway and Sant'Anna (2021) could be adapted in this way. Likewise, the estimators proposed here could be adapted to target group-time ATTs accounting for time-varying covariates. Thus, relative to existing econometric methods, the contributions of the present work are to highlight that existing DID approaches are perhaps more general than commonly believed, propose new estimators tailored to this generality, allow for flexible adjustment for time-varying covariates, and draw connections with the literature on time-varying treatments.

Regarding the example presented in Section 6, care should be taken when assuming parallel trends for pandemic-related outcomes without conditioning on pandemic state variables such as infection rates, as marginal parallel trends are incompatible with standard epidemic models (Callaway & Li, 2021). DID methods have been used to estimate effects of stay-at-home orders on the treated (e.g., Fowler et al., 2021). The methods in this

paper allow for (i) a different target parameter that may more directly correspond to decisions facing policy makers and public health officials (Maldonado & Greenland, 2002), and (ii) adjustment for time-varying pandemic state variables likely affected by prior treatment, which DID methods have only recently begun to consider (Callaway & Li, 2021). That said, assessing effects of stay-at-home orders is complex, and a comprehensive analysis should consider potential biases not factored into the present analysis; for example, there is likely some interference (Haber et al., 2021). Thus, the application results are not meant to inform policy or scientific conclusions.

The approach presented here may have application in many other contexts. Many U.S. state-level policies have changed in such a way as to accommodate a staggered discontinuation design, including in domains other than pandemic mitigation. Outside of staggered discontinuation designs, the proposed methods apply more generally in settings where baseline potential outcomes are identified. For example, the approach could be used to estimate per-protocol effects in a clinical trial of a time-varying treatment regime with nonadherence.

Several areas for future research remain. First, it will be important to explore efficiency for competing estimators in this framework. Notably, the TMLE presented here is only known to be semiparametric efficient for the nuisance parameters $\phi_{jk}$ and not necessarily for the target parameter $\psi_t$ (van der Laan & Gruber, 2012). Second, there is a notable similarity between the proposed parallel trends assumption and the discrete-time independent censoring assumption from linear increments methods for missing data (Diggle et al., 2007); this connection may be informative for suggesting new estimators, particularly in the case of nonmonotonic treatments. Finally, though parallel trends may be considered more plausible than sequential exchangeability in some settings, strategies for evaluating the plausibility of the assumption (including any implications regarding effect heterogeneity) using domain knowledge are needed (e.g., Ghanem et al., 2022). While the parallel trends assumption in this paper avoid restricting effect heterogeneity by considering only one regime, structural models that allow parallel trends for one regime but not others may be difficult to justify in practice (Shahn et al., 2022). Thus, an important focus of future research is to elucidate under what structural conditions parallel trends are plausible and for what regime(s), and to develop diagnostics for evaluating these conditions in practice.

## ACKNOWLEDGMENTS

The authors thank the Associate Editor, two reviewers, and Dr. Whitney Robinson for helpful comments. This research was supported by the NIH grants T32-HD091058-02, T32-AI007001, R01 AI085073, and P2C-HD050924. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

## DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available as part of the R package `didgformula`, available at https://github.com/audreyrenson/didgformula, in the dataset called `stayathome2020`.

## OPEN RESEARCH BADGES

This article has earned Open Data and Open Materials badges. Data and materials are available at http://re3data.org/

## ORCID

*Audrey Renson* https://orcid.org/0000-0003-1603-2587
*Michael G. Hudgens* https://orcid.org/0000-0002-9106-4194
*Alexander P. Keil* https://orcid.org/0000-0002-0955-6107

## REFERENCES

Ashenfelter, O. & Card, D. (1985) Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67, 648–660.

Bang, H. & Robins, J.M. (2005) Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61, 962–973.

Callaway, B. & Li, T. (2021) Policy evaluation during a pandemic. *arXiv preprint arXiv:2105.06927*, 1–37.

Callaway, B., Goodman-Bacon, A. & Sant'Anna, P.H. (2021) Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*, 1–74.

Callaway, B. & Sant'Anna, P.H. (2021) Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225, 200–230.

de Chaisemartin, C. & D'Haultfoeuille, X. (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 11, 2964–2996.

de Chaisemartin, C. & D'Haultfoeuille, X. (2021a) Difference-in-differences estimators of intertemporal treatment effects. *arXiv preprint arXiv:2007.04267*, 1–64.

de Chaisemartin, C. & D'Haultfoeuille, X. (2021b). Two-way fixed effects regressions with several treatments. *arXiv preprint arXiv:2012.10077*, 1–34.

Diggle, P., Farewell, D. & Henderson, R. (2007) Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56, 499–550.

Fowler, J., Hill, S., Levin, R. & Obradovich, N. (2021) Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States. *PLoS One*, 16, e02488 49.

Ghanem, D., Sant'Anna, P. & Wüthrich, K. (2022) Selection and parallel trends. *arXiv Preprint arXiv:2203.09001*.

Goodman-Bacon, A. (2021) Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225, 254–277.

Haber, N.A., Clarke-Deelder, E., Salomon, J.A., Feller, A. & Stuart, E.A. (2021) Impact evaluation of coronavirus disease 2019 policy: a guide to common design issues. *American Journal of Epidemiology*, 190, 2474–2486.

Halloran, M.E. & Hudgens, M.G. (2016) Dependent happenings: a recent methodological review. *Current Epidemiology Reports*, 3, 297–305.

Maldonado, G. & Greenland, S. (2002) Estimating causal effects. *International Journal of Epidemiology*, 31, 422–429.

Marcus, M. & Sant'Anna, P.H. (2021) The role of parallel trends in event study settings: an application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, 8, 235–275.

Rambachan, A. & Roth, J. (2022) *A more credible approach to parallel trends*. Working Paper 1–47

Robins, J.M. (1986) 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, 1393–1512.

Robins, J.M. (1989) The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health Service Research Methodology: A Focus on AIDS*, 113–159.

Robins, J.M. (2000) Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, M.E. & Berry, D.A. (Eds.) *Statistical models in epidemiology, the environment, and clinical trials*. New York: Springer, pp. 95–133.

Robins, J.M., Hernán, M.Á. & Brumback, B. (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11, 550–560.

Robins, J.M., Mark, S.D. & Newey, W.K. (1992) Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48, 479–495.

Roth, J. (2019) *Pre-test with caution: event-study estimates after testing for parallel trends*. Working Paper 1–54.

Roth, J., Sant'Anna, P.H., Bilinski, A. & Poe, J. (2022) What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *arXiv preprint arXiv:2201.01194* .

Sant'Anna, P.H.C. & Zhao, J. (2022) Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219, 101–122.

Shahn, Z., Dukes, O., Richardson, D., Tchetgen Tchetgen, E. & Robins, J. (2022) Structural nested mean models under parallel trends assumptions. *arXiv preprint arXiv:2204.10291* 1–42.

Stefanski, L.A. & Boos, D.D. (2002) The calculus of M-estimation. *The American Statistician*, 56, 29–38.

Tran, L., Yiannoutsos, C., Wools-Kaloustian, K., Siika, A., van der Laan, M. & Petersen, M. (2019) Double robust efficient estimators of longitudinal treatment effects: comparative performance in simulations and a case study. *International Journal of Biostatistics*, 15, 1–27.

van der Laan, M.J. & Gruber, S. (2012) Targeted minimum loss based estimation of causal effects of multiple time point interventions. *International Journal of Biostatistics*, 8, 1–39.

Zeldow, B. & Hatfield, L.A. (2021) Confounding and regression adjustment in difference-in-differences studies. *Health Services Research*, 56, 932–941.

## SUPPORTING INFORMATION

Web appendices referenced in Sections 3, 5, and 7, code for the simulation study in Section 5, and data and code for the application in Section 6, are available with this paper at the Biometrics website on Wiley Online Library. The R package `didgformula|`, available at https://github.com/audreyrenson/didgformula , implements the estimators, as well as the simulation study (in vignette "simulation"), and applications results (in vignette "example").

Data S1

**How to cite this article:** Renson, A., Hudgens, M.G., Keil, A.P., Zivich, P.N. & Aiello, A.E. (2023) Identifying and estimating effects of sustained interventions under parallel trends assumptions. *Biometrics*, 79, 2998–3009.
https://doi.org/10.1111/biom.13862