

# NEW $G$ -FORMULA FOR THE SEQUENTIAL CAUSAL EFFECT AND BLIP EFFECT OF TREATMENT IN SEQUENTIAL CAUSAL INFERENCE

BY XIAOQIN WANG<sup>1</sup> AND LI YIN<sup>2</sup>

<sup>1</sup>*Department of Electrical Engineering, Mathematics and Science, University of Gävle, [xwg@hig.se](mailto:xwg@hig.se)*

<sup>2</sup>*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, [li.yin@ki.se](mailto:li.yin@ki.se)*

In sequential causal inference, two types of causal effects are of practical interest, namely, the causal effect of the treatment regime (called the sequential causal effect) and the blip effect of treatment on the potential outcome after the last treatment. The well-known  $G$ -formula expresses these causal effects in terms of the standard parameters. In this article, we obtain a new  $G$ -formula that expresses these causal effects in terms of the point observable effects of treatments similar to treatment in the framework of single-point causal inference. Based on the new  $G$ -formula, we estimate these causal effects by maximum likelihood via point observable effects with methods extended from single-point causal inference. We are able to increase precision of the estimation without introducing biases by an unsaturated model imposing constraints on the point observable effects. We are also able to reduce the number of point observable effects in the estimation by treatment assignment conditions.

**1. Introduction.** In many economic and medical practices, treatments are assigned or observed in the form of a sequence to influence an observable outcome of interest, which occurs after the last treatment of the sequence in a certain population. In addition to stationary observable covariates, there may often be time-dependent observable covariates between treatments. From the observed data, the following causal effects of practical interest can be estimated: the sequential causal effect and the blip effect of treatment. The sequential causal effect is the causal effect of the treatment regime on the potential outcome, and the blip effect is the causal effect of one treatment on the potential outcome given the history of previous treatments and potential covariates while setting the subsequent treatments at controls (Robins [5])

Under the identifying condition, Robins [4, 5, 7] derived the well-known  $G$ -formula, which expresses the sequential causal effect and the blip effect in terms of the standard parameters, that is, the conditional means of the observable outcome given all treatments and observable covariates. In the expression for the standard parameters, the treatment involves not only the history of previous treatments and observable covariates as possible confounders but also the subsequent treatments and observable covariates as posttreatment variables; therefore, the standard parameters are essentially all different and are all needed in the  $G$ -formula (see the literature on the influence of the posttreatment variables, e.g., Robins [5]). Consequently, based on the  $G$ -formula, the maximum-likelihood (ML) estimation of these causal effects via standard parameters suffers from the curse of dimensionality and the null paradox (Robins [4, 5, 7]). The curse of dimensionality implies that if a treatment sequence is long and/or the number of observable covariates is large, a huge number of standard parameters are needed in the ML estimation. The null paradox implies that an unsaturated model imposing equalities between standard parameters is misspecified. As a result, it is highly difficult to

---

Received January 2018; revised December 2018.

*MSC2010 subject classifications.* Primary 62H12; secondary 62H15, 62F03, 62F30.

*Key words and phrases.* Blip effect, curse of dimensionality, new  $G$ -formula, null paradox, point observable effect, sequential causal effect.

improve the estimation in the usual framework of likelihood inference, where precision of the estimation is increased by an unsaturated model and the number of model parameters is reduced in the estimation by treatment assignment conditions. Without a constraint on standard parameters, the ML estimates may not be consistent if treatment sequences are long and/or the number of observable covariates is large.

If the treatment sequence is short and the observable covariates are few, noticeably, the aforementioned likelihood-based parametric method can be used to estimate the sequential causal effect and the blip effect, although such methods are still subject to the null paradox (Taubman et al. [11]). For more complicated cases, semiparametric methods are instead used, such as the marginal structural model based on inverse probability weighting (Robins [6, 7]) and the  $G$ -estimation based on the structural nested mean model (Robins [5, 7]) and their extensions, which are not based on the usual likelihood of the observed data, and thus yield nonlikelihood-based estimates of these causal effects. Although these nonlikelihood-based estimates are consistent, the likelihood-based estimates remain attractive because they reflect the data-generating mechanism and are the most efficient.

In the framework of single-point causal inference (Rosenbaum and Rubin [8]), every treatment in the treatment sequence has the point causal effect of treatment, which is defined as the causal effect of the treatment on the final potential outcome given a history of previous treatments and potential covariates. Under the assumption of strongly ignorable treatment assignments, namely, the identifying condition, the point causal effect of treatment is equal to the point observable effect of treatment, which is defined as the effect of the treatment on the observable outcome only given a history of previous treatments and observable covariates. In the expression for the point observable effects, the treatment involves only a history of the previous treatments and observable covariates as possible confounders but does not involve the subsequent treatments and observable covariates, unlike the treatment in the expression for the standard parameters. As a result, the point observable effect can be estimated by ML without knowing the influences of the subsequent treatments and observable covariates. Furthermore, the number of point observable effects can be reduced in the ML estimation by treatment assignment conditions, such as the randomization in randomized trials or the subclassification in observational studies (Rosenbaum and Rubin [8]). The precision of the estimation can also be improved by an unsaturated model imposing constraints on the point observable effects. As a result, the ML estimate of the point observable effect can remain consistent even for a long treatment sequence and/or many observable covariates.

In an attempt to extend the methodology from single-point causal inference to sequential causal inference, we derive the new  $G$ -formula, which expresses the sequential causal effect and the blip effect in terms of the point observable effects instead of the standard parameters. In Section 2, we express the sequential causal effect, the blip effect and the point causal effect in terms of potential covariates and outcome. In Section 3, we derive the relationship between these three causal effects. In Section 4, we use the obtained relationship to derive the new  $G$ -formula under the identifying condition. In Section 5, we apply the new  $G$ -formula to estimate the sequential causal effect and the blip effect via the point observable effects by ML and compare our method with other methods in the literature. In Section 6, we conclude the article with discussions on further applications of the new  $G$ -formula.

## 2. Causal effects arising from treatment regimes.

2.1. *Notation for the treatment regime, potential covariates and potential outcome.* Let  $G_t$  be the treatment variable representing a certain rule under which treatments  $z_t$  could potentially be assigned at time  $t = 1, \dots, T$ . Suppose that every treatment could be applied to each unit of the population. Assume that there is no interference between units and no

representative treatment for any unit. For notational simplicity, we use one subpopulation defined by stationary covariates of the population as our population, and henceforth do not consider stationary covariates in the following development.

A treatment *regime* is a sequence of such treatment variables under which treatments  $z_t$  could potentially and consecutively be assigned at times  $t = 1, \dots, T$ , and it is denoted by  $\mathbf{G}_1^T = (G_1, \dots, G_T)$ . A *subregime* from time  $t$  to  $s$  is denoted by  $\mathbf{G}_t^s = (G_t, \dots, G_s)$ . Under regime  $\mathbf{G}_1^T$ , each unit could have a potential (time-dependent) covariate vector  $\mathbf{X}_{t-1}(\mathbf{G}_1^{t-1})$  between  $G_{t-1}$  and  $G_t$  ( $t > 1$ ) and a potential outcome  $Y(\mathbf{G}_1^T)$  of interest after the last treatment variable  $G_T$ . Let  $\mathbf{X}_1^{t-1}(\mathbf{G}_1^{t-1}) = \{\mathbf{X}_1(G_1), \mathbf{X}_2(\mathbf{G}_1^2), \dots, \mathbf{X}_{t-1}(\mathbf{G}_1^{t-1})\}$  be an array of the potential covariates before  $G_t$ . Then the set  $\{\mathbf{G}_1^T, \mathbf{X}_1^{T-1}(\mathbf{G}_1^{T-1}), Y(\mathbf{G}_1^T)\} = \{G_1, \mathbf{X}_1(G_1), \dots, G_{T-1}, \mathbf{X}_{T-1}(\mathbf{G}_1^{T-1}), G_T, Y(\mathbf{G}_1^T)\}$  represents a summary of all treatment and potential variables in the temporal order. These variables have the realizations  $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y) = (z_1, \mathbf{x}_1, \dots, z_{T-1}, \mathbf{x}_{T-1}, z_T, y)$ .

Under *random regime*  $\mathbf{G}_1^T = \mathbf{R}_1^T$ , treatment  $z_t$  is randomly assigned according to the conditional distribution  $\mathbf{P}^{R_t}\{z_t | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}$  of the random treatment variable  $R_t$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ . Let  $\mathbf{P}^{R_1}(z_1) = \mathbf{P}^{R_1}(z_1 | \mathbf{z}_1^{1-1}, \mathbf{x}_1^{1-1})$  at  $t = 1$ . Throughout, we use  $\mathbf{P}(\cdot)$  to denote the probability distribution of discrete potential variables or the density distribution of continuous potential variables and a superscript on  $\mathbf{P}(\cdot)$  to indicate the regime under which these variables are generated.

Under *deterministic regime*  $\mathbf{G}_1^T = \mathbf{D}_1^T$ , treatment  $z_t$  is deterministically assigned according to the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , that is,  $\mathbf{P}^{D_t}\{z_t | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} = 1$  if  $z_t = z(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  and zero otherwise. If each  $z_t$  does not depend on  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , then  $\mathbf{D}_1^T$  is a static regime; otherwise, it is a dynamic regime. Noticeably, treatment  $z_t$  in a dynamic regime often depends only on part of  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ . For *mixed regimes*, the treatments can be assigned deterministically or randomly; for instance,  $\mathbf{G}_1^T = (D_1, \mathbf{R}_2^T)$ , where the first treatment  $z_1$  is deterministically assigned and the subsequent treatments  $\mathbf{z}_2^T$  are randomly assigned.

Under regime  $\mathbf{G}_1^T$ , the treatment and potential variables  $\{\mathbf{G}_1^T, \mathbf{X}_1^{T-1}(\mathbf{G}_1^{T-1}), Y(\mathbf{G}_1^T)\}$  follow the stochastic process described by

$$\begin{aligned} \mathbf{P}^{\mathbf{G}_1^T}(z_1^T, \mathbf{x}_1^{T-1}, y) &= \mathbf{P}^{\mathbf{G}_1}(z_1) \mathbf{P}^{\mathbf{G}_1}(\mathbf{x}_1 | z_1) \cdots \mathbf{P}^{\mathbf{G}_1^T}(z_T | \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) \\ &\quad \times \mathbf{P}^{\mathbf{G}_1^T}(y | \mathbf{z}_1^T, \mathbf{x}_1^{T-1}). \end{aligned}$$

For deterministic treatment variable  $G_t = D_t$ , we still keep  $z_t$  in the distribution despite a slight abuse of notation. From this formula, we have the following remark, which will be applied throughout the article.

REMARK. Under subregime  $\mathbf{G}_t^T$  given the same  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  of different  $\mathbf{G}_1^{t-1}$ , the conditional distributions  $\mathbf{P}^{\mathbf{G}_t^T}(z_t^T, \mathbf{x}_t^{T-1}, y | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  are identical; therefore we will not specify subregime  $\mathbf{G}_1^{t-1}$  for  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  in the distribution, and hence denote the distributions by  $\mathbf{P}^{\mathbf{G}_t^T}(z_t^T, \mathbf{x}_t^{T-1}, y | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ . Specifically, we use  $\mathbf{P}^{\mathbf{G}_t^T}(y | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  and  $\mathbf{P}^{\mathbf{G}_t^s}(z_t^{s-1}, \mathbf{x}_t^{s-1}, z_s | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ .

2.2. *Sequential causal effects, blip effects and point causal effects.* In this article, we study additive causal effects, although we believe that our method can be extended to nonadditive causal effects. Consider the mean  $E\{Y(\mathbf{G}_1^T)\}$  with respect to the distribution  $\mathbf{P}^{\mathbf{G}_1^T}(y)$  of the potential outcome  $Y(\mathbf{G}_1^T)$ . The *sequential causal effect* of deterministic regimes  $\mathbf{G}_1^T = \mathbf{A}_1^T$  relative to  $\mathbf{G}_1^T = \mathbf{B}_1^T$  is according to Robins [4, 5, 7]

$$(1) \quad \text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) = E\{Y(\mathbf{A}_1^T)\} - E\{Y(\mathbf{B}_1^T)\}.$$

We can extend the sequential causal effect from deterministic regimes to random regimes but refrain from doing so for the sake of notational simplicity.

Consider the conditional mean  $E\{Y(\mathbf{G}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}$  with respect to the conditional distribution  $\mathbf{P}^{\mathbf{G}_t^T}(y \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  of  $Y(\mathbf{G}_t^T)$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ . Without a loss of generality, we take  $z_t = 0$  as the control treatment and  $z_t > 0$  as the active treatment. For two static subregimes,  $\mathbf{G}_t^T = (D_t = z_t, \mathbf{D}_{t+1}^T = \mathbf{0})$  and  $\mathbf{G}_t^T = (D_t = 0, \mathbf{D}_{t+1}^T = \mathbf{0})$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , the *blip effect* of treatment  $z_t > 0$  is according to Robins [4, 5, 7]

$$\begin{aligned} & \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &= E\{Y(D_t = z_t, \mathbf{D}_{t+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ (2) \quad & - E\{Y(D_t = 0, \mathbf{D}_{t+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

which at  $t = 1$  is  $\phi(z_1) = E\{Y(D_1 = z_1, \mathbf{D}_2^T = \mathbf{0})\} - E\{Y(\mathbf{D}_1^T = \mathbf{0})\}$ .

For two mixed regimes,  $\mathbf{G}_t^T = (D_t = z_t, \mathbf{R}_{t+1}^T)$  and  $\mathbf{G}_t^T = (D_t = 0, \mathbf{R}_{t+1}^T)$  given  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  where  $\mathbf{R}_{t+1}^T$  is dependent on  $D_t$  and thus may be different in the two regimes, the *point causal effect* of treatment  $z_t > 0$  is

$$\begin{aligned} & \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ (3) \quad &= E\{Y(D_t = z_t, \mathbf{R}_{t+1}^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} - E\{Y(D_t = 0, \mathbf{R}_{t+1}^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

which at  $t = 1$  is  $\theta(z_1) = E\{Y(D_1 = z_1, \mathbf{R}_2^T)\} - E\{Y(D_1 = 0, \mathbf{R}_2^T)\}$ .

**3. Relationship between causal effects.** In the [Appendix](#), we prove the following.

LEMMA 1. *The conditional mean of the potential outcome  $Y(\mathbf{G}_t^T)$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  satisfies*

$$\begin{aligned} & E\{Y(\mathbf{G}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ (4) \quad &= E\{Y(\mathbf{D}_t^T = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} + \sum_{s=t}^T E\{\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

where the expectation in the summation is with respect to  $\mathbf{P}^{\mathbf{G}_t^S}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  and  $\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s)$  is a shorthand for  $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)$  in the context of the conditional expectation. We will use the shorthand throughout the article.

When  $t = 1$ , formula (4) becomes

$$E\{Y(\mathbf{G}_1^T)\} = E\{Y(\mathbf{D}_1^T = \mathbf{0})\} + \sum_{s=1}^T E\{\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}.$$

Applying this expression to the deterministic regimes  $\mathbf{G}_1^T = \mathbf{A}_1^T$  and  $\mathbf{G}_1^T = \mathbf{B}_1^T$  and inserting the obtained equalities into formula (1), we obtain the following.

PROPOSITION 1. *In terms of the blip effects of treatments, the sequential causal effect is expressed by*

$$(5) \quad \text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) = \sum_{s=1}^T E_1\{\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} - \sum_{s=1}^T E_2\{\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\},$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{A}_1^S}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$  and  $E_2(\cdot)$  to  $\mathbf{P}^{\mathbf{B}_1^S}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ .

By applying formula (4) to mixed subregimes  $\mathbf{G}_t^T = (D_t = z_t, \mathbf{R}_{t+1}^T)$  and  $\mathbf{G}_t^T = (D_t = 0, \mathbf{R}_{t+1}^T)$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , inserting the obtained equalities into (3) and using the equality  $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t = 0) = 0$ , we obtain the following.

PROPOSITION 2. *In terms of the blip effects of treatments, the point causal effect is expressed by*

$$(6) \quad \begin{aligned} \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) &= \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &+ \sum_{s=t+1}^T E_1\{\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ &- \sum_{s=t+1}^T E_2\{\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  for  $\mathbf{G}_t^T = (D_t = z_t, \mathbf{R}_{t+1}^T)$  and  $E_2(\cdot)$  for  $\mathbf{G}_t^T = (D_t = 0, \mathbf{R}_{t+1}^T)$ .

In the Appendix, we also prove the following.

LEMMA 2. *For deterministic subregime  $\mathbf{D}_t^T$  and mixed subregime  $(D_t, \mathbf{R}_{t+1}^T)$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , the conditional mean of the potential outcome  $Y(\mathbf{D}_t^T)$  given  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  satisfies*

$$(7) \quad \begin{aligned} E\{Y(\mathbf{D}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} &= E\{Y(D_t = 0, \mathbf{R}_{t+1}^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} + \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &+ \sum_{s=t+1}^T E_1\{\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ &- \sum_{s=t+1}^T E_2\{\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  for  $\mathbf{G}_t^s = \mathbf{D}_t^s$  and  $E_2(\cdot)$  for  $\mathbf{G}_t^s = (\mathbf{D}_t^{s-1}, R_s)$ ; noticeably,  $\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s)$  is a shorthand for  $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_s)$  in the context of the conditional expectation like  $\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s)$  in Lemma 1.

When  $t = 1$ , formula (7) becomes

$$\begin{aligned} E\{Y(\mathbf{D}_1^T)\} &= E\{Y(D_1 = 0, \mathbf{R}_2^T)\} + \theta(z_1) \\ &+ \sum_{s=2}^T E_1\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} - \sum_{s=2}^T E_2\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}. \end{aligned}$$

Applying this expression to deterministic regimes  $\mathbf{D}_1^T = \mathbf{A}_1^T$  and  $\mathbf{D}_1^T = \mathbf{B}_1^T$  and then inserting the obtained equalities into formula (1), we obtain the following.

PROPOSITION 3. *Let  $a_1$  be the treatment assigned by  $A_1$  and  $b_1$  by  $B_1$ . Then in terms of the point causal effects of treatments, the sequential causal effect is expressed by*

$$\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) = \theta(a_1) - \theta(b_1)$$

$$(8) \quad \begin{aligned} & + \sum_{s=2}^T E_1\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} - \sum_{s=2}^T E_2\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} \\ & - \sum_{s=2}^T E_3\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} + \sum_{s=2}^T E_4\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$  for  $\mathbf{G}_1^s = \mathbf{A}_1^s$ ,  $E_2(\cdot)$  for  $\mathbf{G}_1^s = (\mathbf{A}_1^{s-1}, R_s)$ ,  $E_3(\cdot)$  for  $\mathbf{G}_1^s = \mathbf{B}_1^s$  and  $E_4(\cdot)$  for  $\mathbf{G}_1^s = (\mathbf{B}_1^{s-1}, R_s)$ .

Applying formula (7) to static subregimes  $\mathbf{D}_t^T = (D_t = z_t, \mathbf{D}_{t+1}^T = \mathbf{0})$  and  $\mathbf{D}_t^T = (D_t = 0, \mathbf{D}_{t+1}^T = \mathbf{0})$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , inserting the obtained equalities into formula (2) and using the equality  $\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s = 0) = 0$ , we obtain the following.

PROPOSITION 4. *In terms of the point causal effects, the blip effect is expressed by*

$$(9) \quad \begin{aligned} \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) & = \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ & - \sum_{s=t+1}^T E_1\{\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ & + \sum_{s=t+1}^T E_2\{\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  for  $\mathbf{G}_t^s = (D_t = z_t, \mathbf{D}_{t+1}^{s-1} = \mathbf{0}, R_s)$  and  $E_2(\cdot)$  for  $\mathbf{G}_t^s = (\mathbf{D}_t^{s-1} = \mathbf{0}, R_s)$ .

From Propositions 1 to 4, the following statements can be made. First, according to Propositions 2 and 4, the blip effects of treatments are one-to-one linear transformations of the point causal effects of treatments. Second, according to Propositions 1 and 3, all sequential causal effects can be expressed in terms of the blip effects or the point causal effects. Finally, according to Propositions 1–4, these relationships between causal effects do not involve the effects of potential covariates, which describe how the potential outcome changes with the potential covariates. As a result of these statements, we only need to estimate the point causal effect to estimate the sequential causal effect and the blip effect.

In the next section, we will use Propositions 1–4 to derive the new  $G$ -formula, which expresses the sequential causal effect and the blip effect in terms of the point observable effects.

#### 4. New $G$ -formula for causal effects.

4.1. *Identifying condition.* When treatments are consecutively and randomly assigned/observed, the potential covariates and then the potential outcome are consecutively and randomly observed, typically from sequential randomized experiments or observational studies. The observable treatments, covariates and outcome are denoted by  $(\mathbf{Z}_1^T, \mathbf{X}_1^{T-1}, Y)$ . Their realizations are the observed values  $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y)$ . The distribution of the observable  $(\mathbf{Z}_1^T, \mathbf{X}_1^{T-1}, Y)$  is denoted by  $\mathbf{P}^{\mathbf{O}}(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y)$ , where the superscript  $\mathbf{O}$  indicates that these variables are observable. The *standard parameter* for the conditional distribution  $\mathbf{P}^{\mathbf{O}}(y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$  of  $Y$  given  $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$  is the conditional mean  $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = E(Y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ .

In sequential causal inference, we need to link the observable covariates and outcome  $(\mathbf{X}_1^{T-1}, Y)$  under the observable treatments  $\mathbf{Z}_1^T$  to the potential covariates and outcome

$\{\mathbf{X}_1^{T-1}(\mathbf{G}_1^{T-1}), y(\mathbf{G}_1^T)\}$  under regime  $\mathbf{G}_1^T$ , which can be random or deterministic or mixed; see Section 2.1 for the notation of these potential variables. To this end, Robins [4, 5, 7] introduced the following.

**THE IDENTIFYING CONDITION.** (1) *The consistency assumption: if the observed treatments  $\mathbf{z}_1^T$  are equal to realization of regime  $\mathbf{G}_1^T$ , then the observed covariates  $\mathbf{x}_1^{T-1}$  are equal to realization of the potential covariates  $\mathbf{X}_1^{T-1}(\mathbf{G}_1^{T-1})$ , and the observed outcome  $y$  is equal to realization of the potential outcome  $Y(\mathbf{G}_1^T)$ .* (2) *The assumption of no unmeasured confounders: given the observed history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , treatment  $Z_t$  satisfies*

$$\mathbf{X}_t^{T-1}(\mathbf{G}_t^{T-1}), Y(\mathbf{G}_t^T) \perp Z_t \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}$$

for a deterministic or random or mixed regime  $\mathbf{G}_t^T$  given the observed  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , where  $A \perp B \mid C$  means that  $A$  is conditionally independent of  $B$  given  $C$ . (3) *The positivity assumption: if  $P^O(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) > 0$ , then  $P^O(z_t \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) > 0$ .*

Under the identifying condition, a random regime  $\mathbf{G}_1^T = \mathbf{R}_1^T$  can be identified by

$$P^{\mathbf{R}_1^T}(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y) = P^{\mathbf{O}}(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y).$$

One can also use  $P^{\mathbf{O}}(\mathbf{z}_t^T, \mathbf{x}_t^{T-1}, y \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  to identify  $P^{\mathbf{G}_t^T}(\mathbf{z}_t^T, \mathbf{x}_t^{T-1}, y \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , where  $\mathbf{G}_t^T$  is deterministic or mixed. Here, we present the following three well-known formulas (10a)–(10c), which will be used to derive the new  $G$ -formula (e.g., see Robins [5] for the proofs of these formulas). In Supplement I of Supplementary Material [13], we also sketch the proofs. For mixed subregime  $\mathbf{G}_t^s = (\mathbf{D}_t^{s-1}, R_s)$  given the history  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , we have

$$(10a) \quad P^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = P^{\mathbf{O}}(z_s \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) \prod_{k=t}^{s-1} P^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}).$$

For mixed subregime  $\mathbf{G}_t^s = (D_t, \mathbf{R}_{t+1}^s)$  given  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , we have

$$(10b) \quad P^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = P^{\mathbf{O}}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}).$$

For deterministic regime  $\mathbf{G}_t^s = \mathbf{D}_t^s$  given  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ , we have

$$(10c) \quad P^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \prod_{k=t}^{s-1} P^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}).$$

**4.2. Well-known  $G$ -formula.** Under the identifying condition, Robins [4, 5, 7] derived the following  $G$ -formula (11) or (12), which expresses the sequential causal effect or the blip effect in terms of the standard parameters. Let  $\mathbf{a}_1^T$  be the treatments assigned by deterministic regime  $\mathbf{A}_1^T$  and  $\mathbf{b}_1^T$  by  $\mathbf{B}_1^T$ . The  $G$ -formula for the sequential causal effect is then

$$(11) \quad \text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) = E_1\{\mu(\mathbf{a}_1^T, \mathbf{x}_1^{T-1})\} - E_2\{\mu(\mathbf{b}_1^T, \mathbf{x}_1^{T-1})\},$$

where  $E_1(\cdot)$  is an expectation with respect to  $\prod_{t=1}^{T-1} P^{\mathbf{O}}(\mathbf{x}_t \mid \mathbf{z}_t^T, \mathbf{x}_1^{t-1})$  for  $\mathbf{z}_1^{T-1} = \mathbf{a}_1^{T-1}$  and  $E_2(\cdot)$  for  $\mathbf{z}_1^{T-1} = \mathbf{b}_1^{T-1}$ . The  $G$ -formula for the blip effect is

$$(12) \quad \begin{aligned} & \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &= E_1\{\mu(\mathbf{z}_1^t, \mathbf{z}_{t+1}^T = \mathbf{0}, \mathbf{x}_1^{T-1}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ & \quad - E_2\{\mu(\mathbf{z}_1^{t-1}, \mathbf{z}_t^T = \mathbf{0}, \mathbf{x}_1^{T-1}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\prod_{s=t}^{T-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_s | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, \mathbf{z}_t^s, \mathbf{x}_1^{s-1})$  for  $\mathbf{z}_t^{T-1} = (z_t, \mathbf{z}_{t+1}^{T-1} = \mathbf{0})$  and  $E_2(\cdot)$  for  $\mathbf{z}_t^{T-1} = \mathbf{0}$ . The explicit expression of the  $G$ -formula (11) or (12) in the form of summation or integration of the standard parameters is highly complex and difficult to use to estimate the blip and sequential causal effects, which is well known in the literature (Robins [4, 5, 7]) and discussed in the [Introduction](#) of this article.

4.3. *New G-formula.* Let  $\mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = E(Y | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$  ( $1 \leq t \leq T$ ), where the expectation is with respect to the conditional distribution  $\mathbf{P}^{\mathbf{O}}(y | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$  given  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ . At  $t = T$ , it is simply the standard parameter  $\mu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ . As well known in single-point causal inference (e.g., Rosenbaum and Rubin [8]), the *point observable effect* of treatment  $z_t > 0$  is

$$(13) \quad \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) = \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0).$$

The subsequent treatments and covariates  $(\mathbf{z}_{t+1}^T, \mathbf{x}_1^{T-1})$  after  $z_t$  do not appear in this expression; therefore, we can estimate  $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t)$  without knowing influences of these variables. Under the identifying condition, the point causal effect is equal to the point observable effect, that is,

$$(14) \quad \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) = \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t).$$

Formula (14) can be considered the new  $G$ -formula for the point causal effect.

The new  $G$ -formula for the blip and sequential causal effects is given by

**THEOREM 1.** *Under the identifying condition, the blip effect can be expressed in terms of the point observable effects by*

$$(15) \quad \begin{aligned} \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) &= \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &\quad - \sum_{s=t+1}^T E_1\{\vartheta(\mathbf{z}_{t+1}^{s-1} = \mathbf{0}, \mathbf{x}_1^{s-1}; z_s) | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t\} \\ &\quad + \sum_{s=t+1}^T E_2\{\vartheta(\mathbf{z}_{t+1}^{s-1} = \mathbf{0}, \mathbf{x}_1^{s-1}; z_s) | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{O}}(z_s | \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) \prod_{k=t}^{s-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for  $\mathbf{z}_t^{s-1} = (z_t, \mathbf{z}_{t+1}^{s-1} = \mathbf{0})$  and  $E_2(\cdot)$  for  $\mathbf{z}_t^{s-1} = \mathbf{0}$ . The sequential causal effect is expressed in terms of the point observable effects by

$$(16) \quad \begin{aligned} \text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) &= \vartheta(a_1) - \vartheta(b_1) \\ &\quad + \sum_{s=2}^T E_1\{\vartheta(\mathbf{a}_1^{s-1}, \mathbf{x}_1^{s-1}; a_s)\} - \sum_{s=2}^T E_2\{\vartheta(\mathbf{a}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} \\ &\quad - \sum_{s=2}^T E_3\{\vartheta(\mathbf{b}_1^{s-1}, \mathbf{x}_1^{s-1}; b_s)\} + \sum_{s=2}^T E_4\{\vartheta(\mathbf{b}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}. \end{aligned}$$

Here,  $E_1(\cdot)$  is an expectation with respect to  $\prod_{k=1}^{s-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for  $\mathbf{z}_1^{s-1} = \mathbf{a}_1^{s-1}$  and  $E_3(\cdot)$  for  $\mathbf{z}_1^{s-1} = \mathbf{b}_1^{s-1}$ . In addition,  $E_2(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{O}}(z_s | \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) \prod_{k=1}^{s-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for  $\mathbf{z}_1^{s-1} = \mathbf{a}_1^{s-1}$  and  $E_4(\cdot)$  for  $\mathbf{z}_1^{s-1} = \mathbf{b}_1^{s-1}$ .

PROOF. According to (10a) and (14), we have that the expectation

$$E\{\theta(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}$$

with respect to  $\mathbf{P}^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  for  $\mathbf{G}_t^s = (\mathbf{D}_t^{s-1}, R_s)$  is equal to

$$E\{\vartheta(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t\}$$

with respect to  $\mathbf{P}^{\mathbf{O}}(z_s \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) \prod_{k=t}^{s-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$ . Applying this equality to  $E_1(\cdot)$  and  $E_2(\cdot)$  in (9) and using (14), we obtain (15). Similarly, we prove formula (16) by applying (10a) and (10c) with  $t = 1$  and (14) to (8).  $\square$

The converse form of (15) and an alternative form of (16) will be used in the next section and are given by

**THEOREM 2.** *Under the identifying condition, the point observable effect is expressed in terms of the blip effects by*

$$(17) \quad \begin{aligned} \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) &= \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) \\ &+ \sum_{s=t+1}^T E_1\{\phi(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t\} \\ &- \sum_{s=t+1}^T E_2\{\phi(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0\}, \end{aligned}$$

where the expectation  $E_1(\cdot)$  is with respect to  $\mathbf{P}^{\mathbf{O}}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  and  $E_2(\cdot)$  to  $\mathbf{P}^{\mathbf{O}}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)$ . The sequential causal effect is expressed in terms of the blip effects by

$$(18) \quad \begin{aligned} \text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) &= \phi(a_1) - \phi(b_1) \\ &+ \sum_{s=2}^T E_1\{\phi(\mathbf{a}_1^{s-1}, \mathbf{x}_1^{s-1}; a_s)\} - \sum_{s=2}^T E_2\{\phi(\mathbf{b}_1^{s-1}, \mathbf{x}_1^{s-1}; b_s)\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\prod_{k=1}^{s-1} \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for  $\mathbf{z}_1^{s-1} = \mathbf{a}_1^{s-1}$  and  $E_2(\cdot)$  for  $\mathbf{z}_1^{s-1} = \mathbf{b}_1^{s-1}$ . Formula (17)/(18) implies that the point observable effect/the sequential causal effect can be decomposed into the blip effects at times  $s = t, \dots, T / s = 1, \dots, T$ .

PROOF. According to (10b), we have that the expectation

$$E\{\phi(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}$$

with respect to  $\mathbf{P}^{\mathbf{G}_t^s}(\mathbf{z}_t^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$  for  $\mathbf{G}_t^s = \{D_t, \mathbf{R}_{t+1}^s\}$  is equal to the expectation

$$E\{\phi(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}; z_s) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t\}$$

with respect to  $\mathbf{P}^{\mathbf{O}}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ . Applying this equality to  $E_1(\cdot)$  and  $E_2(\cdot)$  in (6) and using (14), we obtain (17). Similarly, we prove formula (18) by applying (10c) with  $t = 1$  and (14) to (5).  $\square$

Wang and Yin [12] proved a special case of formulas (15) and (17) based on the conditional probability  $\mathbf{P}^{\mathbf{O}}(y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$  of  $Y$  given  $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$  and under a restrictive condition, where  $(\mathbf{Z}_1^T, \mathbf{X}_1^{T-1})$  need to be discrete. Their proof is highly complex and does not reflect the underlying causal relationship given by formulas (6) and (9).

## 5. Applications of the new G-formula.

5.1. *Method of estimating the blip and sequential causal effects.* Theorems 1 and 2 imply that we can estimate both the blip and sequential causal effects via the point observable effects. In the future work, we will study comprehensive applications of Theorems 1 and 2 in sequential causal inference. Here, we apply Theorem 2 to develop a method of estimation with the following procedure. Suppose a data of observations  $(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i)$ ,  $i = 1, \dots, N$  from identical and independently distributed random variables  $(\mathbf{Z}_{i1}^T, \mathbf{X}_{i1}^{T-1}, Y_i)$ .

In the first stage, we estimate the point observable effects  $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t)$  by ML. We can apply the assignment conditions of individual treatments to weaken the positivity assumption in the identifying condition and only need to estimate fewer point observable effects that correspond to fewer point causal effects (Rosenbaum and Rubin [8]).

In the second stage, we treat (17) of Theorem 2 as a regression model to estimate the blip effects  $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t)$ , where the response variable is the estimate  $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t)$  and the explanatory variables are the (observed) proportions that correspond to the probabilities  $\mathbf{P}^{\mathbf{O}}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s | \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ . Here, we can apply the *structural nested mean model (SNMM)*, which characterizes the blip effects by

$$\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) = f(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi),$$

where  $f(\cdot)$  is a deterministic function of  $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$  indexed by a parameter vector  $\varphi$  (Robins [5, 7]). Noticeably, there is no restriction on the form of SNMM. However, the dimension of the parametric vector  $\varphi$  is smaller than or equal to that of the estimated point observable effects. SNMM plays two roles in estimating the blip effects. First, SNMM leads to an unsaturated model for point observable effects. This unsaturated model is not necessarily misspecified and so may improve precision of the estimation without causing bias. Second, together with treatment assignment conditions, SNMM also makes it possible to identify and estimate the blip effects via fewer point observable effects (see Petersen et al. [3] for a description of the relationship between the positivity assumption and statistical modeling); therefore, it is possible to avoid the curse of dimensionality in estimating the blip effects. However, possible misspecification of SNMM may lead to bias in estimating the blip effects. Possible violation of the identifying condition, in particular, the assumption of no unmeasured confounders, may lead to violation of (17), which in turn leads to bias in estimating the blip effects.

In the third stage, we use the estimates  $\hat{\varphi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t)$  to estimate the sequential causal effects  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$  based on (18) of Theorem 2, where the probabilities  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for given  $\mathbf{z}_1^k = \mathbf{a}_1^k$  or  $\mathbf{b}_1^k$  is estimated by ML. Because the covariates  $\mathbf{x}_1^{T-1}$  are of secondary interest compared to the outcome  $y$ , we can estimate these probabilities in a flexible manner, for instance, via standard parameters.

Unfortunately, the estimation of these probabilities is subject to the curse of dimensionality in the presence of a long treatment sequence and many covariates. Parametric models in terms of standard parameters are frequently used to alleviate the curse of dimension in estimating these probabilities, although such models are subject to the null paradox and may lead to biases in estimating these probabilities and thus  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . In (18) replacing the probabilities and blip effects by their estimates, we obtain the estimate for  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . When the probabilities and the blip effects have too complex structures to apply (18), it may become necessary to perform the Monte Carlo simulation to estimate  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . Denote the estimates of the probabilities by  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$ . Following Robins [7], we have the following Monte Carlo algorithm:

- Step 1: for each  $v = (1, \dots, V)$ , do the following:

- recursively for  $k = 1, \dots, T - 1$ , predict  $\mathbf{x}_k$  according to  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for  $\mathbf{z}_1^k = \mathbf{a}_1^k$  or  $\mathbf{z}_1^k = \mathbf{b}_1^k$ .
- calculate  $\hat{\Delta}_v^{\mathbf{A}_1^T} = \sum_{s=1}^T \hat{\phi}(\mathbf{a}_1^{s-1}, \mathbf{x}_1^{s-1}; a_s)$  and  $\hat{\Delta}_v^{\mathbf{B}_1^T} = \sum_{s=1}^T \hat{\phi}(\mathbf{b}_1^{s-1}, \mathbf{x}_1^{s-1}; b_s)$ .
- Step 2: calculate  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T) = \frac{1}{V} \sum_{v=1}^V (\hat{\Delta}_v^{\mathbf{A}_1^T} - \hat{\Delta}_v^{\mathbf{B}_1^T})$ .

Under special circumstances, we may further improve the estimation of the probabilities  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$ . For instance, when  $\mathbf{x}_k$  ( $k = 1, \dots, T - 1$ ) are measurements of the same type of covariates, we may sometimes have such conditions as  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{x}_s^{k-1}, \mathbf{z}_s^k)$  with  $1 < s < k$ . Under these conditions, we estimate  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{x}_s^{k-1}, \mathbf{z}_s^k)$  even without a parametric model. In such cases, the estimates for  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{x}_s^{k-1}, \mathbf{z}_s^k)$  is not necessarily subject to biases and the curse of dimensionality, and neither is the estimate for  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

Occasionally, the blip effects are not dependent on the time-dependent covariates, that is,  $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t) = \phi(\mathbf{z}_1^{t-1}; z_t)$ . If both  $\mathbf{A}_1^T$  and  $\mathbf{B}_1^T$  are also static regimes, then (18) becomes

$$\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T) = \sum_{s=1}^T \phi(\mathbf{a}_1^{s-1}; a_s) - \sum_{s=1}^T \phi(\mathbf{b}_1^{s-1}; b_s),$$

which does not contain the probabilities  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$ . In this case, we only need to estimate  $\phi(\mathbf{z}_1^{t-1}; z_t)$  to estimate  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . If the estimates for  $\phi(\mathbf{z}_1^{t-1}; z_t)$  are not subject to biases and the curse of dimensionality, neither is the estimate for  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

In the fourth stage, we apply the bootstrap method to the first three stages to obtain the interval estimates for the blip and sequential causal effects, which incorporates variability of the treatments and covariates and outcome.

In the rest of the article, we will illustrate our method above via a theoretical example, a simulation study and a medical example.

**5.2. Theoretical example.** Here, we consider the case in which treatments  $Z_t = 0, 1$  and covariates  $\mathbf{X}_{t-1}$  are discrete ( $t = 1, \dots, T$ ). The treatment assignment follows the *first-order Markov process*, in which the assignment of  $z_t$  only depends on the latest covariate  $\mathbf{x}_{t-1}$ . Suppose that SNMM is of the form

$$(19) \quad \begin{cases} \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}; z_t = 1) = \varphi_1, & t \leq T - 2, \\ \phi(\mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}; z_{T-1} = 1) = \varphi_2, \\ \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}; z_T = 1) = \varphi_3, \end{cases}$$

where  $\varphi_1$  is the blip effects of distant treatments,  $\varphi_2$  is the blip effects of intermediate treatments, and  $\varphi_3$  is the blip effects of the latest treatments.

Despite the Markov process and SNMM, treatment  $z_t$  still has  $\mathbf{x}_{t-1}$  as a confounder and  $\mathbf{x}_t$  as a posttreatment variable in the expression for the standard parameters  $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ . Therefore, based on the well-known  $G$ -formula (11) or (12), the ML estimation of the blip effects  $\varphi_1, \varphi_2, \varphi_3$  and the sequential causal effect  $\text{sce}(\mathbf{a}_1^T; \mathbf{b}_1^T)$  via the standard parameters suffers from the null paradox and the curse of dimensionality in the case of a long treatment sequence, as has been demonstrated in the literature (Robins [4, 5, 7]) and discussed in the **Introduction**. In the following, we will estimate  $\varphi_1, \varphi_2, \varphi_3$  and  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$  by applying our method introduced in Section 5.1.

In the first stage, we estimate the point observable effects by ML as follows. As well known in single-point causal inference (Rosenbaum and Rubin [8]), due to the assignment condition of the Markov process, we weaken the positivity assumption in the identifying condition by requiring  $0 < \mathbf{P}^{\mathbf{O}}(z_t = 1 | \mathbf{x}_{t-1}) < 1$  instead of  $0 < \mathbf{P}^{\mathbf{O}}(z_t = 1 | \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) < 1$ , and we only need to estimate fewer point observable effects of  $z_t = 1$  in strata  $\mathbf{x}_{t-1}$  instead of strata

$(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ . Let  $\mu(\mathbf{x}_{t-1}, z_t) = E(Y | \mathbf{x}_{t-1}, z_t)$  denote the conditional mean of the observable outcome  $Y$  given the observed  $(\mathbf{x}_{t-1}, z_t)$ . Then the point observable effect of  $z_t = 1$  in strata  $\mathbf{x}_{t-1}$ , denoted by  $\vartheta(\mathbf{x}_{t-1})$ , is

$$(20) \quad \vartheta(\mathbf{x}_{t-1}) = \mu(\mathbf{x}_{t-1}, z_t = 1) - \mu(\mathbf{x}_{t-1}, z_t = 0).$$

The number of  $\vartheta(\mathbf{x}_{t-1})$  is equal to that of strata  $\mathbf{x}_{t-1}$  and does not depend on the length  $T$  of treatment sequence; therefore,  $\vartheta(\mathbf{x}_{t-1})$  are estimable even for a long treatment sequence and the typical size of a sample.

As described in Section 5.1, we will perform a regression of the ML estimate of the point observable effect on the proportions of treatments and covariates in the second stage. Hence, we are going to estimate the point observable effect conditional on covariates and treatments. Let  $n(\mathbf{x}_{t-1}, z_t)$  be the number of observations in stratum  $(\mathbf{x}_{t-1}, z_t)$ , and  $I(\mathbf{x}_{t-1}, z_t)$  be the set of all indexes  $i$  such that  $(\mathbf{z}_{i(t-1)}, \mathbf{x}_{i(t-1)}) = (\mathbf{z}_{t-1}, \mathbf{x}_{t-1})$ . Then, for common distributions, such as the normal, Bernoulli and Poisson, the ML estimate of  $\mu(\mathbf{x}_{t-1}, z_t)$  is simply the average

$$\hat{\mu}(\mathbf{x}_{t-1}, z_t) = \frac{\sum_{i \in I(\mathbf{x}_{t-1}, z_t)} Y_i}{n(\mathbf{x}_{t-1}, z_t)}.$$

For convenience, suppose the same variance  $\sigma^2$  for  $Y$  given any  $\{\mathbf{z}_1^T, \mathbf{x}_1^{T-1}\}$ . Then the variance of  $\hat{\mu}(\mathbf{x}_{t-1}, z_t)$  is

$$\text{var}\{\hat{\mu}(\mathbf{x}_{t-1}, z_t)\} = \frac{\sigma^2}{n(\mathbf{x}_{t-1}, z_t)}.$$

If the variance is not the same, some minor loss of efficiency may occur in the ML estimation. The ML estimates of the point observable effects and their variances are

$$\begin{aligned} \hat{\vartheta}(\mathbf{x}_{t-1}) &= \hat{\mu}(\mathbf{x}_{t-1}, z_t) - \hat{\mu}(\mathbf{x}_{t-1}, z_t = 0), \\ \text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\} &= \text{var}\{\hat{\mu}(\mathbf{x}_{t-1}, z_t)\} + \text{var}\{\hat{\mu}(\mathbf{x}_{t-1}, z_t = 0)\}. \end{aligned}$$

In Supplement I of Supplementary Material [13], we prove that under the Markov process, the covariance between the estimated point observable effects at different times satisfies

$$(21) \quad \text{cov}\{\hat{\vartheta}(\mathbf{x}_{t-1}), \hat{\vartheta}(\mathbf{x}_{s-1})\} = 0, \quad t \neq s.$$

Formula (21) will considerably simplify the regression in the second stage.

In the second stage, we estimate the blip effects by regression as follows. Intuitively, we decompose the point observable effect  $\vartheta(\mathbf{x}_{t-1})$  into the blip effects  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$  at times  $s = t, \dots, T$  in stratum  $(\mathbf{x}_{t-1}, z_t)$  versus  $(\mathbf{x}_{t-1}, z_t = 0)$  and obtain

$$(22) \quad \vartheta(\mathbf{x}_{t-1}) = \varphi_1 c^{(1)}(\mathbf{x}_{t-1}) + \varphi_2 c^{(2)}(\mathbf{x}_{t-1}) + \varphi_3 c^{(3)}(\mathbf{x}_{t-1}),$$

where

$$c^{(1)}(\mathbf{x}_{t-1}) = \sum_{s=t}^{T-2} \{\text{pr}(z_s = 1 | \mathbf{x}_{t-1}, z_t = 1) - \text{pr}(z_s = 1 | \mathbf{x}_{t-1}, z_t = 0)\},$$

$$c^{(2)}(\mathbf{x}_{t-1}) = \text{pr}(z_{T-1} = 1 | \mathbf{x}_{t-1}, z_t = 1) - \text{pr}(z_{T-1} = 1 | \mathbf{x}_{t-1}, z_t = 0),$$

$$c^{(3)}(\mathbf{x}_{t-1}) = \text{pr}(z_T = 1 | \mathbf{x}_{t-1}, z_t = 1) - \text{pr}(z_T = 1 | \mathbf{x}_{t-1}, z_t = 0).$$

Noticeably, if  $t = T - 1$ , then  $c^{(1)}(\mathbf{x}_{T-2}) = 0$ ; if  $t = T$ , then  $c^{(1)}(\mathbf{x}_{T-1}) = 0$ ,  $c^{(2)}(\mathbf{x}_{T-1}) = 0$ . Conditional on treatments and covariates,  $c^{(1)}(\mathbf{x}_{t-1})$ ,  $c^{(2)}(\mathbf{x}_{t-1})$  and  $c^{(3)}(\mathbf{x}_{t-1})$  are constants, which describe the proportion differences in distant, intermediate and the latest treatments in

stratum  $(\mathbf{x}_{t-1}, z_t = 1)$  versus stratum  $(\mathbf{x}_{t-1}, z_t = 0)$ , respectively. In Supplement I of Supplementary Material [13], we derive this formula by applying the Markov process and (19) to (17). Here, we see that the blip effects  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$  are identifiable via the point observable effects  $\vartheta(\mathbf{x}_{t-1})$ . Formula (22) is indexed by only three parameters  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$ , and thus an unsaturated model for all  $\vartheta(\mathbf{x}_{t-1})$  at  $t = 1, \dots, T$ . This unsaturated model reflects the data-generating mechanism and is therefore not misspecified; therefore, bias does not occur in estimating the blip effects despite an unsaturated model. Like  $\vartheta(\mathbf{x}_{t-1})$ , the number of  $\mathbf{P}^{\mathbf{O}}(z_s = 1 \mid \mathbf{x}_{t-1}, z_t)$  does not depend on the length  $T$  of the treatment sequence; therefore, the curse of dimensionality does not occur in the estimation of  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$  as  $T$  increases. Actually, more point observable effects  $\vartheta(\mathbf{x}_{t-1})$  will be observed as  $T$  increases, and thus more information about  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$  will be available, which improves the estimation.

We estimate  $c^{(1)}(\mathbf{x}_{t-1})$ ,  $c^{(2)}(\mathbf{x}_{t-1})$  and  $c^{(3)}(\mathbf{x}_{t-1})$  by  $\hat{c}^{(1)}(\mathbf{x}_{t-1})$ ,  $\hat{c}^{(2)}(\mathbf{x}_{t-1})$  and  $\hat{c}^{(3)}(\mathbf{x}_{t-1})$ , respectively, in which the probabilities  $\mathbf{P}^{\mathbf{O}}(z_s = 1 \mid \mathbf{x}_{t-1}, z_t)$  are replaced by the proportions  $\hat{\mathbf{P}}^{\mathbf{O}}(z_s = 1 \mid \mathbf{x}_{t-1}, z_t)$ . Let  $\Theta$  denote the vector with subvectors  $\Theta_t$  ( $t = 1, \dots, T$ ), where the elements of  $\Theta_t$  are  $\hat{\vartheta}(\mathbf{x}_{t-1})$ . Let  $\Omega$  denote the diagonal matrix with diagonal submatrices  $\Omega_{tt}$  ( $t = 1, \dots, T$ ), where the diagonal elements of  $\Omega_{tt}$  are  $\text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\}$ . Let  $\mathbf{X}$  denote the design matrix with submatrices  $\mathbf{X}_{t1}$  ( $t = 1, \dots, T$ ), where the row vector of  $\mathbf{X}_{t1}$  is  $\{\hat{c}^{(1)}(\mathbf{x}_{t-1}), \hat{c}^{(2)}(\mathbf{x}_{t-1}), \hat{c}^{(3)}(\mathbf{x}_{t-1})\}$ . Based on (22) and using (21), we regress  $\hat{\vartheta}(\mathbf{x}_{t-1})$  on  $\hat{c}^{(1)}(\mathbf{x}_{t-1})$ ,  $\hat{c}^{(2)}(\mathbf{x}_{t-1})$  and  $\hat{c}^{(3)}(\mathbf{x}_{t-1})$  to obtain the ML estimates

$$(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3)' = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\Theta.$$

These ML estimates are consistent under certain weak conditions, even for long treatment sequence and/or when many covariates are present, for example, the  $\hat{\vartheta}(\mathbf{x}_{t-1})$ ,  $\hat{c}^{(1)}(\mathbf{x}_{t-1})$ ,  $\hat{c}^{(2)}(\mathbf{x}_{t-1})$  and  $\hat{c}^{(3)}(\mathbf{x}_{t-1})$  have their variances approaching zero as the sample size approaches infinity. Such conditions are required even in single-point causal inference.

In the third stage, we estimate the sequential causal effect as follows. We replace the blip effect  $\varphi_1$  by  $\hat{\varphi}_1$ ,  $\varphi_2$  by  $\hat{\varphi}_2$ ,  $\varphi_3$  by  $\hat{\varphi}_3$ , and the probabilities  $\mathbf{P}^{\mathbf{O}}(\cdot)$  by the estimates  $\hat{\mathbf{P}}^{\mathbf{O}}(\cdot)$  in formula (18) and obtain

$$\begin{aligned} & \widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T) \\ &= \hat{\varphi}_1 \{I(a_1 = 1) - I(b_1 = 1)\} \\ &+ \hat{\varphi}_1 \left[ \sum_{t=2}^{T-2} \sum_{\mathbf{x}_1^{t-1}} a_t \prod_{k=1}^{t-1} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{a}_1^k, \mathbf{x}_1^{k-1}) - \sum_{t=2}^{T-2} \sum_{\mathbf{x}_1^{t-1}} b_t \prod_{k=1}^{t-1} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{b}_1^k, \mathbf{x}_1^{k-1}) \right] \\ &+ \hat{\varphi}_2 \left[ \sum_{\mathbf{x}_1^{T-2}} a_{T-1} \prod_{k=1}^{T-2} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{a}_1^k, \mathbf{x}_1^{k-1}) - \sum_{\mathbf{x}_1^{T-2}} b_{T-1} \prod_{k=1}^{T-2} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{b}_1^k, \mathbf{x}_1^{k-1}) \right] \\ &+ \hat{\varphi}_3 \left[ \sum_{\mathbf{x}_1^{T-1}} a_T \prod_{k=1}^{T-2} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{a}_1^k, \mathbf{x}_1^{k-1}) - \sum_{\mathbf{x}_1^{T-1}} b_T \prod_{k=1}^{T-2} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{b}_1^k, \mathbf{x}_1^{k-1}) \right], \end{aligned}$$

where  $I(a_1 = 1)$  equals one if  $a_1 = 1$  is true and zero otherwise, and similarly for  $I(b_1 = 1)$ . If  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  is only the proportion of  $\mathbf{x}_k$  in stratum  $(\mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  in the data, then it may not be a consistent estimate of  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  for a long treatment sequence due to the curse of dimensionality, nor may  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T)$  be a consistent estimate for  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . If  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  is obtained by a parametric model in terms of standard parameters, it possesses a higher efficiency and so does  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ . However,  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  is biased due to the null paradox, and so is  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

However, the estimation of  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$  can be improved without introducing bias under certain circumstances. For instance, when  $z_k$  and  $\mathbf{x}_k$  ( $k = 1, \dots, T - 1$ ) are repeated measurements of the same treatment variable and covariates, we may have such conditions as  $\mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = \mathbf{P}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{x}_{k-1}, z_k)$ . In this case,  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{x}_{k-1}, z_k)$  is consistent even for long treatment sequences, and so is  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

We notice that

$$\sum_{\mathbf{x}_1^{t-1}} \prod_{k=1}^{t-1} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{t-1}} \prod_{k=1}^{t-1} \hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = 1.$$

Therefore, for two static regimes  $\mathbf{A}_1^T = \mathbf{a}_1^T$  and  $\mathbf{B}_1^T = \mathbf{b}_1^T$ , we have

$$\begin{aligned} \widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T) &= \hat{\varphi}_1 \sum_{t=1}^{T-2} \{I(a_t = 1) - I(b_t = 1)\} \\ &\quad + \hat{\varphi}_2 \{I(a_{T-1} = 1) - I(b_{T-1} = 1)\} \\ &\quad + \hat{\varphi}_3 \{I(a_T = 1) - I(b_T = 1)\}, \end{aligned}$$

which does not contain the estimates  $\hat{\mathbf{P}}^{\mathbf{O}}(\mathbf{x}_k | \mathbf{z}_1^k, \mathbf{x}_1^{k-1})$ . As a result, if  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$  and  $\hat{\varphi}_3$  are consistent for a long treatment sequence, then so is  $\widehat{\text{sce}}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

In the last stage, we apply the bootstrap method to the first three stages to obtain interval estimates for  $\varphi_1$ ,  $\varphi_2$ ,  $\varphi_3$  and  $\text{sce}(\mathbf{A}_1^T; \mathbf{B}_1^T)$ .

The method described in this subsection can be readily extended to higher-order Markov processes, where the assignments of treatments are dependent on a finite history of treatments and discrete covariates or finite numbers of intervals of the values of continuous covariates and to more general forms of SNMM, where the blip effects are indexed by more than three parameters.

**5.3. Simulation study.** The treatment sequence has a length of  $T = 3$ . The treatments are dichotomous with  $Z_t = 0, 1$  ( $t = 1, 2, 3$ ). Between treatments  $Z_{t-1}$  and  $Z_t$  ( $t > 1$ ), there is a time-dependent covariate vector  $\mathbf{X}_{t-1} = (0, 0), (0, 1), (1, 0), (1, 1)$ . After the last treatment  $Z_3$ , there is an outcome  $Y$  of interest. The treatment assignment follows the Markov process described in the previous subsection, that is, the assignment of  $z_2$  depends only on  $\mathbf{x}_1$  and that of  $z_3$  only on  $\mathbf{x}_2$ . According to (20), there are a total of nine point observable effects: one  $\vartheta$  of  $z_1 = 1$ , four  $\vartheta(\mathbf{x}_1)$  of  $z_2 = 1$  with  $\mathbf{x}_1 = (0, 0), (0, 1), (1, 0), (1, 1)$  and four  $\vartheta(\mathbf{x}_2)$  of  $z_3 = 1$  with  $\mathbf{x}_2 = (0, 0), (0, 1), (1, 0), (1, 1)$ .

We suppose that two structural nested mean models SNMM1 and SNMM2 are available. In SNMM1, there are nine blip effects: one blip effect of  $z_1 = 1$  denoted by  $\varphi$ , four blip effects of  $z_2 = 1$  depending only on  $\mathbf{x}_1 = (0, 0), (0, 1), (1, 0), (1, 1)$  and denoted by  $\varphi(\mathbf{x}_1)$ , and four blip effects of  $z_3 = 1$  depending only on  $\mathbf{x}_2 = (0, 0), (0, 1), (1, 0), (1, 1)$  and denoted by  $\varphi(\mathbf{x}_2)$ . In SNMM2, it is further required that  $\varphi(\mathbf{x}_1) = \varphi(\mathbf{x}_2)$  if  $\mathbf{x}_1 = \mathbf{x}_2$ ; therefore, there are four different blip effects for both  $z_2 = 1$  and  $z_3 = 1$  in addition to one blip effect for  $z_1 = 1$ . In Supplement II of Supplementary Material [13], we construct three data-generating mechanisms for the normal, dichotomous and Poisson outcomes. From each of the three data-generating mechanisms, we generate a total of 1000 data sets, each having 400 independent observations on  $(\mathbf{Z}_1^3, \mathbf{X}_1^2, Y)$ .

In the simulation, we apply five methods to estimate the blip and sequential causal effects. Methods (i) and (ii) are constructed using our method described in Section 5.1. In method (i), we impose SNMM1 and estimate nine blip effects. In method (ii), we impose SNMM2 and thus estimate five different blip effects. Method (iii) is constructed using the well-known  $G$ -formula (11) or (12) (Taubman et al. [11]); method (iv) is constructed using the marginal

structural model based on inverse probability weighting (Robins [6, 7]); method (v) is constructed using the  $G$ -estimation based on SNMM1 (Robins [5, 7]). In methods (iii), (iv) and (v), we estimate nine blip effects compared to four in method (ii). In Supplement II of Supplementary Material [13], we describe methods (i)–(v) and the simulation in detail.

Methods (i) and (ii) need only nine point observable effects to estimate all blip and sequential causal effects. As described in Supplement II of Supplementary Material [13], method (iii) needs 128 standard parameters to estimate these causal effects; method (iv) needs 32 parameters and a specification of the stabilized weights to estimate the blip effects and only sequential causal effects of static regimes without excessive programming; and method (v) needs 18 parameters and a specification of the variability of pseudo outcomes to estimate only the blip effects without excessive programming. The result of the simulation is presented in Table 1, from which the following comparisons can be made between these methods.

Regarding the estimates for the blip and sequential causal effects, the following observations can be made. First, all methods yield unbiased estimates for the blip effects of  $z_3 = 1$  (columns (f)–(i) in Table 1). Noticeably, methods (i), (iii), (iv) and (v) yield nearly the same estimates due to the setting of the simulation. Second, methods (i), (ii), (iv) and (v) yield unbiased estimates for the blip effects of  $z_1 = 1$  and  $z_2 = 1$  (columns (a)–(e) in Table 1) and two sequential causal effects (columns (j) and (k) in Table 1), whereas method (iii) yields estimates of certain biases for these causal effects indicating slow convergence of these estimates to the true values.

The following interesting observations can be made for variances in the estimates of the blip and sequential causal effects as well as actual coverage probabilities for the confidence intervals of these causal effects. First, methods (i), (iii), (iv) and (v) yield nearly the same variances and coverage probabilities for the blip effects of  $z_3 = 1$  (columns (f)–(i) in Table 1) due to the setting of the simulation. Second, method (i) achieves considerably smaller variances and more accurate coverage probabilities than methods (iii) and (iv) do for the blip effects of  $z_1 = 1$  and  $z_2 = 1$  (columns (a)–(e) in Table 1) and two sequential causal effects (columns (j) and (k) in Table 1). Third, method (ii) achieves the smallest variances of all methods and nearly the nominal coverage probabilities, which implies that an unsaturated model may improve the estimation without causing biases. Fourth, method (v) has even smaller variances than method (i), although method (i) is based on ML and theoretically yields the variance bounds under SNMM1. As shown in Supplement II of Supplementary Material [13], this result is due to an inadequate evaluation of the variability of pseudo outcomes in method (v). With a correct specification of this variability, method (v) is equivalent to method (i). Method (i), however, does not need a specification of this variability.

Additionally, we have constructed a misspecified model for the standard parameters  $\mu(\mathbf{z}_1^3, \mathbf{x}_1^2)$  in which equalities are imposed between these parameters and then applied method (iii) to estimate the blip and sequential causal effects by ML in the simulation. We have observed the null paradox, that is, the biases for the ML estimates of these causal effects (see discussion in the [Introduction](#) and references therein). The result is not shown here, although the SAS code is given in Supplement II of Supplementary Material [13] for interested readers.

**5.4. Medical example.** Some HIV-infected gay or bisexual men are users of recreational drugs such as cocaine. Of medical interest is the distant influence of the recreational drug on the CD4 count when the drug is used repeatedly. Our data is a subset of the data from Multicenter AIDS Cohort Study and involves 256 participants who were seronegative at entry and seroconverted during the follow-up period (Zeger and Diggle [14]). Here, we apply our method and three methods in the literature to this data. These methods estimate the blip

TABLE 1

Simulation study of the causal effects in Section 5.3: average estimate (estimate), variance estimate (variance) and actual coverage probability (coverage) of the 95% bootstrap percentile confidence interval. 1000 data sets are used, each having 400 units. The bootstrap percentile confidence interval is obtained using 1000 bootstrap data sets

	True value and (estimate, variance, coverage)' of the causal effect										
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
	Normal outcome										
True value	3	-4	-5	5	4	-4	-5	5	4	3	2/3
Method (i)	2.89	-3.93	-4.99	4.88	3.89	-3.98	-5.04	5.02	4.03	2.87	0.57
	0.73	0.68	1.80	2.42	2.15	2.01	2.68	2.58	2.72	2.30	1.09
	93.90	96.80	95.60	94.70	96.20	95.20	95.10	95.20	95.30	94.90	94.50
Method (ii)	2.94	-3.96	-4.97	4.95	3.96					2.92	0.62
	0.69	0.62	1.27	1.35	1.18	(b)	(c)	(d)	(e)	2.27	0.96
	94.50	94.70	94.80	94.60	94.60					95.10	94.40
Method (iii)	2.11	-4.20	-4.70	5.22	2.46	-3.98	-5.04	5.02	4.03	2.79	0.33
	3.79	0.49	4.23	4.34	7.47	2.01	2.68	2.58	2.72	5.12	3.14
	93.70	86.80	92.60	93.70	71.60	94.80	94.90	95.40	95.70	93.10	93.40
Method (iv)	3.05	-4.08	-5.02	4.96	3.64	-3.98	-5.04	5.02	4.03	3.17	
	3.06	0.67	2.16	2.90	3.02	2.01	2.68	2.58	2.72	5.14	
	92.90	91.50	94.60	93.30	93.20	94.80	94.90	95.40	95.70	94.30	
Method (v)	2.89	-3.96	-4.99	4.84	3.89	-3.98	-5.04	5.02	4.03		
	0.71	0.58	1.40	1.93	2.06	2.01	2.68	2.58	2.72		
	93.80	96.10	95.80	94.20	96.10	94.80	94.90	95.40	95.70		

TABLE 1  
(Continued)

	True value and ( <i>estimate, variance, coverage</i> )' of the causal effect										
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
	Dichotomous outcome										
True value	0.2	-0.15	0.15	0.1	-0.1	-0.15	0.15	0.1	-0.1	0.117	0.225
Method (i)	0.195	-0.149	0.136	0.094	-0.101	-0.156	0.147	0.098	-0.099	0.106	0.218
	0.003	0.010	0.011	0.013	0.012	0.010	0.011	0.011	0.012	0.010	0.004
	95.20	96.70	95.80	94.10	96.30	95.90	94.10	94.80	95.50	95.30	95.50
Method (ii)	0.198	-0.153	0.141	0.096	-0.100					0.108	0.220
	0.003	0.005	0.006	0.006	0.005	(b)	(c)	(d)	(e)	0.009	0.004
	95.10	96.30	95.00	95.20	94.90					95.00	95.60
Method (iii)	0.170	-0.152	0.153	0.110	-0.103	-0.156	0.147	0.098	-0.099	0.109	0.198
	0.015	0.023	0.031	0.034	0.024	0.010	0.011	0.011	0.012	0.018	0.014
	92.40	96.30	89.30	91.40	93.40	95.60	94.40	94.90	95.10	93.00	93.50
Method (iv)	0.198	-0.137	0.147	0.102	-0.086	-0.156	0.147	0.098	-0.099	0.113	
	0.016	0.024	0.030	0.034	0.025	0.010	0.011	0.011	0.012	0.016	
	93.10	95.70	92.40	93.50	93.10	95.60	94.40	94.90	95.10	94.40	
Method (v)	0.197	-0.147	0.143	0.104	-0.099	-0.156	0.147	0.098	-0.099		
	0.003	0.011	0.011	0.013	0.012	0.010	0.011	0.011	0.012		
	95.10	96.30	95.70	94.10	95.20	95.70	94.40	94.80	95.10		

TABLE 1  
(Continued)

	True value and ( <i>estimate, variance, coverage</i> )' of the causal effect										
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
	Poisson outcome										
True value	2	4	3	-3	-4	4	3	-3	-4	2	3.67
Method (i)	1.95	3.94	2.97	-2.95	-3.96	4.02	3.03	-3.00	-3.97	1.98	3.63
	0.18	0.81	0.79	0.86	0.65	0.65	0.59	0.74	0.57	0.67	0.27
	95.90	96.00	95.80	96.00	96.60	95.60	95.10	94.30	94.70	95.80	96.00
Method (ii)	1.97	3.98	2.99	-2.97	-3.96					2.00	3.63
	0.17	0.36	0.37	0.43	0.30	(b)	(c)	(d)	(e)	0.65	0.27
	95.00	94.70	95.20	94.80	95.30					94.90	94.90
Method (iii)	1.66	3.40	3.07	-2.77	-4.41	4.02	3.02	-3.00	-3.97	1.96	3.18
	1.52	2.55	1.72	2.71	1.56	0.65	0.59	0.74	0.57	1.97	1.62
	95.00	81.80	93.90	89.60	87.80	95.80	95.70	94.30	94.50	94.40	94.30
Method (iv)	2.01	3.99	2.98	-3.00	-4.03	4.02	3.02	-3.00	-3.97	2.04	
	0.89	1.58	1.38	1.70	1.30	0.65	0.59	0.74	0.57	1.10	
	95.30	92.90	94.10	92.90	93.80	95.80	95.70	94.30	94.50	95.10	
Method (v)	1.95	3.96	2.97	-2.97	-3.96	4.02	3.02	-3.00	-3.97		
	0.18	0.72	0.62	0.69	0.58	0.65	0.59	0.74	0.57		
	95.10	94.60	94.30	95.00	95.10	95.80	95.70	94.30	94.50		

- Three outcome types: normal, dichotomous and Poisson outcomes.
- Nine blip effects: (a)  $\varphi$ ; (b)  $\varphi\{\mathbf{x}_1 = (0, 0)\}$ ; (c)  $\varphi\{\mathbf{x}_1 = (0, 1)\}$ ; (d)  $\varphi\{\mathbf{x}_1 = (1, 0)\}$ ; (e)  $\varphi\{\mathbf{x}_1 = (1, 1)\}$ ; (f)  $\varphi\{\mathbf{x}_2 = (0, 0)\}$ ; (g)  $\varphi\{\mathbf{x}_2 = (0, 1)\}$ ; (h)  $\varphi\{\mathbf{x}_2 = (1, 0)\}$ ; (i)  $\varphi\{\mathbf{x}_2 = (1, 1)\}$ .
- Two sequential causal effects  $\text{sce}(\mathbf{A}_1^3; \mathbf{B}_1^3)$ : (j)  $\mathbf{A}_1^3 = (1, 1, 1)$  and  $\mathbf{B}_1^3 = (0, 0, 0)$ ; (k)  $\mathbf{A}_1^3 = (1, 0, A_3)$  and  $\mathbf{B}_1^3 = (0, 0, 0)$  where  $A_3 = 1$  when  $\mathbf{x}_2 = (0, 0)$  or  $(0, 1)$  and  $A_3 = 0$  otherwise.
- Five estimation methods: method (i) our method without constraint on (a)–(i); method (ii) our method with constraint (b) = (f), (c) = (g), (d) = (h), (e) = (i); method (iii) the parametric method based on the well-known G-formula; method (iv) the marginal structural model based on inverse probability weighting; method (v) the G-estimation based on SNMM1. Empty cells imply that they are not easily estimable by the method.

TABLE 2

*Medical study of the causal effects of recreational drugs on CD4 count in Section 5.4: estimate, bootstrap variance estimate (variance) and 95% bootstrap percentile confidence interval (95% CI). The variance and confidence intervals are estimated using 1000 bootstrap data sets*

	(Estimate, variance, 95% CI) <sup>a</sup> of the causal effect			
	(a)	(b)	(c)	(d)
Method (i)	0.044 0.004 (-0.080, 0.162)	0.082 0.004 (-0.034, 0.197)	0.126 0.007 (-0.027, 0.289)	0.083 0.004 (-0.047, 0.211)
Method (iii)	-0.147 0.050 (-0.667, 0.192)	0.082 0.004 (-0.036, 0.212)	-0.126 0.050 (-0.664, 0.209)	-0.128 0.049 (-0.664, 0.198)
Method (iv)	0.032 0.019 (-0.246, 0.304)	0.121 0.012 (-0.067, 0.384)	0.110 0.009 (-0.069, 0.295)	
Method (v)	0.045 0.004 (-0.072, 0.159)	0.082 0.004 (-0.035, 0.212)		

- Outcome: the logarithm of CD 4 count after drug use  $z_2$ .
- Two blip effects: (a) the blip effect of  $z_1 = 1$ ; (b) the blip effect of  $z_2 = 1$ .
- Two sequential causal effects  $\text{sce}(\mathbf{A}_1^2, \mathbf{B}_1^2)$ : (c)  $\mathbf{A}_1^2 = (1, 1)$  and  $\mathbf{B}_1^2 = (0, 0)$ ; (d)  $\mathbf{A}_1^2 = (1, A_2)$  and  $\mathbf{B}_1^2 = (0, 0)$  where  $A_2 = 1$  when  $x_{11} = 0$  and  $A_2 = 0$  otherwise.
- Four estimation methods: method (i) our method; method (iii) the parametric method based on the well-known  $G$ -formula; method (iv) the marginal structural model based on inverse probability weighting; method (v) the  $G$ -estimation based on SNMM. Empty cells imply that they are not easily estimable by the method.

effects separately at different times and are similar to methods (i), (iii) and (iv) and (v) in the simulation study of the previous subsection; therefore, we still refer to them as methods (i), (iii), (iv) and (v). Thus, method (i) is constructed using our method described in Section 5.1; method (iii) is constructed using the well-known  $G$ -formula (11) or (12) (Taubman et al. [11]); method (iv) is constructed using the marginal structural model based on inverse probability weighting (Robins [6, 7]); method (v) is constructed using the  $G$ -estimation based on SNMM (Robins [5, 7]).

Our treatment variables are two consecutive drug uses  $Z_1$  and  $Z_2$ , and both are dichotomized. The outcome  $Y$  is the logarithm of CD4 count after  $Z_2$ . We wish to estimate the blip and sequential causal effects. Using method (i), we model the point observable effects and find that the relevant covariates are two earlier CD4 counts,  $X_{01}$  before  $Z_1$  and  $X_{11}$  between  $Z_1$  and  $Z_2$ , and both are dichotomized as high versus low. Given the small sample of this study, we assume a simple SNMM: the blip effects  $\phi(x_{01})$  of  $z_1 = 1$  are the same for all  $x_{01}$ ; the blip effects  $\phi(x_{01}, z_1, x_{11})$  of  $z_2 = 1$  are the same for all  $(x_{01}, z_1, x_{11})$ . Using two covariates  $X_{01}$  and  $X_{11}$  and two treatment variables  $Z_1$  and  $Z_2$ , we readily apply methods (i), (iii), (iv) and (v) to estimate the blip and sequential causal effects.

The analysis is presented in detail in Supplement III of Supplementary Material [13]. The result is presented in Table 2, which shows that method (i) estimates all the blip and sequential causal effects while achieving the smallest variances. Furthermore, method (i) estimates these causal effects via only two point observable effects of  $z_1 = 1$  and  $z_2 = 1$ .

**6. Conclusions.** In the framework of causal inference based on the concept of potential variables, there are two major parts: one part expressing the causal effect in terms of potential variables and another part linking the potential variables, and thus the causal effect to

observed data by a certain treatment assignment mechanism satisfying the identifying condition. Neyman [10] and Rubin [9] and others pointed out that the separation of the two parts clarifies the relationship between causal effects and observed data.

There are also these two parts in sequential causal inference of this article. In the first part, we not only express the blip and sequential causal effects in terms of potential variables in Section 2.2 but also obtain the relationship between these causal effects in Section 3. This relationship turns out to be simple and useful expressions of one another, as described by Propositions 1–4. In the second part, we obtain the new  $G$ -formula, which links the blip and sequential causal effects to observed data via the point observable effect under the identifying condition, as described in Theorems 1 and 2 of Section 4.3. Based on the new  $G$ -formula, we are able to estimate the blip and sequential causal effects via the point observable effects in the framework of the likelihood-based inference, as described in Section 5.1.

Given a relatively simple setting of the simulation and medical example, our method shows certain advantages over the parametric method based on the well-known  $G$ -formula (Taubman et al. [11]) in terms of efficiency and modeling conditions. However, the semiparametric methods, such as the marginal structural model based on inverse probability weighting (Robins [6, 7]) and the  $G$ -estimation based on SNMM (Robins [5, 7]) and their extensions, are able to deal with complex problems under different premises.

To conclude the article, we discuss possible applications of the new  $G$ -formula to three difficult statistical problems. The first problem concerns observational studies with numerous covariates including time-dependent ones. In observational studies of single-point treatment with numerous covariates, the treatment assignment is often approximated by a number of subrandomized trials called subclasses (Rosenbaum and Rubin [8]). In observational studies, the assignments of treatment sequence may be approximated by the first-order Markov process in which the assignment of each treatment depends only on the latest subclasses. Based on the new  $G$ -formula, we believe that the approximate first-order Markov process can be used to estimate the point observable effects and then the blip and sequential causal effects.

The second problem arises when the likelihood of the observed data cannot be specified. In single-point causal inference, semiparametric and nonparametric approaches can be used to estimate the point observable effect (e.g., McCullagh and Nelder [2]). Recall that the new  $G$ -formula is true without specific assumptions for the probability model. Therefore, in sequential causal inference based on the new  $G$ -formula, we believe that semiparametric and nonparametric approaches can be used to estimate the point observable effects and then the blip and sequential causal effects.

The third problem concerns the missing time-dependent covariates between treatments in the sequence. In single-point causal inference, the likelihood-based methods are used, such as the expectation-maximization algorithm or multiple imputation method, to deal with the missing covariates when estimating the point observable effect (Little and Rubin [1]). In sequential causal inference based on the new  $G$ -formula, the missing time-dependent covariates are only the problems for relevant treatments. We believe that the likelihood-based methods can be used to address the problems in estimating the point observable effects of relevant treatments.

## APPENDIX

**A.1. Proof of formula (4).** For notational simplicity, we only prove (4) when  $t = 1$ , that is,

$$(23) \quad E\{Y(\mathbf{G}_1^T)\} = E\{Y(\mathbf{D}_1^T = \mathbf{0})\} + \sum_{s=1}^T E\{\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\},$$

where the expectation for  $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)$  is with respect to  $\mathbf{P}^{\mathbf{G}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ . However, interested readers may readily extend the proof to the case of  $t > 1$ .

Let  $Q_s = E\{Y(\mathbf{G}_1^s, \mathbf{D}_{s+1}^T = \mathbf{0})\}$  for  $T \geq s \geq 0$ . Noticeably,  $Q_T = E\{Y(\mathbf{G}_1^T)\}$  and  $Q_0 = E\{Y(\mathbf{D}_1^T = \mathbf{0})\}$ . Then we write  $Q_T$  as the following telescoping sum:

$$Q_T = Q_T - Q_{T-1} + Q_{T-1} - \cdots + Q_s - Q_{s-1} + Q_{s-1} - \cdots + Q_1 - Q_0 + Q_0.$$

By rewriting

$$\begin{aligned} Q_s &= E[E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\}], \\ Q_{s-1} &= E[E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}], \end{aligned}$$

we obtain

$$\begin{aligned} Q_s - Q_{s-1} &= E[E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\} \\ &\quad - E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}], \end{aligned}$$

where the outer expectation is with respect to  $\mathbf{P}^{\mathbf{G}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ .

We rewrite (2) at  $t = s$  as

$$\begin{aligned} \phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \\ &= E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\} - E\{Y(\mathbf{D}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}. \end{aligned}$$

Inserting this expression into  $Q_s - Q_{s-1}$ , we obtain

$$Q_s - Q_{s-1} = E\{\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}.$$

Now, by inserting these  $Q_s - Q_{s-1}$  ( $s = T, \dots, 1$ ) and  $Q_0 = E\{Y(\mathbf{D}_1^T = \mathbf{0})\}$  into  $Q_T$  above, we obtain formula (23), which is (4) when  $t = 1$ .

**A.2. Proof of formula (7).** For notational simplicity, we only prove (7) when  $t = 1$ , that is,

$$\begin{aligned} E\{Y(\mathbf{D}_1^T)\} &= E\{Y(D_1 = 0, \mathbf{R}_2^T)\} + \theta(z_1) \\ (24) \quad &+ \sum_{s=2}^T E_1\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} - \sum_{s=2}^T E_2\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}, \end{aligned}$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$  for  $\mathbf{G}_1^s = \mathbf{D}_1^s$  and  $E_2(\cdot)$  for  $\mathbf{G}_1^s = (\mathbf{D}_1^{s-1}, R_s)$ . However, interested readers may readily extend the proof to the case of  $t > 1$ . Let  $W_s = E\{Y(\mathbf{D}_1^s, \mathbf{R}_{s+1}^T)\}$  for  $T \geq s \geq 1$ . Then we write  $W_T = E\{Y(\mathbf{D}_1^T)\}$  as the following telescoping sum:

$$W_T = W_T - W_{T-1} + W_{T-1} - \cdots + W_s - W_{s-1} + W_{s-1} - \cdots + W_2 - W_1 + W_1.$$

We rewrite  $W_s$  as

$$\begin{aligned} W_s &= E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\}] \\ &= E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\} - E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}] \\ &\quad + E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}]. \end{aligned}$$

We rewrite (3) at  $t = s$  as

$$\begin{aligned} \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s) \\ &= E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\} - E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}. \end{aligned}$$

Inserting this into  $W_s$ , we obtain

$$W_s = E_1\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} + E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}],$$

where  $E_1(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{D}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ .

Then we rewrite  $W_{s-1} = E\{Y(\mathbf{D}_1^{s-1}, R_s, \mathbf{R}_{s+1}^T)\}$  as

$$W_{s-1} = E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s\}]$$

and follow the same procedure above to obtain

$$W_{s-1} = E_2\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} + E[E\{Y(\mathbf{R}_{s+1}^T) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0\}],$$

where  $E_2(\cdot)$  is an expectation with respect to  $\mathbf{P}^{\mathbf{G}_1^s}(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$  for  $\mathbf{G}_1^s = (\mathbf{D}_1^{s-1}, R_s)$ .

Therefore, we have

$$W_s - W_{s-1} = E_1\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\} - E_2\{\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}; z_s)\}.$$

We also rewrite  $W_1 = E\{Y(D_1, \mathbf{R}_2^T)\}$  as

$$\begin{aligned} W_1 &= E\{Y(D_1, \mathbf{R}_2^T)\} - E\{Y(D_1 = 0, \mathbf{R}_2^T)\} + E\{Y(D_1 = 0, \mathbf{R}_2^T)\} \\ &= \theta(z_1) + E\{Y(D_1 = 0, \mathbf{R}_2^T)\}, \end{aligned}$$

where  $z_1$  is the treatment assigned by  $D_1$ . Now, by inserting these  $W_s - W_{s-1}$  ( $s = T, \dots, 2$ ) and  $W_1$  into  $W_T$ , we obtain formula (24), which is (7) when  $t = 1$ .

**Acknowledgments and conflicts of interest.** The authors of the article wish to thank one Associate Editor, two anonymous reviewers for their comments and advice, which were tremendously helpful and have greatly improved the article. This research received no specific grant from any funding agency in the public, commercial or nonprofit sectors.

## SUPPLEMENTARY MATERIAL

**Supplement to “New  $G$ -formula for the sequential causal effect and blip effect of treatment in sequential causal inference”** (DOI: 10.1214/18-AOS1795SUPP; .zip). The material includes the following three sections:

Supplement I: Proofs for formulas (10a), (10b), (10c), (21), (22).

Supplement II: Description of the simulation study in Section 5.3.

Supplement III: Description of the medical example in Section 5.4.

In addition, the material contains the SAS code for Supplement II as well as the SAS code and data for Supplement III.

## REFERENCES

- [1] LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1925014 <https://doi.org/10.1002/9781119013563>
- [2] MCCULLAGH, P. and NELDER, J. A. (1999). *Generalized Linear Models*. Chapman & Hall/CRC, Boca Raton, FL.
- [3] PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Stat. Methods Med. Res.* **21** 31–54. MR2867537 <https://doi.org/10.1177/0962280210386207>
- [4] ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. MR0877758 [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6)

- [5] ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (Los Angeles, CA, 1994) (M. Berkane, ed.). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 [https://doi.org/10.1007/978-1-4612-1842-5\\_4](https://doi.org/10.1007/978-1-4612-1842-5_4)
- [6] ROBINS, J. M. (1999). Association, causation, and marginal structural models. *Synthese* **121** 151–179. MR1766776 <https://doi.org/10.1023/A:1005285815569>
- [7] ROBINS, J. M. (2009). Longitudinal data analysis. In *Handbooks of Modern Statistical Methods* (G. Fitzmaurice, ed.) 553–599. Chapman & Hall/CRC, Boca Raton, FL.
- [8] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 <https://doi.org/10.1093/biomet/70.1.41>
- [9] RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 <https://doi.org/10.1198/016214504000001880>
- [10] SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. Dąbrowska and T. P. Speed. MR1092986
- [11] TAUBMAN, S. L., ROBINS, J. M., MITTLEMAN, M. A. and HERNÁN, M. A. (2009). Intervening on risk factors for coronary heart disease: An application of the parametric g-formula. *Int. J. Epidemiol.* **38** 1599–1611.
- [12] WANG, X. and YIN, L. (2015). Identifying and estimating net effects of treatments in sequential causal inference. *Electron. J. Stat.* **9** 1608–1643. MR3379004 <https://doi.org/10.1214/15-EJS1046>
- [13] WANG, X. and YIN, L. (2020). Supplement to “New G-formula for the sequential causal effect and blip effect of treatment in sequential causal inference.” <https://doi.org/10.1214/18-AOS1795SUPP>.
- [14] ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.