Electronic Journal of Statistics
Vol. 9 (2015) 1608–1643
ISSN: 1935-7524
DOI: 10.1214/15-EJS1046

Identifying and estimating net effects of treatments in sequential causal inference

Xiaoqin Wang

Department of Electronics, Mathematics and Natural Sciences
University of Gävle
SE 801 76, Gävle
Sweden
e-mail: xwg@hig.se

and

Li Yin

Department of Medical Epidemiology and Biostatistics
Karolinska Institute
Box 281, SE 171 77, Stockholm
Sweden
e-mail: Li.Yin@ki.se

Abstract: Suppose that a sequence of treatments are assigned to influence an outcome of interest that occurs after the last treatment. Between treatments, there are time-dependent covariates that may be post-treatment variables of the earlier treatments and confounders of the subsequent treatments. In this article, we study identification and estimation of the net effect of each treatment in the treatment sequence. We construct a point parametrization for the joint distribution of treatments, time-dependent covariates and the outcome, in which the point parameters of interest are the point effects of treatments considered as single-point treatments. We identify net effects of treatments by their expressions in terms of point effects of treatments and express patterns of net effects of treatments by constraints on point effects of treatments. We estimate net effects of treatments through their point effects under the constraint by maximum likelihood and reduce the number of point parameters in the estimation by the treatment assignment condition. As a result, we obtain an unbiased consistent maximum-likelihood estimate for the net effect of treatment even in a long treatment sequence. We also show by simulation that the interval estimation of the net effect of treatment achieves the nominal coverage probability.

MSC 2010 subject classifications: Primary 62H12; secondary 62H15, 62F03, 62F30.

Keywords and phrases: Net effect of treatment, pattern of net effects of treatments, point effect of treatment, constraint on point effects of treatments, treatment assignment condition, sequential causal inference.

Received August 2014.

Contents

1. Introduction

In many economic and medical practices, a sequence of treatments are assigned to influence an outcome of interest that occurs after the last treatment of the sequence. Between consecutive treatments, time-dependent covariates are present that may be post-treatment variables of the earlier treatments (Rosenbaum [16], Robins [8], Frangakis & Rubin [1]) and confounders of the subsequent treatments. In this article, we aim at the net effect of each treatment in the sequence, i.e. the causal effect of the treatment on the subpopulation defined by the previous treatments and time-dependent covariates while setting the subsequent treatments at controls. The net effect of treatment is also called the blip effect in the context of semi-parametric sequential causal inference (Robins [9, 10, 14, 15]).

By using the G-computation algorithm formula (also called the G-formula), Robins [7, 10, 12, 14, 15] identified the net effect of treatment by standard parameters, which are usually the means of the outcome in subpopulations defined by all the treatments and time-dependent covariates. Robins illustrated that any constraint imposing equalities among standard parameters may lead to automatic rejection of the null hypothesis of net effects if the time-dependent covariate is a post-treatment variable of the earlier treatments and a confounder of the subsequent treatments. As the treatment sequence gets long, the number of standard parameters becomes huge, and with no constraint on these parameters, the maximum-likelihood (ML) estimates of net effects may not be consistent (Robins & Ritov [11]). To obtain non-genuine likelihood-based estimates of net effects, two semi-parametric approaches have been developed: the structural nested model (Robins [9, 10, 14, 15], Robins et al. [13], Murphy [6], Henderson et al. [4]) and the marginal structural model (Robins [12, 15], Murphy et al. [5]).

On the other hand, every treatment in the sequence can be considered as a single-point treatment. In the framework of single-point causal inference (Rosenbaum & Rubin [17], Rosenbaum [18], Rubin [19]), a treatment has a point causal effect on a subpopulation defined by the previous treatments and time-dependent covariates. The point causal effect of the treatment can be identified by the point effect of the treatment, which corresponds to differences between the means of the outcome in the sub-subpopulations assigned with different treatments. In this case, the time-dependent covariates act only as confounders. As a result, a constraint on point effects does not necessarily lead to automatic rejection of the null hypothesis of point causal effects.

Intuitively, the point causal effect of each treatment in the sequence results from the net effects of this and all the subsequent treatments. If one can identify the point causal effects by the point effects, then one should be able to identify the net effects by the point effects. Because a constraint on point effects does not necessarily lead to automatic rejection of the null hypothesis of point causal effects, it should not lead to automatic rejection of the null hypothesis of net effects. The purpose of this article is to identify and estimate the net effects through their point effects in line of the above observation.

In Section 2, we introduce the background and notation of the identification and estimation of the net effects of treatments. In Section 3, we construct a point parametrization for the joint distribution of treatments, time-dependent covariates and an outcome after the last treatment, in which the point parameters of interest are the point effects of treatments. In Section 4, we identify net effects by their point effects, express patterns of net effects by constraints on point effects and analyze the role of treatment assignment conditions in the identification. In Section 5, we estimate net effects through their point effects under the constraint and reduce the number of point parameters in the estimation by the treatment assignment condition. In Section 6, we illustrate our method via an analytical example, a simulation study and a real example. In Section 7, we conclude the article with remarks and discussion.

2. Background and notation

2.1. Treatment sequence, potential covariates and potential outcome

Let z_t indicate the treatments at time t $(t = 1, \ldots, T)$. Assume that all z_t are discrete variables and have the values $0, 1, \ldots$. We take $z_t = 0$ as the control treatment and $z_t = 1, 2, \ldots$ as active treatments. Let the set $\mathbf{z}_1^t = (z_1, \ldots, z_t)$ indicate the treatment sequences from times 1 to t. Suppose that every treatment sequence $\mathbf{z}_1^T$ could be applied to each unit of a population. Assume that there is no interference between units and no represented treatment sequence 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.

In the framework of sequential causal inference, each unit could have a potential (time-dependent) covariate vector $\mathbf{x}_t(\mathbf{z}_1^t)$ between treatments z_t and z_{t+1} and a potential outcome $y(\mathbf{z}_1^T)$ of our interest after the last treatment z_T under treatment sequence $\mathbf{z}_1^T$. Assume that $\mathbf{x}_t(\mathbf{z}_1^t)$ is a discrete vector with nonnegative components. We take $\mathbf{x}_t(\mathbf{z}_1^t) = \mathbf{0}$ as the reference level. Let $\mathbf{x}_1^t(\mathbf{z}_1^t) = \{\mathbf{x}_1(z_1), \mathbf{x}_2(\mathbf{z}_1^2), \ldots, \mathbf{x}_t(\mathbf{z}_1^t)\}$ be the potential covariate array between treatments z_1 and z_{t+1}.

In the above definition of the potential covariates and outcome, every z_t in $\mathbf{z}_1^T$ is a deterministic function of the earlier treatments and potential covariates, i.e. $z_t = z_t\{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}(\mathbf{z}_1^{t-1})\}$. If each z_t in $\mathbf{z}_1^T$ does not depend on the earlier treatments and potential covariates, then $\mathbf{z}_1^T$ is a static treatment sequence, and otherwise, it is a dynamic treatment sequence.

2.2. Net effects of treatments

In most practical cases, treatments z_t $(t = 1, \ldots, T)$ are consecutively and randomly assigned, so the potential covariate vectors $\mathbf{x}_t(\mathbf{z}_1^t)$ $(t = 1, \ldots, T-1)$ and the potential outcome $y(\mathbf{z}_1^T)$ become consecutively and randomly observable. Denote the observable covariate vector simply by $\mathbf{x}_t$ $(t = 1, \ldots, T-1)$ and the

observable outcome by y. Let $\mathbf{x}_1^t = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$. The assigned treatments $\mathbf{z}_1^{t-1}$ are assumed to be a treatment sequence, which can be static or dynamic and leads to the observable covariate array $\mathbf{x}_1^{t-1}$. This assumption is known as the consistency assumption (Robins [7, 8, 9, 10, 12, 14, 15]). Let $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ be a set of observable variables in which the treatment sequence $\mathbf{z}_1^{t-1}$ leads to the observable covariate array $\mathbf{x}_1^{t-1}$.

Let $\mathbf{z}_t^T = (z_t, \ldots, z_T)$ be the treatment sequence given the observable variables $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. Under $\mathbf{z}_t^T$ given $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$, each unit could have potential covariate vectors $\mathbf{x}_t(\mathbf{z}_1^t), \ldots, \mathbf{x}_{T-1}(\mathbf{z}_1^{T-1})$ and a potential outcome $y(\mathbf{z}_1^T)$. Let $\mathbf{x}_t^{T-1}(\mathbf{z}_t^{T-1}) = \{\mathbf{x}_t(\mathbf{z}_1^t), \ldots, \mathbf{x}_{T-1}(\mathbf{z}_1^{T-1})\}$ and $y(\mathbf{z}_t^T) = y(\mathbf{z}_1^T)$ for a given $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$.

Robins [9, 10, 14, 15] defined the net effect of $z_t > 0$ given the observable variable $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ as

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

for $t = 1, \ldots, T$. Here the expectation $E(b \mid a)$ is with respect to the conditional distribution of b given a in the population.

Throughout the article, we adopt the following notational conventions. First, the notations $\mathbf{z}_u^v$, $\mathbf{x}_u^v$ and $\mathbf{x}_u^v(\mathbf{z}_u^v)$ with $u > v$ or $u = v = 0$ should be omitted from relevant expressions. For instance, the notations $\mathbf{z}_1^0$ and $\mathbf{x}_1^0$ in (1) for $t = 1$ should be omitted, and so (1) becomes

$$\phi(z_1) = E\{y(z_1, \mathbf{z}_2^T = \mathbf{0})\} - E\{y(\mathbf{z}_1^T = \mathbf{0})\}.$$

Similarly, the notation $z_{T+1}^T = \mathbf{0}$ in (1) for $t = T$ should be omitted, and so (1) becomes

$$\phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = E\{y(z_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} - E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\}.$$

Second, the sigma notation $\sum_{i=u}^v a_i$ with $v < u$ should be omitted from relevant expressions. Third, the notations $\mathbf{z}_u^v$, $\mathbf{x}_u^v$, $\mathbf{x}_u^v(\mathbf{z}_u^v)$ and $\sum_{i=u}^v a_i$ with $u < 1$ and $v \geq 1$ are treated as $\mathbf{z}_1^v$, $\mathbf{x}_1^v$, $\mathbf{x}_1^v(\mathbf{z}_1^v)$ and $\sum_{i=1}^v a_i$. Fourth, the notation $(\mathbf{z}_u^v, \mathbf{x}_u^{v-1})$ is equal to $(\mathbf{z}_u^{v-1}, \mathbf{x}_u^{v-1}, z_v)$, and $(\mathbf{z}_u^v, \mathbf{x}_u^v)$ to $(\mathbf{z}_u^v, \mathbf{x}_u^{v-1}, \mathbf{x}_v)$; we may use one or another notation in different contexts.

2.3. G-formula for net effects of treatments

In the following assumption, let z_t^* indicate the treatments randomly assigned at time t. We assume that the assignment of z_t^* $(t = 1, \ldots, T)$ satisfies

$$\begin{cases} \mathbf{x}_t^{T-1}(\mathbf{z}_t^{T-1}), y(\mathbf{z}_t^T) \perp z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1} \\ 0 < \mathrm{pr}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) < 1 \end{cases} \tag{2}$$

for any treatment sequence $\mathbf{z}_t^T$ given the observable variable $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. Here $A \perp B \mid C$ means that A is conditionally independent of B given C. The z_t is a specified treatment at time t in the specified treatment sequence $\mathbf{z}_t^T$.

The first part of (2) is known as the assumption of no unmeasured confounders (Robins [7, 8, 9, 10, 12, 14, 15]). The second part is known as the positivity assumption. There may exist other observable covariates than $\mathbf{x}_1^{T-1}$ but the assignment of z_t^* does not depend on them and no further information is available about them.

Standard parameters for the conditional distribution of y given $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ are the means $E(y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ denoted by $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. Standard parameters for the conditional distribution of the discrete variable $\mathbf{x}_t$ given $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ are the probabilities $\text{pr}(\mathbf{x}_t \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$. Using assumption (2), Robins (1986, 1997) derived the G-formula

$$E\{y(\mathbf{z}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} = \sum_{\mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \prod_{s=t}^{T-1} \text{pr}(\mathbf{x}_s \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) \tag{3}$$

for $t = 1, \ldots, T-1$ and $E\{y(z_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} = \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$.

Combining (1) and (3), we obtain the G-formula for the net effect

$$\begin{aligned}
&\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\quad = \sum_{\mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^t, \mathbf{z}_{t+1}^T = \mathbf{0}, \mathbf{x}_1^{T-1}) \prod_{s=t}^{T-1} \text{pr}(\mathbf{x}_s \mid \mathbf{z}_1^t, \mathbf{z}_{t+1}^s = \mathbf{0}, \mathbf{x}_1^{s-1}) \\
&\qquad - \sum_{\mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^{t-1}, \mathbf{z}_t^T = \mathbf{0}, \mathbf{x}_1^{T-1}) \prod_{s=t}^{T-1} \text{pr}(\mathbf{x}_s \mid \mathbf{z}_1^{t-1}, \mathbf{z}_t^s = \mathbf{0}, \mathbf{x}_1^{s-1})
\end{aligned} \tag{4}$$

for $t = 1, \ldots, T-1$ and

$$\phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) - \mu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T = 0),$$

which identifies the net effect $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ by the standard parameters $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ and $\text{pr}(\mathbf{x}_s \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1})$ $(s = t, \ldots, T-1)$ under assumption (2).

If $\mathbf{x}_t$ $(t = 1, \ldots, T-1)$ are post-treatment variables of z_s $(s \leq t)$, then the standard parameters essentially do not have any pattern (Rosenbaum [16], Robins [8], Frangakis & Rubin [1]). If $\mathbf{x}_t$ are simultaneously confounders of z_s $(s > t)$, then one needs to use all these standard parameters to identify $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$. With a long treatment sequence, the number of these parameters is huge. Without a constraint on standard parameters, the ML estimates of $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ may not be consistent (Robins [7, 10, 12, 14, 15], Robins & Ritov [11]).

In this article, instead of standard parameters, we are going to use point effects of treatments to identify the net effects of treatments under assumption (2).

3. Point parametrization for the joint distribution of observable variables

3.1. Joint distribution of observable variables

Instead of one set $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}, y)$ of the observable variables, we consider N independent and identically distributed sets, $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i\}$, $i = 1, \ldots, N$. The joint distribution of $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{t-1}, y_i\}_{i=1}^N$ can be factorized into

$$f(\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i\}_{i=1}^N) \tag{5}$$

$$= \prod_{i=1}^N f(y_i \mid \mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}) \tag{6}$$

$$\prod_{t=1}^{T-1} \prod_{i=1}^N f(\mathbf{x}_{it} \mid \mathbf{z}_{i1}^t, \mathbf{x}_{i1}^{t-1}) \tag{7}$$

$$\prod_{t=1}^{T} \prod_{i=1}^N f(z_{it} \mid \mathbf{z}_{i1}^{t-1}, \mathbf{x}_{i1}^{t-1}). \tag{8}$$

Here $f(u \mid v)$ is a conditional probability distribution of u given v if u is discrete, and a conditional density distribution of u given v if u is continuous. We are going to construct point parameterizations for (6) in Section 3.2 and for (7), (8) and then (5) in Section 3.3.

3.2. Point parametrization for the conditional outcome distribution (6)

Given N sets $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, a stratum is a set of those sets satisfying certain conditions. For instance, a stratum $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ is a set of those sets satisfying $(\mathbf{z}_{i1}^t, \mathbf{x}_{i1}^{t-1}) = (\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$. Let $\text{prop}(A)$ denote the proportion of stratum A in the N sets and $\text{prop}(A \mid B)$ denote the conditional proportion of stratum A in stratum B.

Consider the mean of y in stratum $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$

$$\nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) = \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{prop}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$$

for $t = 1, \ldots, T-1$ and $\nu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. The point effect of treatment $z_t > 0$ on stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ is then

$$\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0). \tag{9}$$

Consider the mean of y in stratum $(\mathbf{z}_1^t, \mathbf{x}_1^t)$

$$\nu(\mathbf{z}_1^t, \mathbf{x}_1^t) = \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{prop}(\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^t)$$

for $t = 1, \ldots, T-1$. The point effect of covariate $\mathbf{x}_t > \mathbf{0}$ on stratum $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ is then

$$\zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t) = \nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t) - \nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t = \mathbf{0}). \tag{10}$$

Let

$$\nu = \sum_{\mathbf{z}_1^T, \mathbf{x}_1^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{prop}(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \tag{11}$$

which is the grand mean of y in all the N sets $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$. Given $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, the proportions of treatments and covariates can be treated as constants. Therefore the point effects of treatments, the point effects of covariates and the grand mean are linear functions of the standard parameters $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ and thus are parameters of (6) which are called point parameters.

According to (9), (10) and (11), each point parameter can be expressed in terms of the standard parameters $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. Conversely, we show in Appendix A.1 that each standard parameter can be expressed in terms of the point parameters by

$$\begin{aligned}
&\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \\
&\quad = \sum_{t=1}^{T} \left\{ \sum_{z_t^* > 0} -\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \text{prop}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \right\} \\
&\qquad + \sum_{t=1}^{T-1} \left\{ \sum_{\mathbf{x}_t^* > \mathbf{0}} -\zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t^*) \text{prop}(\mathbf{x}_t^* \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}) + \zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t) \right\} + \nu
\end{aligned} \tag{12}$$

where we take $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0) = 0$ and $\zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t = \mathbf{0}) = 0$.

Let Ψ_y be the set of all the point parameters, i.e. $\Psi_y = \{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t), t = 1, \ldots, T; \zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t), t = 1, \ldots, T-1; \nu\}$. Then Ψ_y forms a parametrization for (6).

Definition 1. The set Ψ_y is called a point parametrization for the conditional outcome distribution (6) of $\{y_i\}_{i=1}^N$ given $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$.

3.3. Point parametrization for the joint distribution (5)

Treating $\mathbf{x}_t$ as an outcome conditional on the observable variable $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ and repeating the procedure of the previous subsection, we construct point parameters and the point parametrization for the conditional distribution $\prod_{i=1}^N f(\mathbf{x}_{it} \mid \mathbf{z}_{i1}^t, \mathbf{x}_{i1}^{t-1})$ in (7). Denote the point parametrization by $\Psi_{\mathbf{x}_t}$ $(t = 1, \ldots, T-1)$.

In observational studies, treatments z_t $(t = 1, \ldots, T)$ can be exposures, whose assignments are unknown and need to be estimated. Treating z_t as an outcome conditional on the observable variable $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ and repeating the procedure of the previous subsection, we construct point parameters and the point parametrization for the conditional distribution $\prod_{i=1}^N f(z_{it} \mid \mathbf{z}_{i1}^{t-1}, \mathbf{x}_{i1}^{t-1})$ in (8). Denoted the point parametrization by Ψ_{z_t}.

Let $\Psi = \{\Psi_y; \Psi_{z_t}, t = 1, \ldots, T; \Psi_{\mathbf{x}_t}, t = 1, \ldots, T-1\}$. Then Ψ forms a parametrization for the joint distribution (5).

Definition 2. The set Ψ is called a point parametrization for the joint distribution (5) of $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i\}_{i=1}^N$.

4. Identification of net effects of treatments by point effects of treatments

4.1. Limits of point parameters of Ψ_y

As described in Section 3, the point parameters of Ψ_y are conditional on N sets $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$ of treatments and covariates. Given different $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, the proportions of treatments and covariates are different and therefore these parameters may be different. On the other hand, the limits of these parameters as N approaches infinity exist and are properties of the population.

As N approaches infinity, the proportion $\text{prop}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ converges to the probability $\text{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$, and $\nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ to

$$\mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) = \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}),$$

which is the mean of y in subpopulation $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$. Formula (9) implies that $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ converges to

$$\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0), \tag{13}$$

which is the point effect of treatment $z_t > 0$ on subpopulation $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$.

As N approaches infinity, $\text{prop}(\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^t)$ converges to $\text{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^t)$, and $\nu(\mathbf{z}_1^t, \mathbf{x}_1^t)$ to

$$\mu(\mathbf{z}_1^t, \mathbf{x}_1^t) = \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_{t+1}^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^t).$$

Formula (10) implies that $\zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t)$ converges to

$$\xi(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t) = \mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t) - \mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t = \mathbf{0}), \tag{14}$$

which is the point effect of covariate $\mathbf{x}_t > \mathbf{0}$ on subpopulation $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$. Formula (11) implies that ν converges to

$$\mu = \sum_{\mathbf{z}_1^T, \mathbf{x}_1^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \text{pr}(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}), \tag{15}$$

which is the grand mean in the population.

Letting N approach infinity in (12), we obtain

$$\begin{aligned}
&\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \\
&\quad = \sum_{t=1}^{T} \left\{ \sum_{z_t^*>0} -\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \mathrm{pr}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \right\} \\
&\qquad + \sum_{t=1}^{T-1} \left\{ \sum_{\mathbf{x}_t^*>\mathbf{0}} -\xi(\mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}, \mathbf{x}_t^*) \mathrm{pr}(\mathbf{x}_t^* \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) + \xi(\mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}, \mathbf{x}_t) \right\} + \mu.
\end{aligned} \tag{16}$$

We are going to show that we only need the point effects $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ to identify the net effects $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$.

4.2. Net effects of treatments versus point effects of treatments

Inserting (16) into (3), we derive the following formula in Appendix A.2

$$\begin{aligned}
&E\{y(\mathbf{z}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\
&= \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \\
&\quad + \sum_{z_t^*>0} -\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \mathrm{pr}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\quad + \sum_{s=t+1}^{T} \sum_{\mathbf{x}_t^{s-1}} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\} \\
&\qquad \prod_{k=t}^{s-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^{k}, \mathbf{x}_1^{k-1})
\end{aligned} \tag{17}$$

for $t = 1, \ldots, T-1$ and

$$\begin{aligned}
&E\{y(\mathbf{z}_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} \\
&\quad = \mu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) \\
&\qquad + \sum_{z_T^*>0} -\theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T^*) \mathrm{pr}(z_T^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) + \theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T).
\end{aligned}$$

Inserting (17) into (1), using $\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s = 0) = 0$ and $\sum_{\mathbf{x}_t} \mathrm{pr}(\mathbf{x}_t \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) = 1$, and then dropping the unnecessary superscript *, we express each net effect $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ in terms of the point effects of treatments and the probabilities of covariates and treatments by

$$\begin{aligned}
&\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\quad = \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - [Q(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - Q(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)]
\end{aligned} \tag{18}$$

for $t =, \ldots, T-1$ and

$$\phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T),$$

where

$$\begin{aligned} &Q(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\ &\quad = \sum_{s=t+1}^{T} \sum_{\mathbf{x}_t^{s-1}} \sum_{z_s>0} \theta(\mathbf{z}_1^t, \mathbf{z}_{t+1}^{s-1} = \mathbf{0}, \mathbf{x}_1^{s-1}, z_s) \mathrm{pr}(z_s \mid \mathbf{z}_1^t, \mathbf{z}_{t+1}^{s-1} = \mathbf{0}, \mathbf{x}_1^{s-1}) \\ &\quad\quad \prod_{k=t}^{s-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^t, \mathbf{z}_{t+1}^k = \mathbf{0}, \mathbf{x}_1^{k-1}). \end{aligned}$$

Formula (18) shows that each net effect $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ can be expressed in terms of the point effects of treatments and the probabilities of covariates and treatments.

In Appendix A.3, we derive the following formula, under assumption (2),

$$\begin{aligned} &\mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \\ &\quad = E\{y(\mathbf{z}_t^T = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} + \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\ &\quad\quad + \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s>0} \phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \mathrm{pr}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \end{aligned} \tag{19}$$

for $t = 1, \ldots, T-1$ and

$$\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} + \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T).$$

Formula (19) implies that the mean $\mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ arises from the net effects of active treatments $z_s > 0$ at times $s \geq t$ on substrata $(\mathbf{z}_1^s, \mathbf{x}_1^{s-1})$ in stratum $(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})$. Formula (19) can also be derived from formula (8.3) of Robins [10].

Inserting (19) into (13), we obtain

$$\begin{aligned} &\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\ &\quad = \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) + R(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) - R(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0) \end{aligned} \tag{20}$$

for $t = 1, \ldots, T-1$ and

$$\theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T),$$

where

$$\begin{aligned} &R(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\ &\quad = \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s>0} \phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \mathrm{pr}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}). \end{aligned}$$

Formula (20) shows that each point effect $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ can be expressed in terms of the net effects of treatments and the probabilities of covariates and treatments. The formula decomposes $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ into the net effects of treatments $z_s > 0$ at times $s \geq t$ in subpopulation $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ versus subpopulation $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)$.

In summary, we have the following identification theorem:

Theorem 1. *Under assumption (2), for the given probabilities of covariates and treatments, the point effects* $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ $(t = 1, \ldots, T)$ *are a one-to-one linear transformation of the net effects* $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ $(s = 1, \ldots, T)$ *according to (18) and (20).*

Remark 1. The identification of $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$ in this theorem does not involve the point effects $\xi(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t)$ $(t = 1, \ldots, T-1)$ and the grand mean μ. Although the G-formula (4) also identifies $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$, the identifying standard parameters $\mu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T)$ are not a one-to-one transformation of $\phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s)$.

4.3. Patterns of net effects of treatments versus constraints on point effects of treatments

Suppose that the data-generating mechanism is such that the net effects $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ follow a certain pattern. One example of such patterns is

$$\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \varphi_1 z_t + \varphi_2 z_{t-1} + \varphi_3 \mathbf{x}'_{t-1}$$

for any $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ at $t = 1, \ldots, T$, where the parameter vector $\varphi = (\varphi_1, \varphi_2, \varphi_3)$ indexes all the net effects. Generally, we consider a pattern of the net effects described by a function

$$\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi) \tag{21}$$

where the k-dimensional parameter vector $\varphi = (\varphi_1, \ldots, \varphi_k)$ indexes all the net effects. We call φ the net effect vector.

Inserting (21) into (20), we obtain a constraint on the point effects $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$.

Corollary 1. *For the given probabilities of covariates and treatments, under pattern (21), all the point effects* $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ *are indexed by the net effect vector* φ *and given by*

$$\begin{aligned}
&\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\quad = \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi) \\
&\qquad + \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s > 0} \phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s; \varphi) \mathrm{pr}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})
\end{aligned}$$

$$- \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s > 0} \phi(\mathbf{z}_1^{t-1}, z_t = 0, \mathbf{z}_{t+1}^{s-1}, \mathbf{x}_1^{s-1}, z_s; \varphi)$$
$$\text{pr}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, z_t = 0, \mathbf{x}_1^{t-1}) \tag{22}$$

for $t = 1, \ldots, T-1$ *and* $\theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T; \varphi)$.

Despite pattern (21), the covariate $\mathbf{x}_t$ can still be a post-treatment variable of $\mathbf{z}_1^t$ (Robins [7, 10, 12, 14, 15]). In this case, the standard parameter $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ does not have any pattern for treatments z_t $(t < T)$ (Rosenbaum [16], Robins [8], Frangakis & Rubin [1]). If $\mathbf{x}_t$ is additionally a confounder of $\mathbf{z}_{t+1}^T$, then one needs all the standard parameters to identify the net effect.

Remark 2. Generally, pattern (21) does not simplify the *G*-formula (4).

4.4. Role of treatment assignment conditions in the identification of net effects of treatments

In the framework of single-point causal inference, it is well known that treatment assignment conditions may reduce the number of parameters in the identification of the causal effect of a single-point treatment (Rosenbaum & Rubin [17], Rosenbaum [18], Rubin [19]). In randomized trials, one does not need to use covariates to identify the single-point causal effect. In observational studies even with a large number of covariates, one can approximate the treatment assignment by several sub-randomized trials called subclasses, and one only needs to use the subclasses to identify the single-point causal effect.

In sequential randomized trials, the covariate vector $\mathbf{x}_t$ typically has a small dimension, but the array $\mathbf{x}_1^{t-1}$ has a huge dimension at large t. In observational sequential studies, the vector $\mathbf{x}_t$ can easily have a dimension of several dozens. On the other hand, every treatment in the treatment sequence can be treated as a single-point treatment. We may use assignment conditions of individual treatments z_t to reduce the number of point parameters in constraint (22). For illustration, we consider the Markov process, a common assignment mechanism of a treatment sequence, in which the assignment of z_t only depends on a limited history of previous treatments and covariates, for instance, $\mathbf{x}_{t-1}$. In this case, the following proportion equality is satisfied, at least approximately:

$$\text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t) = \text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}). \tag{23}$$

Assuming equality (23), given $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, the mean of y in stratum $(\mathbf{x}_{t-1}, z_t)$ is

$$\begin{aligned}\nu(\mathbf{x}_{t-1}, z_t) &= \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t) \\ &= \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}).\end{aligned}$$

Taking the average on both sides of (9) with respect to $\text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1})$, we then obtain the following point effect of treatment $z_t > 0$ on stratum $\mathbf{x}_{t-1}$

$$\begin{aligned}
&\vartheta(\mathbf{x}_{t-1}, z_t) \\
&\quad = \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}) \\
&\quad = \nu(\mathbf{x}_{t-1}, z_t) - \nu(\mathbf{x}_{t-1}, z_t = 0). \qquad (24)
\end{aligned}$$

As N approaches infinity, $\text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t)$ converges to $\text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t)$ and $\text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1})$ to $\text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1})$. Then (23) converges to the following probability equality:

$$\text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t) = \text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}).$$

Hence $\nu(\mathbf{x}_{t-1}, z_t)$ converges to

$$\begin{aligned}
\mu(\mathbf{x}_{t-1}, z_t) &= \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \mu(\mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) \text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t) \\
&= \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \mu(\mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) \text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}),
\end{aligned}$$

which is the mean of y in subpopulation $(\mathbf{x}_{t-1}, z_t)$, and $\vartheta(\mathbf{x}_{t-1}, z_t)$ to

$$\begin{aligned}
\theta(\mathbf{x}_{t-1}, z_t) &= \sum_{\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2}} \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}) \\
&= \mu(\mathbf{x}_{t-1}, z_t) - \mu(\mathbf{x}_{t-1}, z_t = 0),
\end{aligned}$$

which is the point effect of $z_t > 0$ on subpopulation $\mathbf{x}_{t-1}$.

Now we consider a constraint on $\theta(\mathbf{x}_{t-1}, z_t)$. For illustration, suppose the following pattern of the net effects

$$\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \phi(\mathbf{x}_{t-1}, z_t; \varphi) \qquad (25)$$

for $t = 1, \ldots, T$. Firstly inserting (25) into (20), secondly taking the average on both sides of (20) with respect to $\text{pr}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1})$ and noticing the probability equality above, and finally using the equality

$$\sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-2}} \text{pr}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid z_t, \mathbf{x}_{t-1}) = \text{pr}(\mathbf{x}_{s-1}, z_s \mid z_t, \mathbf{x}_{t-1}),$$

we obtain a constraint on the point effects $\theta(\mathbf{x}_{t-1}, z_t)$.

Corollary 2. *For the given probabilities of covariates and treatments, under the Markov process and pattern (25), all $\theta(\mathbf{x}_{t-1}, z_t)$ are indexed by the net effect vector φ and given by*

$$\begin{aligned}\theta(\mathbf{x}_{t-1}, z_t) &= \phi(\mathbf{x}_{t-1}, z_t; \varphi) \\ &\quad + \sum_{s=t+1}^{T} \sum_{\mathbf{x}_{s-1}} \sum_{z_s>0} \phi(\mathbf{x}_{s-1}, z_s; \varphi) \mathrm{pr}(\mathbf{x}_{s-1}, z_s \mid \mathbf{x}_{t-1}, z_t) \\ &\quad - \sum_{s=t+1}^{T} \sum_{\mathbf{x}_{s-1}} \sum_{z_s>0} \phi(\mathbf{x}_{s-1}, z_s; \varphi) \mathrm{pr}(\mathbf{x}_{s-1}, z_s \mid \mathbf{x}_{t-1}, z_t = 0) \end{aligned} \tag{26}$$

for $t = 1, \ldots, T-1$ *and* $\theta(\mathbf{x}_{T-1}, z_T) = \phi(\mathbf{x}_{T-1}, z_T; \varphi)$.

Despite the Markov process and pattern (25), the covariate $\mathbf{x}_t$ can still be a post-treatment variable of z_t (Robins [7, 10, 12, 14, 15]). In this case, the standard parameter $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$ does not have any pattern for treatments z_t $(t < T)$ (Rosenbaum [16], Robins [8], Frangakis & Rubin [1]). If $\mathbf{x}_t$ is additionally a confounder of z_{t+1}, then one needs all the standard parameters to identify the net effect.

Remark 3. Generally, the Markov process and pattern (25) do not simplify the G-formula (4).

5. ML estimation of net effects of treatments through point effects of treatments

5.1. Likelihoods of point parameters

The data set comprises independent observations $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i\}$ on units $i = 1, \ldots, N$. Using distribution (5) in Section 3.1, we obtain the following likelihood of the point parameters

$$\begin{aligned} L(\Psi; \{y_i, \mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N) &= \prod_{i=1}^{N} f(y_i \mid \mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}; \Psi_y) && (27) \\ &\quad \prod_{t=1}^{T-1} \prod_{i=1}^{N} f(\mathbf{x}_{it} \mid \mathbf{z}_{i1}^t, \mathbf{x}_{i1}^{t-1}; \Psi_{\mathbf{x}_t}) && (28) \\ &\quad \prod_{t=1}^{T} \prod_{i=1}^{N} f(z_{it} \mid \mathbf{z}_{i1}^{t-1}, \mathbf{x}_{i1}^{t-1}; \Psi_{z_t}). && (29) \end{aligned}$$

Ignoring the sampling variability of the treatments and covariates $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, we treat the proportions as the probabilities and $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ as $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ $(t = 1, \ldots, T)$. Hence we can use (27) alone to estimate the net effect $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$. In some circumstances, it is also of interest to incorporate this sampling variability into the estimation of the net effect. In this case, we can use (28) and (29) to estimate this variability. With the estimate of

this variability, we can obtain the estimate of the net effect incorporating this sampling variability of treatments and covariates.

In the rest of the article, we are going to focus on the estimation of the net effect based on (27), thus ignoring the sampling variability of treatments and covariates.

5.2. ML estimation of net effects of treatments under a constraint

The outcome model is

$$\mu_i = \mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}) \tag{30}$$

where $\mu_i = E(y_i \mid \mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1})$ and $\mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}) = \mu(\mathbf{z}_1^T = \mathbf{z}_{i1}^T, \mathbf{x}_1^{T-1} = \mathbf{x}_{i1}^{T-1})$ which is expressed by (12) in terms of the point parameters in Ψ_y defined in Section 3.2.

Ignoring the sampling variability of $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, we replace the probabilities by the proportions and the point parameter $\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ of the population by the point parameter $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ of the sample in constraint (22), leading to the following constraint on $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$

$$\begin{aligned}
&\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\quad = \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi) \\
&\qquad + \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s > 0} \phi(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s; \varphi) \mathrm{prop}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \\
&\qquad - \sum_{s=t+1}^{T} \sum_{\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}} \sum_{z_s > 0} \phi(\mathbf{z}_1^{t-1}, z_t = 0, \mathbf{z}_{t+1}^{s-1}, \mathbf{x}_1^{s-1}, z_s; \varphi) \\
&\qquad\quad \mathrm{prop}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, z_t = 0, \mathbf{x}_1^{t-1})
\end{aligned} \tag{31}$$

for $t = 1, \ldots, T-1$ and $\vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) = \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T; \varphi)$. Formula (31) decomposes $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ into the net effects of treatments $z_s > 0$ at times $s \geq t$ in stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ versus stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)$.

For common distributions, the net effects can be estimated according to the following procedure. First, we estimate the mean $\nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ $(t = 1, \ldots, T)$ by using likelihood (27) and model (30). The estimate $\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ is the average of y in stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$. Second, we use $\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ to calculate the estimate of the point effect $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ according to formula (9). Third, we perform a regression of $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ on $\mathrm{prop}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ and $\mathrm{prop}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^{t-1}, z_t = 0, \mathbf{x}_1^{t-1})$ according to (31) to obtain the estimate of the net effect vector φ. Finally, we replace φ by $\hat{\varphi}$ in pattern (21) to obtain the estimate of $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$, that is, $\hat{\phi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \hat{\varphi})$. This procedure will be further illustrated in Section 6. Here we discuss the unbiasedness and consistency of the estimate.

The estimate $\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ is unbiased for finite sample and so is $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$. If $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi)$ is a linear function of φ, then the estimate $\hat{\varphi}$ is unbiased according to (31) treated as a regression model, and so is $\hat{\phi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$. If $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi)$ is not linear in φ, then $\hat{\varphi}$ is biased. However, we believe that $\hat{\phi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ is still unbiased if $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi)$ satisfies simple regularity conditions, for instance, that it is a smooth and monotone function of φ.

Oftentimes, the dimension k of φ is finite, that is, the net effects have a pattern of finite dimension. Treating (31) as a regression model, we see that $\hat{\varphi}$ is consistent and so is $\hat{\phi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ if there exist at least k different point effects $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ which contain the k-dimensional vector φ and whose estimates have zero covariance matrices as the sample size N approaches infinity. This condition can be satisfied even with a long treatment sequence, where the variables z_t and $\mathbf{x}_{t-1}$ take finite numbers of values.

5.3. ML estimation of net effects of treatments using treatment assignment conditions

At large t, the number of possible strata $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ is huge, and with a finite sample, most of these strata do not have both active and control values of the treatment variable z_t, so most of the point effects $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ are not estimable. In this case, it is highly difficult to estimate the net effects $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ through $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$.

However, as described in Section 4.4, we may estimate the net effects $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ through the point effects on large strata under treatment assignment conditions, for instance, $\vartheta(\mathbf{x}_{t-1}, z_t)$ under the Markov process. Even with a small sample and large t, stratum $\mathbf{x}_{t-1}$ is large enough for z_t to take both active and control values so that $\vartheta(\mathbf{x}_{t-1}, z_t)$ is estimable. Furthermore, the proportion equality (23), or equivalently,

$$\text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t > 0) = \text{prop}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}, z_t = 0)$$

can be approximately satisfied. For instance, a small sample is sufficient to allow approximately the same marginal distribution of each variable of $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-2})$ in stratum $(\mathbf{x}_{t-1}, z_t > 0)$ versus stratum $(\mathbf{x}_{t-1}, z_t = 0)$. In some applications, correlations between the variables, often the nearest ones in the sequence, have an influence on the net effect. With a slightly larger sample, we may obtain approximately the same correlations between the nearest variables in stratum $(\mathbf{x}_{t-1}, z_t > 0)$ versus stratum $(\mathbf{x}_{t-1}, z_t = 0)$.

The outcome model is still (30), i.e.

$$\mu_i = \mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}).$$

Ignoring the sampling variability of $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}\}_{i=1}^N$, we replace the probabilities by the proportions and $\theta(\mathbf{x}_{t-1}, z_t)$ by $\vartheta(\mathbf{x}_{t-1}, z_t)$ in constraint (26), leading to the following constraint on $\vartheta(\mathbf{x}_{t-1}, z_t)$

$$\begin{aligned}\vartheta(\mathbf{x}_{t-1}, z_t) &= \phi(\mathbf{x}_{t-1}, z_t; \varphi) \\ &\quad + \sum_{s=t+1}^{T} \sum_{\mathbf{x}_{s-1}} \sum_{z_s>0} \phi(\mathbf{x}_{s-1}, z_s; \varphi)\text{prop}(\mathbf{x}_{s-1}, z_s \mid \mathbf{x}_{t-1}, z_t) \\ &\quad - \sum_{s=t+1}^{T} \sum_{\mathbf{x}_{s-1}} \sum_{z_s>0} \phi(\mathbf{x}_{s-1}, z_s; \varphi)\text{prop}(\mathbf{x}_{s-1}, z_s \mid \mathbf{x}_{t-1}, z_t = 0) \qquad (32)\end{aligned}$$

for $t = 1, \ldots, T-1$ and $\vartheta(\mathbf{x}_{T-1}, z_T) = \phi(\mathbf{x}_{T-1}, z_T; \varphi)$. Formula (32) decomposes $\vartheta(\mathbf{x}_{t-1}, z_t)$ into the net effects of treatments $z_s > 0$ at times $s \geq t$ in stratum $(\mathbf{x}_{t-1}, z_t)$ versus stratum $(\mathbf{x}_{t-1}, z_t = 0)$.

All arguments and statements about the estimation procedure, unbiasedness and consistency of the ML estimate $\hat{\phi}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ are carried over from those in the previous subsection if we replace $\nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ by $\nu(\mathbf{x}_{t-1}, z_t)$, $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ by $\vartheta(\mathbf{x}_{t-1}, z_t)$, $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t; \varphi)$ by $\phi(\mathbf{x}_{t-1}, z_t; \varphi)$, $\text{prop}(\mathbf{z}_{t+1}^{s-1}, \mathbf{x}_t^{s-1}, z_s \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$ by $\text{prop}(\mathbf{x}_{s-1}, z_s \mid \mathbf{x}_{t-1}, z_t)$, (9) by (24), and (31) by (32).

6. Illustrations

6.1. An analytical example

6.1.1. The setting

Suppose that two types of treatments are alternately assigned. Let $z_t = 1$ $(t = 1, 3, \ldots, T)$ be a treatment of type one, $z_t = 1$ $(t = 2, 4, \ldots, T-1)$ be a treatment of type two, and $z_t = 0$ $(t = 1, \ldots, T)$ be the control treatment, i.e. no treatment. Between treatments z_t and z_{t+1} there is a covariate vector $\mathbf{x}_t = (x_{t1}, x_{t2})$ of two dichotomized components. Let y be the outcome after the last treatment z_T. Given $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$, there is only one net effect $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1)$ denoted by $\phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ and one point effect $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1)$ denoted by $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. In particular, $\phi(z_1 = 1) = \phi$ and $\vartheta(z_1 = 1) = \vartheta$ at $t = 1$.

Suppose that all treatments of type one have the same net effect φ_1 and all treatments of type two have the same net effect φ_2. Then the pattern of the net effects is

$$\begin{cases} \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \varphi_1, \ t = 1, 3, \ldots, T \\ \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \varphi_2, \ t = 2, 4, \ldots, T-1. \end{cases} \qquad (33)$$

Let $T^{(1)}(t)$ denote the set of times since t at which treatments of type one are assigned and $T^{(2)}(t)$ the set of times since t at which treatments of type two are assigned. Decomposing the point effect $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ of $z_t = 1$ into the net effects of $z_s = 1$ $(s \geq t)$ in stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1)$ versus stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)$ and using the pattern above, we obtain the following constraint on $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ $(t = 1, \ldots, T)$

$$\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \varphi_1 c^{(1)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \varphi_2 c^{(2)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \qquad (34)$$

with

$$c^{(q)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \sum_{s \in T^{(q)}(t)} \{\text{prop}(z_s = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1) - \text{prop}(z_s = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)\}$$

where $q = 1, 2$ indicates treatments of types one and two respectively.

6.1.2. ML estimates of net effects of treatments under a constraint

Now suppose that y is normally distributed. For simplicity, additionally suppose that the variance is equal to one for any given $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. Given the data set $\{\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}, y_i\}_{i=1}^N$, likelihood (27) then becomes

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\{y_i - \mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1})\}^2\right]$$

where $\mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1})$ is expressed in terms of point parameters by (12). The outcome model we use is (30), i.e.

$$\mu_i = \mu(\mathbf{z}_{i1}^T, \mathbf{x}_{i1}^{T-1}).$$

First we estimate the point effect $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. Let $s(A)$ be the set of units in stratum A and $n(A)$ be the number of units in stratum A. Using the outcome model and the likelihood above, we obtain the estimate for the mean of y in stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$

$$\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) = \frac{\sum_{i \in s(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)} y_i}{n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)}$$

and its variance

$$\text{var}\{\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)\} = \frac{1}{n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)}.$$

Using (9), we then obtain the estimate for the point effect of $z_t = 1$ on stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$

$$\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1) - \hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0) \tag{35}$$

and its variance

$$\text{var}\{\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})\} = \text{var}\{\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1)\} + \text{var}\{\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 0)\}. \tag{36}$$

In Appendix A.4, we prove

Proposition 1. *Suppose that the outcome y is normal and has the same known variance for all given $(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. Then the score function $U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})}$ depends*

only on $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. *Therefore the estimate* $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ *is independent of the estimates of all other point parameters.*

Remark 4. For the outcome of other distributions, the estimate $\hat{\nu}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t)$ and thus $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ are highly robust to point parameters at time $s > t$. Therefore $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ at time t is weakly correlated with estimates of the point parameters at the other times and the correlation may be omitted.

As a result of the proposition, we treat constraint (34) as a linear regression with unequal variances $\text{var}\{\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})\}$. Using techniques of the linear regression, we regress $\hat{\vartheta}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ on $c^{(1)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ and $c^{(2)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ to obtain the estimates of φ_1 and φ_2. The closed form for these estimates can easily be derived and are not presented here due to space restrictions. The estimates $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are both unbiased and consistent.

6.1.3. ML estimates of net effects of treatments under the Markov process

Suppose that the net effects have pattern (33) while the treatment assignment mechanism is a Markov process, described in Sections 4.4 and 5.3, in which the assignment of z_t depends only on $\mathbf{x}_{t-1}$. Given $\mathbf{x}_{t-1}$, there is only one net effect $\phi(\mathbf{x}_{t-1}, z_t = 1)$ denoted by $\phi(\mathbf{x}_{t-1})$ and one point effect $\vartheta(\mathbf{x}_{t-1}, z_t = 1)$ denoted by $\vartheta(\mathbf{x}_{t-1})$, and noticeably, $\phi(z_1 = 1) = \phi$ and $\vartheta(z_1 = 1) = \vartheta$ at $t = 1$. Notice that stratum $\mathbf{x}_{t-1}$ is much larger than stratum $(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ and so $\vartheta(\mathbf{x}_{t-1})$ is almost always estimable even with a long treatment sequence and a finite sample.

Decomposing $\vartheta(\mathbf{x}_{t-1})$ into the net effects of $z_s = 1$ $(s \geq t)$ in stratum $(\mathbf{x}_{t-1}, z_t = 1)$ versus stratum $(\mathbf{x}_{t-1}, z_t = 0)$ and using pattern (33), we obtain the following constraint on $\vartheta(\mathbf{x}_{t-1})$ $(t = 1, \ldots, T)$

$$\vartheta(\mathbf{x}_{t-1}) = \varphi_1 c^{(1)}(\mathbf{x}_{t-1}) + \varphi_2 c^{(2)}(\mathbf{x}_{t-1}) \tag{37}$$

with

$$c^{(q)}(\mathbf{x}_{t-1}) = \sum_{s \in T^{(q)}(t)} \{\text{prop}(z_s = 1 \mid \mathbf{x}_{t-1}, z_t = 1) - \text{prop}(z_s = 1 \mid \mathbf{x}_{t-1}, z_t = 0)\}$$

where $q = 1, 2$ indicates treatments of types one and two respectively.

Using the likelihood and the outcome model given in the beginning of the previous subsection, we obtain the estimate for the mean of y in stratum $(\mathbf{x}_{t-1}, z_t)$

$$\hat{\nu}(\mathbf{x}_{t-1}, z_t) = \frac{\sum_{i \in s(\mathbf{x}_{t-1}, z_t)} y_i}{n(\mathbf{x}_{t-1}, z_t)}$$

and its variance

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

Under the Markov process, we use (24) to obtain the estimate for point effect $\vartheta(\mathbf{x}_{t-1})$ of $z_t = 1$ on stratum $\mathbf{x}_{t-1}$

$$\begin{aligned}\hat{\vartheta}(\mathbf{x}_{t-1}) &= \sum_{\mathbf{z}_1^{t-1},\mathbf{x}_1^{t-2}} \hat{\vartheta}(\mathbf{z}_1^{t-1},\mathbf{x}_1^{t-1})\text{prop}(\mathbf{z}_1^{t-1},\mathbf{x}_1^{t-2} \mid \mathbf{x}_{t-1}) \\ &= \hat{\nu}(\mathbf{x}_{t-1}, z_t = 1) - \hat{\nu}(\mathbf{x}_{t-1}, z_t = 0)\end{aligned}$$

and its variance

$$\text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\} = \text{var}\{\hat{\nu}(\mathbf{x}_{t-1}, z_t = 1)\} + \text{var}\{\hat{\nu}(\mathbf{x}_{t-1}, z_t = 0)\}.$$

From Proposition 1, we see that $\hat{\vartheta}(\mathbf{x}_{t-1})$ at time t is independent of the estimates of point parameters at the other times, i.e. $\vartheta(\mathbf{z}_1^{s-1},\mathbf{x}_1^{s-1})$ including $\vartheta(\mathbf{x}_{s-1})$ $(t \neq s)$, the point effects $\zeta(\mathbf{z}_1^{s},\mathbf{x}_1^{s-1},\mathbf{x}_s)$ $(s = 1,\ldots,T-1)$ and the grand mean ν.

Therefore we can treat constraint (37) as a linear regression with unequal variances $\text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\}$. Using techniques of the linear regression, we regress $\hat{\vartheta}(\mathbf{x}_{t-1})$ on $c^{(1)}(\mathbf{x}_{t-1})$ and $c^{(2)}(\mathbf{x}_{t-1})$ to obtain a closed form for the estimates of φ_1 and φ_2, which are not presented here given space considerations. The estimates $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are both unbiased and consistent.

In the estimation above, we have only used the likelihood, the outcome model and constraint (37), albeit in the point parametrization. It is theoretically possible to do the same estimation in the standard parametrization by using the G-computation algorithm, but this is practically difficult due to the high dimension of standard parameters and the complex expression of constraint (37) in terms of standard parameters. Furthermore, if equalities are imposed on standard parameters, then the estimation is biased (Robins [7, 10, 12, 14, 15]).

6.2. A simulation study

Here we show by simulation that interval estimation of the net effect of treatment achieves the nominal coverage probability in the example of the previous subsection. We consider the case of $T = 3$, so pattern (33) of the net effects is

$$\begin{cases} \phi(\mathbf{z}_1^{t-1},\mathbf{x}_1^{t-1}) = \varphi_1, \ t = 1,3 \\ \phi(z_1,\mathbf{x}_1) = \varphi_2. \end{cases}$$

We are going to study three situations with the net effect $(\varphi_1,\varphi_2) = (20,10)$, $(20,0)$, $(0,0)$ respectively. In particular, the vector $(20,0)$ represents the null net effect of treatment z_2 in the presence of non-zero net effects of treatments z_1 and z_3; the vector $(0,0)$ represents the null net effects of all treatments.

To generate the data with the above pattern of net effects, we construct the standard parameters $\mu(z_1,\mathbf{x}_1,\ z_2,\ \mathbf{x}_2,z_3)$, as explained here. First, we construct the proportions of z_1, $\mathbf{x}_1$, z_2, $\mathbf{x}_2$, and z_3 which give a sample size of 648 units yielding integer frequencies for all $(z_1,\mathbf{x}_1,z_2,\mathbf{x}_2,z_3)$. Second, with the obtained proportions and a given value of (φ_1,φ_2), we calculate the point effect $\vartheta(\mathbf{z}_1^{t-1},\mathbf{x}_1^{t-1})$ $(t = 1,2,3)$ according to formula (34). The second part of

(34) is used to calculate the constants $c^{(1)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ and $c^{(2)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ while the first part is used to calculate $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. Third, we arbitrarily choose the point effect $\zeta(\mathbf{z}_1^t, \mathbf{x}_1^t)$ $(t = 1, 2)$ and the grand mean ν, which according to Theorem 1 do not affect the net effect. Finally, we insert the obtained point parameters $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$, $\zeta(\mathbf{z}_1^t, \mathbf{x}_1^t)$ and ν into (12) to obtain the standard parameters $\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. In this setting, we have also made the time-dependent covariates $\mathbf{x}_1$ and $\mathbf{x}_2$ post-treatment variables of the earlier treatments and confounders of the subsequent treatments. Furthermore, the treatment assignment follows a Markov process such that the assignment of z_2 only depends on $\mathbf{x}_1$ and that of z_3 on $\mathbf{x}_2$. A detailed description of the procedure is presented in Table 1. The standard parameters and the relevant SAS code are given in the supplementary material [20].

Because we have ignored the sampling variability of treatments and covariates since Section 5.1, we only generate the outcome y given $(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3)$. With the obtained standard parameter $\mu(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3)$, assuming that the variance of the outcome y given $(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3)$ is one, we generate y to form a data set of 648 observations on $(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3, y)$. A total of 2000 data sets are generated. For each data set, we calculate the confidence interval of the net effect as follows. Using the method described in Section 6.1.3, we calculate the estimate $\hat{\vartheta}(\mathbf{x}_{t-1})$ and the constants $c^{(1)}(\mathbf{x}_{t-1})$ and $c^{(2)}(\mathbf{x}_{t-1})$ and then regress $\hat{\vartheta}(\mathbf{x}_{t-1})$ on $c^{(1)}(\mathbf{x}_{t-1})$ and $c^{(2)}(\mathbf{x}_{t-1})$ according to (37) to obtain the estimates $\hat{\varphi}_1$ and $\hat{\varphi}_2$. With $\hat{\varphi}_1$ and its variance, we calculate the confidence interval of φ_1. With 2000 data sets, we obtain 2000 confidence intervals. By counting how many confidence intervals contain the given value of φ_1, we obtain the actual coverage probability for the confidence interval of φ_1. The same procedure is followed for φ_2. The SAS code generating the data set and calculating the actual coverage probability is presented in the supplementary material. The mean and standard deviation of $\hat{\varphi}_i$ $(i = 1, 2)$ and the actual coverage probability of the 95 % confidence interval of φ_i are presented in Table 2, together with the given value of φ_i.

Table 2 shows that for the three simulations with $(\varphi_1, \varphi_2) = (20, 10)$, $(20, 0)$, $(0, 0)$ respectively, the actual coverage probability for the 95% confidence interval is the same: 95.20% for φ_1 and 94.85% for φ_2.

6.3. A real example

Here we use a medical example to illustrate the practical procedure of estimating the net effect. Many HIV-infected patients use recreational drugs such as cocaine. A relevant medical question is how the recreational drug influences the count of CD4 cells which reflects progression of the disease. We attempt to estimate the net effect of the recreational drug on the CD4 count when the drug is used repeatedly.

The Multicenter AIDS Cohort Study enrolled nearly 5000 gay or bisexual men from Baltimore, Pittsburgh, Chicago and Los Angeles between 1984 and 1991, and required these men to return every 6 months to complete a questionnaire

TABLE 1

The procedure of constructing standard parameters given net effects of treatments as described in Section 6.2. The SAS code and the obtained standard parameters are given in the supplementary material

(1) Set proportions of $z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3$ where $\mathbf{x}_1 = (x_{11}, x_{12})$ and $\mathbf{x}_2 = (x_{21}, x_{22})$	
$\text{prop}(z_1 = 1)$	$1/2$
$\text{prop}(\mathbf{x}_1 \mid z_1)$	$\text{prop}(0,0 \mid 0) = 1/6,\ \text{prop}(0,0 \mid 1) = 2/6,$ $\text{prop}(1,0 \mid 0) = 2/6,\ \text{prop}(1,0 \mid 1) = 1/6,$ $\text{prop}(0,1 \mid 0) = 2/6,\ \text{prop}(0,1 \mid 1) = 1/6$
$\text{prop}(z_2 \mid \mathbf{x}_1) = \text{prop}(z_2 \mid z_1, \mathbf{x}_1)$	$\text{prop}(1 \mid 0,0) = 1/3,\ \text{prop}(1 \mid 0,1) = 2/3,$ $\text{prop}(1 \mid 1,0) = 2/3,\ \text{prop}(1 \mid 1,1) = 1/3$
$\text{prop}(\mathbf{x}_2 \mid z_2) = \text{prop}(\mathbf{x}_2 \mid z_1, \mathbf{x}_1, z_2)$	$\text{prop}(0,0 \mid 0) = 1/6,\ \text{prop}(0,0 \mid 1) = 2/6,$ $\text{prop}(1,0 \mid 0) = 2/6,\ \text{prop}(1,0 \mid 1) = 1/6,$ $\text{prop}(0,1 \mid 0) = 2/6,\ \text{prop}(0,1 \mid 1) = 1/6$
$\text{prop}(z_3 \mid \mathbf{x}_2) = \text{prop}(z_3 \mid z_1, \mathbf{x}_2, z_2, \mathbf{x}_2)$	$\text{prop}(1 \mid 0,0) = 1/3,\ \text{prop}(1 \mid 0,1) = 2/3,$ $\text{prop}(1 \mid 1,0) = 2/3,\ \text{prop}(1 \mid 1,1) = 1/3$

Using the above proportions, we obtain other proportions of $z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3$, for instance, $\text{prop}(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3) = \text{prop}(z_1)\text{prop}(\mathbf{x}_1 \mid z_1)\text{prop}(z_2 \mid \mathbf{x}_1)\text{prop}(\mathbf{x}_2 \mid z_2)\text{prop}(z_3 \mid \mathbf{x}_2)$. The sample size is set at 648, which yields integer frequencies for all $(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3)$.

(2) Calculate the point effects of treatments according to (34)

Using the above proportions, we calculate the following constants:

$$c^{(1)} = 1 + \text{prop}(z_3 = 1 \mid z_1 = 1) - \text{prop}(z_3 = 1 \mid z_1 = 0);$$
$$c^{(2)} = \text{prop}(z_2 = 1 \mid z_1 = 1) - \text{prop}(z_2 = 1 \mid z_1 = 0);$$
$$c^{(1)}(z_1, \mathbf{x}_1) = \text{prop}(z_3 = 1 \mid z_1, \mathbf{x}_1, z_2 = 1) - \text{prop}(z_3 = 1 \mid z_1, \mathbf{x}_1, z_2 = 0);\ c^{(2)}(z_1, \mathbf{x}_1) = 1;$$
$$c^{(1)}(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2) = 1;\ c^{(2)}(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2) = 0.$$

With a given value of (φ_1, φ_2), we then calculate the point effects of treatments

$$\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = \varphi_1 c^{(1)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \varphi_2 c^{(2)}(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \text{ for } t = 1, 2, 3.$$

(3) Choose the point effects of covariates and the grand mean arbitrarily

For $z_1 = 0, 1$, we have

$$\zeta(z_1, \mathbf{x}_1) = \begin{cases} 10 + 5z_1, & \mathbf{x}_1 = (0,1) \\ 12 + 5z_1, & \mathbf{x}_1 = (1,0) \\ 13 + 5z_1, & \mathbf{x}_1 = (1,1) \end{cases}$$

For $z_1 = 0, 1$, $z_2 = 0, 1$, and $f(\mathbf{x}_1) = 0, 3, 6, 9$ when $\mathbf{x}_1 = (0,0), (0,1), (1,0), (1,1)$ respectively, we have

$$\zeta(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2) = \begin{cases} 10 - 5z_1 - 2z_2 + f(\mathbf{x}_1), & \mathbf{x}_2 = (0,1) \\ 12 - 5z_1 - 2z_2 + f(\mathbf{x}_1), & \mathbf{x}_2 = (1,0) \\ 10 - 5z_1 - 3z_2 + f(\mathbf{x}_1), & \mathbf{x}_2 = (1,1) \end{cases}$$

For the grand mean, we have $\nu = 15$

(4) Calculate the standard parameters according to (12)

$$\begin{aligned} &\mu(z_1, \mathbf{x}_1, z_2, \mathbf{x}_2, z_3) \\ &\quad = -\textstyle\sum_{t=1}^{3} \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})\{\text{prop}(z_t^* = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) - I(z_t)\} \\ &\quad = -\textstyle\sum_{t=1}^{2} \{\sum_{\mathbf{x}_t^* > \mathbf{0}} \zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t^*)\text{prop}(\mathbf{x}_t^* \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}) - \zeta(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}, \mathbf{x}_t)\} \\ &\qquad + \nu \end{aligned}$$

where $I(z_t)$ equals one when $z_t = 1$ and zero otherwise.

TABLE 2

Results from a simulation study of the net effect vector (φ_1, φ_2) in Section 6.2: (1) mean and (2) standard deviation (sd) for the ML estimate of the net effect and (3) actual coverage probability (acp) for the 95 % confidence interval of the net effect. 2000 data sets are used, each having 648 units. The SAS code is given in the supplementary material

	Point and interval estimations of the net effect					
	φ_1			φ_2		
(φ_1, φ_2)	mean	sd	acp	mean	sd	acp
(20,10)	20.00	0.06	95.20	10.00	0.09	94.85
(20,0)	20.00	0.06	95.20	−0.00	0.09	94.85
(0,0)	−0.00	0.06	95.20	−0.00	0.09	94.85

and undergo various examinations (Kaslow et al. [3]). Our data was a subset of the data from the study which involved 375 participants who were seronegative at entry and seroconverted during the follow-up (Zeger and Diggle [21]). In the initial period of seroconversion, these participants were not exposed to anti-HIV drugs, which might complicate the net effect of the recreational drug. Hence we restricted our study to the visit before seroconversion ($t = 0$) and the first and second visits after seroconversion ($t = 1, 2$). Furthermore, some participants lost their follow-ups due to unknown non-ignorable missing data mechanism, so we excluded these patients and finally obtained a data set of 256 participants. The data set is presented in the supplementary material.

At each visit $t = 0, 1, 2$, a participant recalled or was examined for the drug use (z_t), CD4 count (x_{t1}), the number of packs of cigarettes smoked daily (x_{t2}), the number of sexual partners (x_{t3}) and a mental illness score (x_{t4}). We assumed that z_t occurred prior to $x_{t1}, \ldots, x_{t4}$. Age (x_{05}) was also included as a covariate at visit $t = 0$. Consequently, the temporal order of these variables is $\{z_0, (x_{01}, \ldots, x_{05}), z_1, (x_{11}, \ldots, x_{14}), z_2, (x_{21}, \ldots, x_{24})\}$.

The treatment variables are drug uses z_1 and z_2. Due to incomplete information about covariates prior to drug use z_0, it is not possible to obtain the net effect of z_0. Instead, we use z_0 and $x_{01}, \ldots, x_{05}$ as stationary covariates to adjust for participants' differences. The time-dependent covariates between z_1 and z_2 are $x_{11}, \ldots, x_{14}$. The outcome is the logarithm of CD4 count at $t = 2$, i.e. $y = \log(x_{21})$. All variables prior to y are dichotomized, with ones meaning 'yes' or 'high' and zeros meaning 'no' or 'low'. We assume that y is normally distributed.

Let $\mathbf{x}_0 = (x_{01}, x_{02}, x_{03}, x_{04}, x_{05})$ and $\mathbf{x}_1 = (x_{11}, x_{12}, x_{13}, x_{14})$. We are going to estimate the net effect $\phi(z_0, \mathbf{x}_0)$ of $z_1 = 1$ and the net effect $\phi(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1)$ of $z_2 = 1$. Given the small sample size, it is reasonable to assume the pattern $\phi(z_0, \mathbf{x}_0) = \varphi_1$ and $\phi(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1) = \varphi_2$.

As described in Section 5.2, we use the model, for $i = 1, \ldots, 256$,

$$\nu_i = \nu(z_{i0}, \mathbf{x}_{i0}, z_{i1}, \mathbf{x}_{i1}, z_{i2})$$

to estimate the point effect $\vartheta(z_0, \mathbf{x}_0)$ of drug use $z_1 = 1$ and the point effect $\vartheta(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1)$ of drug use $z_2 = 1$. When implementing this estimation by common statistical softwares like SAS, we (1) use the model to estimate the variance of y given $(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1, z_2)$, i.e. $\text{var}(y \mid z_0, \mathbf{x}_0, z_1, \mathbf{x}_1, z_2)$, (2) use the

same model to estimate $\vartheta(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1)$, and (3) use the model

$$\nu_i = \nu(z_{i0}, \mathbf{x}_{i0}, z_{i1})$$

to estimate $\vartheta(z_0, \mathbf{x}_0)$ but we use the variance estimated from step (1).

We improve the estimation in the usual framework of statistical modeling. By the likelihood ratio-based significance test of the model parameters at the significance level of 10 %, we find that only the CD4 count x_{01} is significant of all covariates and so $\vartheta(z_0, \mathbf{x}_0)$ is considered equal to the point effect $\vartheta(x_{01})$ of drug use $z_1 = 1$. Furthermore, x_{01} does not have a significant interaction with z_1. As a result, we use the model

$$\nu_i = \nu(x_{i01}, z_{i1}) = \alpha_1 + z_{i1}\beta_1 + x_{i01}\gamma_{01} \tag{38}$$

to estimate $\vartheta(x_{01}) = \beta_1$. Similarly, we find that only the CD4 counts x_{01} and x_{11} are significant of all covariates and so $\vartheta(z_0, \mathbf{x}_0, z_1, \mathbf{x}_1)$ is considered equal to the point effect $\vartheta(x_{01}, x_{11})$ of drug use $z_2 = 1$. Furthermore, x_{01} and x_{11} do not have a significant interaction with z_2. As a result, we use the model

$$\nu_i = \nu(x_{i01}, x_{i11}, z_{i2}) = \alpha_2 + z_{i2}\beta_2 + x_{i01}\lambda_{01} + x_{i11}\lambda_{11} \tag{39}$$

to estimate $\vartheta(x_{01}, x_{11}) = \beta_2$. In the estimation above, the variance of y given $(x_{01}, z_1, x_{11}, z_2)$, i.e. $\mathrm{var}(y \mid x_{01}, z_1, x_{11}, z_2)$, is estimated by using the model

$$\nu_i = \nu(x_{i01}, z_{i1}, x_{i11}, z_{i2}). \tag{40}$$

We decompose the point effects β_1 and β_2 into the net effects φ_1 and φ_2 according to their proportions and obtain

$$\begin{cases} \beta_1 = \varphi_1 + \varphi_2\{\mathrm{prop}(z_2 = 1 \mid z_1 = 1) - \mathrm{prop}(z_2 = 1 \mid z_1 = 0)\}, \\ \beta_2 = \varphi_2. \end{cases} \tag{41}$$

According to this decomposition, we regress $\hat{\beta}_1$ and $\hat{\beta}_2$ on the proportions to estimate φ_1 and φ_2. The estimate (variance) of φ_1 is 0.04(0.08) whereas that of φ_2 is 0.08(0.06). This result shows that the recreational drug has an immediate positive effect on CD4 count (i.e. φ_2), which is well-known in the medical literature, and a decreasing distant effect on CD4 count (i.e. φ_1), which is observed in this study. The estimation procedure and the results are presented in Table 3.

In the procedure above, we estimated the variances $\mathrm{var}(\hat{\beta}_1)$ and $\mathrm{var}(\hat{\beta}_2)$ through $\mathrm{var}(y \mid x_{01}, z_1, , x_{11}, z_2)$ estimated from model (40). On the other hand, we can estimate $\mathrm{var}(\hat{\beta}_1)$ through $\mathrm{var}(y \mid x_{01}, z_1)$ estimated from model (38), and $\mathrm{var}(\hat{\beta}_2)$ through $\mathrm{var}(y \mid x_{01}, x_{11}, z_2)$ estimated from model (39). Because (40) avoids dispersion of the covariates, $\mathrm{var}(y \mid x_{01}, z_1, x_{11}, z_2)$ is smaller than $\mathrm{var}(y \mid x_{01}, x_{11}, z_2)$ and $\mathrm{var}(y \mid x_{01}, z_1)$, yielding smaller estimates for $\mathrm{var}(\hat{\beta}_1)$ and $\mathrm{var}(\hat{\beta}_2)$.

However, the latter method can be useful in the case of a long treatment sequence. As an example, recall the example described in Section 6.1.3, where

TABLE 3

The procedure and results of estimating the net effects of the recreational drugs z_1 and z_2 on the logarithm y of the CD4 count after z_2. The medical background is described in Section 6.3. The SAS code and the data set are given in the supplementary material

(1) Estimating the point effects of treatments according to (38) and (39)	
Point effect of treatment	Estimate (standard deviation)
$\vartheta(x_{01})$	0.10 (0.07)
$\vartheta(x_{01}, x_{11})$	0.08 (0.06)
(2) Decomposing the point effects into the net effects according to (41)	
$\vartheta(x_{01}) = \varphi_1 + 0.73\varphi_2$	
$\vartheta(x_{01}, x_{11}) = \varphi_2$	
(3) Estimate the net effects according to the decomposition above	
Net effect of treatment	Estimate (standard deviation)
φ_1	0.04 (0.08)
φ_2	0.08 (0.06)

we calculated the variance $\text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\}$ through the variance $\text{var}(y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. However, with a long treatment sequence and a finite sample, it is highly difficult to estimate $\text{var}(y \mid \mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. In this case, we can estimate $\text{var}\{\hat{\vartheta}(\mathbf{x}_{t-1})\}$ through $\text{var}(y \mid \mathbf{x}_{t-1}, z_t)$ estimated from the model

$$\nu_i = \nu(\mathbf{x}_{i(t-1)}, z_{it}).$$

7. Conclusion

In this article we have identified net effects of treatments by point effects of treatments and estimated net effects of treatments via point effects of treatments by maximum likelihood. Using methods in the single-point causal inference (Rosenbaum & Rubin [17], Rosenbaum [18], Rubin [19]), we improve estimation of the net effects by using constraints on their point effects and treatment assignment conditions.

Given a data, model and the likelihood, our estimation of the net effects is most efficient due to the nature of maximum likelihood. The point estimation is unbiased while the interval estimation achieves the nominal coverage probability. Furthermore, the ML estimate of the net effect is consistent in many practical applications where the net effect has a pattern of finite dimension while treatments and covariates take finite numbers of values. The consistency is true even when the treatment sequence gets long and the number of point parameters increases exponentially. It is interesting to compare this consistency with the inconsistency of the ML estimate of the causal effect of a single-point treatment when adjusting for a confounder of infinite dimension (Robins & Ritov [11]). In the latter case, the ML estimate of the single-point causal effect is highly correlated with that of the confounder of infinite dimension.

There are rich literatures on semi-parametric approaches to estimation of the net effect, for instance, the marginal structural model and the structural nested model. These approaches are also used to estimate the causal effects of treatment sequences including dynamic ones under special conditions where,

for instance, study units remain on active treatments after treatment initiation and the net effects are the same for all treatments (Lok & DeGruttola [2]). In comparison, our approach estimates the net effects of all treatments in the sequence, though in a relatively simple setting. Together with the proportions of treatments and covariates, the ML estimates of these net effects can be used to obtain the ML estimate of the causal effect of any treatment sequence (for instance, Robins [10]). Due to the scope of this article, we have only focused on the net effect.

Due to the scope of this article, we have only considered the situation where treatments are assigned at fixed times, treatments and covariates are discrete, there is no missing data, the outcome model is linear and the point and net effects are measured by differences. However, there are methods available for estimating the causal effect of one single-point treatment in more complex settings. We believe that analogous methods can be developed to estimate the net effects in more complex settings in the context of a treatment sequence.

Acknowledgment and conflict of interest

The authors wish to thank the reviewer and Professor Michailidis for the important comments and suggestions, which have greatly improved the article. This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

Appendix A

A.1. Proof of formula (12)

Using (9) at $t = T$, we obtain

$$\nu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T = 0) + \vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) \tag{42}$$

where we take $\vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T = 0) = 0$. Taking the average on both sides of (42) with respect to $\text{prop}(z_T \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})$, we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})\\ &\quad = \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T = 0) + \sum_{z_T > 0} \vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T)\text{prop}(z_T \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})\end{aligned}$$

which implies that

$$\begin{aligned}&\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T = 0)\\ &\quad = -\sum_{z_T > 0} \vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T)\text{prop}(z_T \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) + \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}).\end{aligned}$$

Inserting this into (42), we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})\\&\quad= \sum_{z_T^*>0} -\vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T^*)\text{prop}(z_T^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})\\&\qquad+ \vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) + \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}).\end{aligned} \tag{43}$$

Using (10) at $t = T-1$ and then following the procedure above, we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})\\&\quad= \sum_{\mathbf{x}_{T-1}^*>0} -\zeta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}, \mathbf{x}_{T-1}^*)\text{prop}(\mathbf{x}_{T-1}^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2})\\&\qquad+ \zeta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}, \mathbf{x}_{T-1}) + \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}).\end{aligned} \tag{44}$$

Inserting (44) into (43), we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})\\&\quad= \sum_{z_T^*>0} -\vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T^*)\text{prop}(z_T^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) + \vartheta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T)\\&\qquad+ \sum_{\mathbf{x}_{T-1}^*>0} -\zeta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}, \mathbf{x}_{T-1}^*)\text{prop}(\mathbf{x}_{T-1}^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2})\\&\qquad+ \zeta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}, \mathbf{x}_{T-1}) + \nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}).\end{aligned}$$

We continue with the same procedure for $\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}), \ldots, \nu(z_1)$ consecutively and then use $\nu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1})$. Finally we obtain formula (12).

Formula (43) is true for any T. Taking $T = t$, we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1})\\&\quad= \sum_{z_t^*>0} -\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*)\text{prop}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})\\&\qquad+ \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) + \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})\end{aligned} \tag{45}$$

which will be used in Appendix A.4.

Starting from formula (44) and following the same procedure for $\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}), \ldots, \nu(z_1)$ consecutively, we obtain

$$\begin{aligned}&\nu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1})\\&\quad= \sum_{s=1}^{T-1} \left\{ \sum_{z_s^*>0} -\vartheta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*)\text{prop}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \vartheta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\}\\&\qquad+ \sum_{s=1}^{T-1} \left\{ \sum_{\mathbf{x}_s^*>\mathbf{0}} -\zeta(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*)\text{prop}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \zeta(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} + \nu.\end{aligned}$$

This expression is true for any T. Taking $T = t$ and then letting N approach infinity, we obtain

$$\begin{aligned}\mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) &= \sum_{s=1}^{t-1} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\} \\ &\quad + \sum_{s=1}^{t-1} \left\{ \sum_{\mathbf{x}_s^*>\mathbf{0}} -\xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} + \mu \end{aligned} \tag{46}$$

which will be used in Appendix A.2.

A.2. Proof of formula (17)

We rewrite (16) as

$$\mu(\mathbf{z}_1^{T}, \mathbf{x}_1^{T-1}) = q_1 + q_2 + q_3 \tag{47}$$

with

$$\begin{aligned} q_1 &= \sum_{s=1}^{t-1} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\} \\ &\quad + \sum_{s=1}^{t-1} \left\{ \sum_{\mathbf{x}_s^*>\mathbf{0}} -\xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} + \mu, \\ q_2 &= \sum_{s=t}^{T} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\}, \\ q_3 &= \sum_{s=t}^{T-1} \left\{ \sum_{\mathbf{x}_s^*>\mathbf{0}} -\xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^{s}, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} \end{aligned}$$

for $t = 1, \ldots, T-1$ and $\mu(\mathbf{z}_1^{T}, \mathbf{x}_1^{T-1}) = q_1 + q_2$ for $t = T$. According to (46), we have

$$q_1 = \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}).$$

Combining formula (3) with (47) for $t = T$, we obtain

$$\begin{aligned} E\{y(\mathbf{z}_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} &= \mu(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) \\ &\quad + \sum_{z_T^*>0} -\theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T^*) \mathrm{pr}(z_T^* \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}) + \theta(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T), \end{aligned}$$

which is (17) for $t = T$.

Now we derive (17) for $t = 1, \ldots, T-1$. Combining (3) with (47) for $t = 1, \ldots, T-1$, we obtain

$$E\{y(\mathbf{z}_t^T) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} = \sum_{\mathbf{x}_t^{T-1}} (q_1 + q_2 + q_3) \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}).$$

Because q_1 does not involve $\mathbf{x}_t^{T-1}$ and $\sum_{\mathbf{x}_t^{T-1}} \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = 1$, the term involving q_1 is equal to

$$\sum_{\mathbf{x}_t^{T-1}} q_1 \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = q_1 = \mu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}).$$

The term involving q_3 is equal to

$$\begin{aligned}
&\sum_{\mathbf{x}_t^{T-1}} q_3 \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) \\
&\quad = \sum_{s=t}^{T-1} \sum_{\mathbf{x}_t^{T-1}} \left\{ \sum_{\mathbf{x}_s^* > \mathbf{0}} -\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} \\
&\qquad \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) \\
&\quad = \sum_{s=t}^{T-1} \sum_{\mathbf{x}_t^{s}} \left\{ \sum_{\mathbf{x}_s^* > \mathbf{0}} -\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} \\
&\qquad \prod_{k=t}^{s} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) \\
&\quad = \sum_{s=t}^{T-1} \sum_{\mathbf{x}_t^{s-1}} \sum_{\mathbf{x}_s} \left\{ \sum_{\mathbf{x}_s^* > \mathbf{0}} -\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} \\
&\qquad \mathrm{pr}(\mathbf{x}_s \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) \prod_{k=t}^{s-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = 0
\end{aligned}$$

where the second equality is due to $\sum_{\mathbf{x}_{s+1}^{T-1}} \prod_{k=s+1}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = 1$ and the last equality is due to

$$\begin{aligned}
&\sum_{\mathbf{x}_s} \left\{ \sum_{\mathbf{x}_s^* > \mathbf{0}} -\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \right\} \mathrm{pr}(\mathbf{x}_s \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) \\
&= \sum_{\mathbf{x}_s^* > \mathbf{0}} -\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s^*) \mathrm{pr}(\mathbf{x}_s^* \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) + \sum_{\mathbf{x}_s} \xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s) \mathrm{pr}(\mathbf{x}_s \mid \mathbf{z}_1^s, \mathbf{x}_1^{s-1}) = 0
\end{aligned}$$

where we have used $\xi(\mathbf{z}_1^s, \mathbf{x}_1^{s-1}, \mathbf{x}_s = \mathbf{0}) = 0$.

The term involving q_2 is equal to

$$\begin{aligned}
&\sum_{\mathbf{x}_t^{T-1}} q_2 \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) \\
&\quad = \sum_{s=t}^{T} \sum_{\mathbf{x}_t^{T-1}} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\} \\
&\qquad \prod_{k=t}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) \\
&\quad = \sum_{z_t^*>0} -\theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \mathrm{pr}(z_t^* \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) + \theta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) \\
&\qquad + \sum_{s=t+1}^{T} \sum_{\mathbf{x}_t^{s-1}} \left\{ \sum_{z_s^*>0} -\theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s^*) \mathrm{pr}(z_s^* \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}) + \theta(\mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}, z_s) \right\} \\
&\qquad \prod_{k=t}^{s-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}),
\end{aligned}$$

where the last equality is due to $\sum_{\mathbf{x}_s^{T-1}} \prod_{k=s}^{T-1} \mathrm{pr}(\mathbf{x}_k \mid \mathbf{z}_1^k, \mathbf{x}_1^{k-1}) = 1$ for $s = t, \ldots, T-1$. Adding the three terms together, we obtain (17) for $t = 1, \ldots, T-1$.

A.3. Proof of formula (19)

Formula (3) for $t = T$ is

$$\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = E\{y(z_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\}. \tag{48}$$

Using definition (1) for $t = T$, we then obtain

$$\mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) = E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} + \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T)$$

which is (19) for $t = T$.

Now we derive formula (19) for $t = 1, \ldots, T-1$. Using (48), we obtain

$$\begin{aligned}
\mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) &= \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1}} \mu(\mathbf{z}_1^T, \mathbf{x}_1^{T-1}) \mathrm{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}) \\
&= \sum_{\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1}} E\{y(z_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} \mathrm{pr}(\mathbf{z}_{t+1}^T, \mathbf{x}_t^{T-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1}). \qquad (49)
\end{aligned}$$

Let $A(t) = E\{y(z_t, \mathbf{z}_{t+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}$ and

$$A(s) = \sum_{\mathbf{z}_{t+1}^s, \mathbf{x}_t^{s-1}} E\{y(z_s, \mathbf{z}_{s+1}^T = \mathbf{0}) \mid \mathbf{z}_1^{s-1}, \mathbf{x}_1^{s-1}\} \mathrm{pr}(\mathbf{z}_{t+1}^s, \mathbf{x}_t^{s-1} \mid \mathbf{z}_1^t, \mathbf{x}_1^{t-1})$$

for $s = t+1, \ldots, T$. Comparing (49) with $A(T)$, we see that $\mu(\mathbf{z}_1^t, \mathbf{x}_1^{t-1}) = A(T)$.

We rewrite $A(T)$ by

$$\begin{aligned} A(T) = & \sum_{\mathbf{z}_{t+1}^{T}, \mathbf{x}_{t}^{T-1}} [E\{y(z_T) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} - E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\}] \\ & \text{pr}(\mathbf{z}_{t+1}^{T}, \mathbf{x}_{t}^{T-1} \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) \\ & + \sum_{\mathbf{z}_{t+1}^{T}, \mathbf{x}_{t}^{T-1}} E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}\} \text{pr}(\mathbf{z}_{t+1}^{T}, \mathbf{x}_{t}^{T-1} \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) \qquad (50) \\ = & \sum_{\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-1}} \sum_{z_T > 0} \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) \text{pr}(\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-1}, z_T \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) \\ & + \sum_{\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-2}} E\{y(z_T = 0) \mid \mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-2}\} \text{pr}(\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-2} \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}). \qquad (51) \end{aligned}$$

Here the first summation term in (50) is equal to the first summation term in (51) according to definition (1) for $t = T$; the second summation term in (50), after being summed over z_T and then $\mathbf{x}_{T-1}$, is equal to the second summation term in (51).

Assumption (2) for $t = T - 1$ implies

$$y(z_{T-1}, z_T) \perp z_{T-1}^{*} \mid \mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}$$

which implies

$$\begin{aligned} E\{y(z_{T-1}, z_T = 0) \mid \mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}\} &= E\{y(z_{T-1}, z_T = 0) \mid \mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}, z_{T-1}\} \\ &= E\{y(z_T = 0) \mid \mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}, z_{T-1}\}. \end{aligned}$$

Hence the second summation term in (51) is equal to

$$\sum_{\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-2}} E\{y(z_{T-1}, z_T = 0) \mid \mathbf{z}_1^{T-2}, \mathbf{x}_1^{T-2}\} \text{pr}(\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-2} \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1})$$

which is $A(T-1)$.

Therefore we obtain

$$\begin{aligned} & A(T) \\ & = \sum_{\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-1}} \sum_{z_T > 0} \phi(\mathbf{z}_1^{T-1}, \mathbf{x}_1^{T-1}, z_T) \text{pr}(\mathbf{z}_{t+1}^{T-1}, \mathbf{x}_{t}^{T-1}, z_T \mid \mathbf{z}_1^{t}, \mathbf{x}_1^{t-1}) + A(T-1). \end{aligned} \qquad (52)$$

We continue with the same procedure to rewrite $A(T-1), \cdots, A(t+1)$ consecutively and rewrite

$$\begin{aligned} A(t) &= E\{y(z_t, \mathbf{z}_{t+1}^{T} = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} - E\{y(z_t = 0, \mathbf{z}_{t+1}^{T} = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ &\quad + E\{y(z_t = 0, \mathbf{z}_{t+1}^{T} = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\} \\ &= \phi(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t) + E\{y(\mathbf{z}_t^{T} = \mathbf{0}) \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}\}. \end{aligned}$$

Finally we obtain formula (19) for $t = 1, \ldots, T-1$.

A.4. Proof of Proposition 1

In Section 6, treatment z_t takes values one or zero and $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t = 1)$ is denoted by $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$. Using the chain rule, the score function for $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$ based on model (30) is equal to

$$U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})} = \sum_{\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)}} U_{\mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})} \frac{\partial \mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})}{\partial \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})}. \tag{53}$$

Let $I_a(x)$ be an indicator function taking the value one if $x = a$ and the value zero otherwise. Using formula (12), which has been proved in Appendix A.1, we obtain

$$\begin{aligned}\mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)}) = \sum_{t=1}^{T} \Big\{ & -\vartheta(\mathbf{z}_1^{*(t-1)}, \mathbf{x}_1^{*(t-1)}) \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{*(t-1)}, \mathbf{x}_1^{*(t-1)}) \\ & + \vartheta(\mathbf{z}_1^{*(t-1)}, \mathbf{x}_1^{*(t-1)}) I_1(z_t^*) \Big\} + A\end{aligned}$$

where A is some function of the terms that do not depend on $\vartheta(\mathbf{z}_1^{*(t-1)}, \mathbf{x}_1^{*(t-1)})$ $(t = 1, \ldots, T)$. Hence we obtain

$$\begin{aligned}&\frac{\partial \mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})}{\partial \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})} \\ &\quad = I_{(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})}(\mathbf{z}_1^{*(t-1)}, \mathbf{x}_1^{*(t-1)}) \{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \}. \end{aligned} \tag{54}$$

Furthermore, the score function $U_{\mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})}$ for $\mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})$ is

$$U_{\mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})} = \sum_{i \in s(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})} \{ y_i - \mu(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)}) \} \tag{55}$$

because y is normal, where the variance of y given $(\mathbf{z}_1^{*T}, \mathbf{x}_1^{*(T-1)})$ is assumed to be one for notational simplicity.

Inserting (54) and (55) into (53) and then summing the expression over $(\mathbf{x}_t^{*(T-1)}, \mathbf{z}_{t+1}^{*T})$, we obtain

$$\begin{aligned}&U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})} \\ &= \sum_{z_t^*=0,1} \big\{ I_1(z_t^*) - \text{prop}(z_t^{**}{=}1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \big\} \sum_{i \in s(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*)} \{ y_i - \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \} \\ &= \sum_{z_t^*=0,1} \big\{ I_1(z_t^*) - \text{prop}(z_t^{**}{=}1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \big\} \\ &\quad \left\{ \sum_{i \in s(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*)} y_i - n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \right\}. \end{aligned}$$

Replacing z_t by z_t^* and z_t^* by z_t^{**} in formula (45) and noticing that z_t^* and z_t^{**} take values one or zero, we obtain

$$\nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) = \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} + \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}).$$

Hence we obtain

$$\begin{aligned} & U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})} \\ &= \sum_{z_t^*=0,1} \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} \\ &\quad \Bigg[\sum_{i \in s(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*)} y_i - n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \\ &\quad \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} \\ &\quad - n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \Bigg]. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} & \sum_{z_t^*=0,1} \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \\ &= \{ n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^* = 1) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \} \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \\ &= \{ n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^* = 1) - n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^{**} = 1) \} \nu(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) = 0. \end{aligned}$$

Therefore we obtain

$$\begin{aligned} & U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})} \\ &= \sum_{z_t^*=0,1} \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} \\ &\quad \Bigg[\sum_{i \in s(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*)} y_i - n(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}, z_t^*) \\ &\quad \vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \left\{ I_1(z_t^*) - \text{prop}(z_t^{**} = 1 \mid \mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1}) \right\} \Bigg]. \end{aligned} \tag{56}$$

From this formula, we see that $U_{\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})}$ depends only on $\vartheta(\mathbf{z}_1^{t-1}, \mathbf{x}_1^{t-1})$, thus proving the proposition.

Supplementary Material

SAS codes and SAS data sets

(doi: 10.1214/15-EJS1046SUPP; .zip). The supplementary material contains (1) SAS codes and SAS data sets for the simulation study in Section 6.2 and (2) SAS code and SAS data set for the illustrative study in Section 6.3. (Zip file)

References

[1] Frangakis, C. E. & Rubin, D. B. (2002). Principal Stratification in Causal Inference. *Biometrics*, **58**, 21–29. MR1891039

[2] Lok, J. J. & DeGruttola, V. (2012). Impact of Time to Start Treatment Following Infection with Application to Initiating HAART in HIV-Positive Patients. *Biometrics*, **68**, 745–754. MR3055179

[3] Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. & Rinaldo, C. R. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristic of the participants. *American Journal of Epidemiology*, **126**, 310–318.

[4] Henderson, R., Ansell, P. & Alshibani, D. (2010). Regret-Regression for Optimal Dynamic Treatment Regimes. *Biometrics*, **66**, 1192–1201. MR2758507

[5] Murphy, S. A., van der Laan, J., Robins, J. M. & CPPRG (2001). Marginal Mean Model for Dynamic Regimes. *Journal of American Statistical Association*, **96**, 1410–1423. MR1946586

[6] Murphy, S. A. (2003). Optimal dynamic treatment regions. *Journal of the Royal Statistical Society: Series B*, **62**, 331–354. MR1983752

[7] Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survival effect. *Mathematical Modeling*, **7**, 1393–1512. MR0877758

[8] Robins, J. M. (1989). The control of confounding by intermediate variables. *Statistics in Medicine*, **8**, 679–701.

[9] Robins, J. M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, **79**, 321–334. MR1185134

[10] Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality, Lecture notes in Statistics (120)*, (Ed. Berkane, M.), pp. 69–117. New York: Springer-Verlag. MR1601279

[11] Robins, J. M. & Ritov, Y. (1997). Towards a curse of dimensionality asymptotic theory for semi-parametric models. *Statistics in Medicine*, **16**, 285–319.

[12] Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese*, **121**, 151–179. MR1766776

[13] Robins, J. M., Rotnitzky, A. & Scharfstein, D. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, (Eds. Halloran, M. E. & Berry, D.), IMA Volume 116, pp. 1–92. NY: Springer-Verlag. MR1731681

[14] Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Lecture notes in Statistics (179)*, (Ed. Berkane, M.), pp. 189–326. New York: Springer-Verlag. MR2129402

[15] Robins, J. M. (2009). Longitudinal Data Analysis. In *Handbooks of Modern Statistical Methods*, (Ed. Fitzmaurice, G.), pp. 553–599. Chapman & Hall / CRC. MR1500110

[16] Rosenbaum, P. R. (1984). The consequence of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A*, **147**, 656–666.

[17] Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55. MR0742974

[18] Rosenbaum, P. R. (1995). *Observational studies.* New York, NY: Springer. MR1353914

[19] Rubin, D. B. (2005). Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*, **100**, 322–331. MR2166071

[20] Wang, X. & Yin, L. (2015). Supplement to "Identifying and Estimating Net Effects of Treatments in Sequential Causal Inference" DOI: 10.1214/15-EJS1046SUPP.

[21] Zeger, S. L. & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, **50**, 689–699.