Featured Article

Statistics in Medicine

Received 9 July 2013, Accepted 26 March 2014 Published online 29 May 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6181

Identification and estimation of survivor average causal effects

Eric J. Tchetgen Tchetgen[*†]

In longitudinal studies, outcomes ascertained at follow-up are typically undefined for individuals who die prior to the follow-up visit. In such settings, outcomes are said to be truncated by death and inference about the effects of a point treatment or exposure, restricted to individuals alive at the follow-up visit, could be biased even if as in experimental studies, treatment assignment were randomized. To account for truncation by death, the survivor average causal effect (SACE) defines the effect of treatment on the outcome for the subset of individuals who would have survived regardless of exposure status. In this paper, the author nonparametrically identifies SACE by leveraging post-exposure longitudinal correlates of survival and outcome that may also mediate the exposure effects on survival and outcome. Nonparametric identification is achieved by supposing that the longitudinal data arise from a certain nonparametric structural equations model and by making the monotonicity assumption that the effect of exposure on survival agrees in its direction across individuals. A novel weighted analysis involving a consistent estimate of the survival process is shown to produce consistent estimates of SACE. A data illustration is given, and the methods are extended to the context of time-varying exposures. We discuss a sensitivity analysis framework that relaxes assumptions about independent errors in the nonparametric structural equations model and may be used to assess the extent to which inference may be altered by a violation of key identifying assumptions. © 2014 The Authors. Statistics in Medicine published by John Wiley & Sons, Ltd.

Keywords: truncation by death; principal stratification; double robust; sensitivity analysis

1. Introduction

In longitudinal studies, it sometimes happens that individuals die between follow-up visits, in which case, unobserved outcomes that would have been ascertained during follow-up are said to be truncated by death. It is well known that inference about the effects of a point treatment or exposure, restricted to individuals alive at a follow-up visit, could be biased even if as in experimental studies, treatment assignment were randomized. Similarly, it may be that a vaccine studied in a randomized trial has a protective effect against a viral infection for some but not all individuals in the study. Then, viral load associated with the infection would not be observed unless a person became infected, which is a post-randomization event. As for truncation by death, an evaluation of the effects of the vaccine on viral load among infected individuals in the study likewise could be biased. Such bias may be present, if as we expect is likely the case in the aforementioned settings, there are downstream effects of the exposure or treatment, which affect survival or post-randomization infection, and the outcome of interest. A more fundamental issue is that the outcome may not be well defined for individuals who die or remain uninfected by the virus, under either exposure status, and therefore, it is not clear that a causal effect of the exposure can be defined for such individuals. In order to appropriately account for truncation by death, one can define the survivor average causal effect (SACE), which is the effect of exposure on the outcome for the subset of individuals that would have survived regardless of exposure status [1, 2]. SACE is an instance of what is sometimes referred to as a principal strata causal effect [2–4]. An analogous principal strata causal effect is likewise defined for the effect of vaccine on viral load, among infected individuals for whom the vaccine has no effect on HIV infection [5–8]. Throughout the paper, we refer to these two types of effects

Departments of Epidemiology and Biostatistics, Harvard University, MA, U.S.A.
[*]*Correspondence to: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115, U.S.A.*
[†]*E-mail: etchetgen@gmail.com*

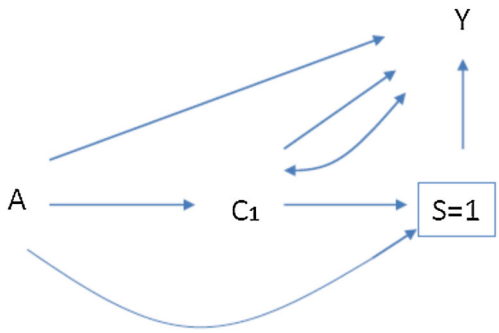

Figure 1. Causal diagram associated with the nonparametric structural equations model Equations (1)–(4).

as SACE without further distinguishing between their respective contexts. In this paper, SACE is shown to be nonparametrically identified by leveraging post-exposure longitudinal correlates of survival and outcome that may also mediate exposure effects on survival and outcome. Nonparametric identification is achieved by supposing that the longitudinal data arise from a certain nonparametric structural equations model (NPSEM) [9] and by making the monotonicity assumption that the effect of exposure on survival agrees in its direction across individuals. Under these assumptions, a novel yet simple weighted analysis with weights involving the survival process is shown to produce consistent estimates of SACE, provided that the survival process is estimated consistently. A number of alternative estimators are also described, some with interesting theoretical properties. However, it is argued that these other estimators may be more difficult to implement in practice, particularly in studies with prolonged follow-up, where the simple weighted analysis extends with little additional computational difficulty. An illustration of the simple weighted analysis is given in an application concerning the effects of smoking history on decline of cognitive function in an aging population subject to truncation by death. A general sensitivity analysis technique is described, to assess the extent to which inference might be affected by a violation of an assumption that all common causes of survival and the outcome are fully observed. Finally, in the context of time-updated exposures, the survivor marginal structural model (SMSM) is introduced, which amounts to a standard marginal structural model (MSM) for the subpopulation that would survive irrespective of treatment history. A weighted approach is described for estimating the parameters of an SMSM.

2. A simple three-occasion study

2.1. Causal diagram interpretation of biased analyses

By way of introduction, first consider a simplified version of the study of the effects of smoking on cognitive function decline, here restricted to only three longitudinal occasions, as depicted in the causal diagram in Figure 1.

The simplified design consists of a baseline $j = 0$ at which binary smoking status A is observed, and two follow-up contacts, with cognitive function and other covariates assessed at each $j = 1, 2$. Until otherwise stated, assume that all respondents participate at the first follow-up, and thus data $\mathbf{C}_1$ are collected for all subjects at $j = 1$, but some individuals die before the second follow-up, with $S = 1$ indicating survival. Suppose that in addition to cognitive function, $\mathbf{C}_1$ includes all causes of survival (S) that are also correlated with the outcome Y, with the latter defined as cognitive decline between $j = 1$ and $j = 2$. The assumption that we have measured important correlates of S and Y will be key to our developments. As indicated in the diagram of Figure 1, any variable in $\mathbf{C}_1$ may itself contain effects of smoking that simultaneously affect survival and the outcome Y among survivors. Potential examples of exposure-induced correlates of S and Y to consider for $\mathbf{C}_1$ include measures obtained at time $j = 1$ of alcohol consumption, disability score, cognitive score, self-rated health, and smoking status. In addition, as indicated in Figure 1, we will allow survival to be directly affected by A not through $\mathbf{C}_1$, and we will also assume that death is the only source of attrition in this study. Throughout, we assume no measurement error. The double-headed arrow between $\mathbf{C}_1$ and Y encodes possible unmeasured common causes of say cognitive function at follow-up and change in cognitive function Y, the presence of which cannot be ruled out with certainty. For instance, there is evidence for genetic determinants of Alzheimer's disease that suggests a genetic basis for an individual's cognitive function over time [10]. Thus, we would expect a

genetic-induced association between cognitive function measured at time $j = 1$ and cognitive decline between $j = 1$ and $= 2$, because such genetic information was not available for adjustment in this study. Note that while the aforementioned genetic determinant of cognitive decline cannot be directly affected by smoking behavior, other unmeasured common causes of cognitive function and cognitive decline might also include unknown epigenetic effects of smoking behavior and, therefore, could also be directly affected and thus mediate the effects of smoking. For simplicity, we further assume until later sections that all analyses are stratified by pre-exposure confounders of A, which is suppressed in the notation, such that the effects of A on $\left(\mathbf{C}_1, Y, S\right)$ are unconfounded, and therefore, A behaves as if it were randomized.

The causal diagram in Figure 1 can be used to formalize the bias associated with an analysis of the effects of baseline smoking status A on decline of cognitive function Y, conditional on being alive at the end of follow-up. To understand how this bias arises, it suffices to note that by d-separation [9], such an analysis would unblock the noncausal pathways $A \rightarrow S \leftarrow \mathbf{C}_1 \rightarrow Y$ and $A \rightarrow \mathbf{C}_1 \leftrightarrow Y$, thus indicating an effect of smoking on cognitive decline even if there were none. One should also note that further conditioning on $\mathbf{C}_1$ does not resolve the difficulty, because doing so does not block the pathway $A \rightarrow \mathbf{C}_1 \leftrightarrow Y$. Both of these strategies essentially fail because conditioning on S implies conditioning on a collider on the pathway $A \rightarrow S \leftarrow \mathbf{C}_1 \rightarrow Y$ as in the first strategy, but also implies conditioning on a direct descendant of a collider on the pathway $A \rightarrow \mathbf{C}_1 \leftrightarrow Y$, which induces noncausal associations between A and Y [9]. Further conditioning on $\mathbf{C}_1$ as in the second strategy earlier does not really help resolve this issue because it also implies conditioning on a collider on the pathway $A \rightarrow \mathbf{C}_1 \leftrightarrow Y$. Collider bias is invariably the pitfall, and a primary source of bias, for most analyses involving conditioning on a post-exposure event.

2.2. *Nonparametric structural equations model*

The following exposition is framed around a structural equation theory of causal inference, described by Pearl [9]. Structural equations provide a nonparametric algebraic interpretation of the diagram of Figure 1 corresponding to four functions, one for each variable on the causal graph:

$$A = g_A(\varepsilon_A) \tag{1}$$

$$\mathbf{C}_1 = \mathbf{g}_{\mathbf{C}_1}\left(A, \boldsymbol{\varepsilon}_{\mathbf{C}_1}\right) \tag{2}$$

$$S = g_S\left(A, \mathbf{C}_1, \varepsilon_S\right) \tag{3}$$

$$Y = \begin{cases} g_Y\left(A, \mathbf{C}_1, \varepsilon_Y\right) & \text{if } S = 1 \\ \text{undefined} & \text{if } S = 0 \end{cases} \tag{4}$$

Each of the nonparametric functions $\{g_A, \ldots, g_Y\}$ represents a causal mechanism that determines the value of the left variable, known as the output, from variables on the right, known as the inputs [9]. The errors $\left(\varepsilon_A, \boldsymbol{\varepsilon}_{\mathbf{C}_1}, \varepsilon_S, \varepsilon_Y\right)$ stand for all factors not included on the graph that could possibly affect their corresponding outputs when all other inputs are held constant. For instance, ε_S includes all causes of death unrelated to cognitive function decline. To be consistent with the causal graph presented in Figure 1, we require that the errors $\left(\varepsilon_A, \varepsilon_S\right)$ be mutually independent, and we require that they be jointly independent of $\left(\boldsymbol{\varepsilon}_{\mathbf{C}_1}, \varepsilon_Y\right)$. However, as indicated by the double arrow edge in Figure 1, $\boldsymbol{\varepsilon}_{\mathbf{C}_1}$ may not be independent of ε_Y. We allow all error distributions to otherwise remain arbitrary. Lack of a causal effect of a given variable on an output is encoded by an absence of the variable from the right-hand side. For example, the absence of a direct effect of smoking on cognitive function at the first follow-up would imply removing A from the arguments of $\mathbf{g}_{\mathbf{C}_1}$ encoding the assumption that variations in A leave $\mathbf{C}_1$ unchanged, as long as $\boldsymbol{\varepsilon}_{\mathbf{C}_1}$ remains constant, which is consistent with the assumption that there is no unmeasured common cause of smoking and cognitive function.

The last equation makes explicit the fact that Y is observed only among survivors with $(S = 1)$, with corresponding structural equation $g_Y\left(A, \mathbf{C}_1, \varepsilon_Y\right)$. As stated by Pearl [9], the invariance of structural equations permits their use as a basis for modeling causal effects and counterfactuals. In fact, to emulate

the intervention in which one sets $\{A = a\}$ for all individuals simply amounts to replacing the equation for A with $A = a$, producing the following set of modified equations:

$$A = a \tag{5}$$

$$\mathbf{C}_1(a) = \mathbf{g}_{\mathbf{C}_1}(a, \boldsymbol{\varepsilon}_{\mathbf{C}_1}) \tag{6}$$

$$S(a) = g_S\left(a, \mathbf{C}_1(a), \varepsilon_{\mathbf{S}}\right) \tag{7}$$

$$Y(a) = \begin{cases} g_Y\left(a, \mathbf{C}_1(a), \varepsilon_Y\right) & \text{if } S(a) = 1 \\ \text{undefined} & \text{if } S(a) = 0 \end{cases} \tag{8}$$

with $\left(\mathbf{C}_1(a), S(a), Y(a)\right)$ denoting the counterfactual outcomes had smoking status been set to a (possibly contrary to fact). We emphasize that while the model specifies a structural equation for survival, survival is not manipulable and, together with Y, should be understood as part of the outcome produced by the system of equations. As previously observed [11], structural equations are particularly helpful to clarify the difficulty with interpreting the effect of smoking when truncation by death is present. Specifically, we note that the individual effect of smoking is recovered by taking the contrast $Y(a = 1) - Y(a = 0)$, which clearly is defined only for individuals in the principal stratum $\{S(0) = S(1) = 1\}$ and is equal to

$$g_Y\left(1, \mathbf{C}_1(1), \varepsilon_Y\right) - g_Y\left(0, \mathbf{C}_1(0), \varepsilon_Y\right)$$

and the associated population average gives the SACE estimand denoted β:

$$\begin{aligned} \beta &= \mathbb{E}\left\{Y(a = 1) - Y(a = 0) | S(a = 0) = S(a = 1) = 1\right\} \\ &= \mathbb{E}\left\{g_Y\left(1, \mathbf{C}_1(1), \varepsilon_Y\right) - g_Y\left(0, \mathbf{C}_1(0), \varepsilon_Y\right) | S(a = 0) = S(a = 1) = 1\right\} \end{aligned}$$

The SACE is generally not identified without additional assumptions, even under an NPSEM. At one end of the spectrum of possible identifying assumptions, one might assume that the sharp null hypothesis holds that for all individuals in the population, A has no individual causal effect on survival, that is, $S(a = 1) = S(a = 0) = 1$ almost surely. The assumption implies that individuals who survive under an exposure status constitute a random sample of individuals who would survive irrespective of exposure value. Then, it is straightforward to verify that

$$\beta = \mathbb{E}(Y | S = 1, A = 1) - \mathbb{E}(Y | A = 0, S = 1)$$

In light of existing scientific evidence on harmful effects of smoking on human health, the above identifying assumption of no individual causal effect of smoking on survival is clearly inappropriate, and therefore, the aforementioned equation is unlikely to be correct.

At the opposite end of the spectrum of possible identifying assumptions, a strategy that is sometimes adopted entails performing a sensitivity analysis, using data on (A, SY, S) [7, 12–14], possibly also incorporating pre-exposure covariates [9]. A sensitivity analysis then typically involves recovering an estimate of SACE upon making a monotonicity assumption about the effects of exposure on survival and by fixing certain nonidendified parameters involving the joint distribution of potential outcomes to some hypothetical value, which is then varied over a certain range to assess the degree to which the estimate of SACE changes as a function of these parameters. The monotonicity assumption may sometimes be relaxed by introducing additional sensitivity parameters [15, 16]. Worst-case scenarios of a sensitivity analysis give rise to bounds for SACE, but such bounds typically only apply in rather simple settings [3, 17]. For instance, it is not clear whether such bounds can easily be constructed in the presence of numerous baseline and/or time-dependent covariates and with a continuous possibly unbounded outcome.

An alternative identifying assumption that is sometimes made in the principal strata literature and that falls somewhere in between the aforementioned extremes involves assuming that certain potential outcome independencies about the outcome of interest and survival can be obtained upon conditioning

on enough pre-exposure covariates, such that SACE becomes identified within levels of such covariates Hayden *et al.* [18]. Zhang *et al*. [19] proposed to identify SACE under a parametric model using the maximum likelihood estimation. They discussed identifiability for a mixture normal model; however, as pointed out by Ding *et al.* [20], the mixture normal model may not be identifiable under the extreme case that the probability of each latent component is the same. For instance, when the outcome of interest is binary, even with a parametric binary mixture model, the causal parameter of interest is still not identifiable without some further assumptions [20]. An alternative approach is given by Ding *et al.* [20] who exploit a fairly strong form of exclusion restriction for an observed pre-exposure covariate to obtain nonparametric identification of SACE. They essentially require that an association between the covariate in question and survival exists solely because the former is an effect of principal strata and thus cannot be a causal risk factor for death nor can it be affected by a common unmeasured risk factor for death. A variable satisfying these rather stringent conditions may be hard to find in most health-related applications, thus limiting the extent to which their proposed framework may be of practical use. Nolen and Hudgens [21] proposed an elegant randomization-based approach about causal effects within principal strata—possibly within pre-exposure covariate levels—that is particularly useful for testing the null hypothesis of no principal strata causal effect but relies for identification and estimation away from the null, on the assumption of a constant individual principal strata causal effect.

The aforementioned methods all share a notable limitation, in that none appears to appropriately incorporate risk factors of survival available in post-exposure follow-up. In our hypothetical example, it is unclear whether these methods, including sensitivity analysis techniques, can make use of follow-up data collected in $\mathbf{C}_1$ such as post-exposure cognitive function, a correlate of survival that is likely affected by smoking status. A somewhat related context is considered by Dai *et al.* [22], who develop a partially hidden Markov model for time-varying principal stratification in HIV prevential trials. Their proposed approach however relies on a categorization of the intermediate variables into discrete event types and therefore does not easily generalize for continuous or high-dimensional intermediate variables. Tchetgen Tchetgen *et al.* [11] provide an alternative interpretation of standard inverse probability weighting for dependent censoring, incorporating time-updated covariates. They show that under the NPSEM defined by Equations (1–4), the causal effect identified by inverse probability weighting survivors, in fact, naturally incorporates principal strata causal effects therefore formally establishing a previously unknown relation between inverse probability weighting and principal stratification, two seemingly unrelated analytic frameworks. However, their proposed model does not exactly identify SACE without further assumptions. In the next section, an alternative approach is proposed that does not suffer from these possible limitations. But first, we refine the usual definition of SACE to rule out certain pathological situations. As discussed earlier, SACE is typically defined to be the average exposure effect for individuals that would survive irrespective of exposure. One might further refine the definition of SACE by considering a person's survival status $S(a = 1, \mathbf{C}_1(a = 0))$ in the hypothetical situation in which the person smoked, but $\mathbf{C}_1$ behaved as if the individual did not smoke; likewise, one could consider a person's survival status $S(a = 0, \mathbf{C}_1(a = 1))$ in the hypothetical scenario where the person did not smoke, but $\mathbf{C}_1$ behaved as if the person smoked. Such 'cross-world' potential outcomes feature prominently in recent literature on causal mediation analysis, where they are used for formal causal definitions of direct and indirect effects of an exposure [23, 24]. They are formally defined by composition of functions defining the NPSEM each possibly evaluated under different exposure values. Consider the following set of an individual's potential outcomes

$$\left\{ \begin{array}{l} S(a=0) = S(a=0, \mathbf{C}_1(a=0)), S(a=1, \mathbf{C}_1(a=0)) \\ S(a=0, \mathbf{C}_1(a=1)), S(a=1) = S(a=1, \mathbf{C}_1(a=1)) \end{array} \right\}$$

which allows further distinction between individuals who would survive irrespective of exposure, that is, $S(a = 0) = S(a = 1) = 1$. For instance, it may be that a person that would survive whether exposed or not would not survive in certain cross-world situations $S(a = 1, \mathbf{C}_1(a = 0)) = S(a = 0, \mathbf{C}_1(a = 1)) = 0$. Such an individual would be considered rather unusual, and hereafter, the causal effect of exposure for such a person is not further considered, and SACE is redefined to be the causal effect of exposure for individuals that would survive regardless of exposure, including under cross-world situations:

$$\psi = \mathbb{E}\left\{ Y(a=1) - Y(a=0) | S(a, \mathbf{C}_1(a^*)) = 1; a, a^* \in \{0,1\} \right\}$$

If one were to *a priori* rule out the possibility that an individual that would survive irrespective of exposure status could die under cross-world conditions, then it would also be that $\psi = \beta$, and therefore, the

more stringent definition of SACE would match the more common definition. An alternative and perhaps more intuitive condition for this equality is that for every survivor, absence of an individual total effect of exposure on survival implies absence of both an individual direct effect not mediated by $\mathbf{C}_1$ and an individual indirect effect of exposure mediated by $\mathbf{C}_1$, thus ruling out the possibility of countervailing direct and indirect effects resulting in a null individual total effect.

2.3. *Identification of survivor average causal effect*

Identification of SACE requires, in addition to the NPSEM assumptions, that we make the following assumptions:

Monotonicity assumption:

$$S\left(a=1,\mathbf{C}_1(a^*)\right) \leqslant S\left(a=0,\mathbf{C}_1(a^*)\right) \text{ almost surely, } a^*=0,1$$

The contrast $S(a=1,\mathbf{C}_1(a^*))-S(a=0,\mathbf{C}_1(a^*))$ is known in the causal mediation literature as the pure or natural direct effect of exposure, in a hypothetical situation where $\mathbf{C}_1$ is set to what it would be under smoking status a^*, and captures the direct effects of exposure not mediated by $\mathbf{C}_1$. Thus, the assumption states that there is no individual in the population, for whom smoking provides a protective individual direct effect on survival.

In order to state the second assumption, consider the following subsets of individuals. Let $\mathcal{P}_0$ denote the subset of individuals that would survive regardless of smoking, in a hypothetical situation where $\mathbf{C}_1$ would behave as if they did not smoke, that is, $S(a=0,\mathbf{C}_1(a=0))=S(a=1,\mathbf{C}_1(a=0))=1$, and let $\mathcal{P}_1$ denote the subset of individuals that would survive irrespective of smoking, in a hypothetical situation where $\mathbf{C}_1$ would behave as if they smoked, that is, $S(a=0,\mathbf{C}_1(a=1))=S(a=1,\mathbf{C}_1(a=1))=1$.

Concordant survivorship assumption: $\mathcal{P}_0=\mathcal{P}_1$ almost surely.

This second assumption states that individuals that would survive irrespective of smoking status, in a hypothetical situation where $\mathbf{C}_1$ behaved as if they smoked, would also survive irrespective of smoking status, in the hypothetical situation where $\mathbf{C}_1$ behaved as if they did not smoke and that the converse also holds. In terminology used in causal mediation analysis, the assumption states that there is no survivor for whom the natural direct effect of A on S is null when $\mathbf{C}_1$ is held at the value it would have been under no exposure, and yet the natural direct effect when $\mathbf{C}_1$ is held at the value it would have been under active exposure is non-null. Under the assumption, the converse is also ruled out. A sufficient condition for the assumption is that A and $\mathbf{C}_1$ do not interact at the individual level (on the additive scale) in causing S, however the assumption may still hold even if this were not the case. Another interesting special case is if $\mathbf{C}_1$ intercepts all causal pathways between A and S, then the assumption (and monotonicity) holds trivially because $S(a=0,\mathbf{C}_1(a^*))-S(a=1,\mathbf{C}_1(a^*))=0$ for all a^*. This latter situation highlights the importance of including in $\mathbf{C}_1$ mediating factors of the effects A on S, which may potentially interact with exposure.

We are now ready to state our identification result:

Theorem 1
Under the NPSEM given by Equations (1)–(4), and under the monotonicity assumption, and the concordant survivorship assumption, we have that SACE :

$$\psi=\mathbb{E}\left\{Y(a=1)-Y(a=0)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}$$

is nonparametrically identified and is given by $\mu_1-\mu_0$, where

$$\begin{aligned}\mu_1&=\mathbb{E}\left\{Y(a=1)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}\\&=\mathbb{E}\left\{Y(a=1)|S(a=0,\mathbf{C}_1(a=1))=S(a=1,\mathbf{C}_1(a=1))=1\right\}\\&=\mathbb{E}\left\{Y|A=1,S=1\right\}\end{aligned}$$

and

$$\begin{aligned}\mu_0 &= \mathbb{E}\left\{Y(a=0)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}\\ &= \mathbb{E}\left\{Y(a=0)|S(a=0,\mathbf{C}_1(a=0))=S(a=1,\mathbf{C}_1(a=0))=1\right\}\\ &= \frac{\int \mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}\right)\Pr(S=1|A=1,\mathbf{C}_1=\mathbf{c})dF(\mathbf{c}|A=0)}{\int \Pr(S=1|A=1,\mathbf{C}_1=\mathbf{c})dF(\mathbf{c}|A=0)}\end{aligned} \tag{9}$$

where $F(u_1|u_2)$ stands for the CDF of U_1 given U_2, evaluated at $U_1=u_1$, $U_2=u_2$, and it is assumed that $\Pr(S=1,\mathbf{C}_1=\mathbf{c}|A=1)/\Pr(S=1,\mathbf{C}_1=\mathbf{c}|A=0)<\infty$.

Proofs of all theorems are given in the Appendix. According to the theorem, estimation of the average survivor outcome for smoking presents no particular difficulty and can be obtained by the simple average outcome for the exposed individuals who survived. The situation is quite different for the average survivor outcome for the nonsmoking exposure status. The theorem states that this average can be obtained by using the expression given in Equation (9). Intuition is gained by comparing this expression with that of the average outcome for unexposed individuals who actually survived:

$$\mathbb{E}(Y|A=0,S=1)=\frac{\int \mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}\right)\Pr(S=1|A=0,\mathbf{C}_1=\mathbf{c})dF(\mathbf{c}|A=0)}{\int \Pr\left(S=1|A=0,\mathbf{C}_1=\mathbf{c}\right)dF(\mathbf{c}|A=0)}$$

Then, one may note that the primary distinction between these two expressions is that the conditional survival probability for the unexposed used in both numerator and denominator of the second expression is replaced by that for the exposed in the first expression. This substitution essentially amounts to a form of standardization of unexposed individuals who survived, by the survival probability of exposed individuals with similar covariate history. One should note as well that in the special instance where S is independent of A given $\mathbf{C}_1$, as one might expect, the two aforementioned expressions coincide. However, in general, the two functionals do not coincide, and the subtle difference between them has nontrivial implications for inference. Specifically, whereas estimation of $\mathbb{E}(Y|A=1,S=1)$ is straightforward and does not require estimating nuisance parameters, estimation of μ_0 is somewhat more involved.

2.4. *Estimation of survivor average causal effect*

As explained in the previous section, we only need to consider estimation of μ_0. Ideally, we may wish to estimate the latter nonparametrically, so as to avoid potential modeling bias; however, this may not be possible in practice. This is because, as shown below, estimation of μ_0 is generally not possible without involving an estimate of a subset of the following quantities $\{\mathbb{E}\left(Y|A,S=1,\mathbf{C}_1\right),\Pr(S=1|A,\mathbf{C}_1),dF(\mathbf{C}_1|A)\}$. In practice, one would probably seek to enrich to the extent possible the set of covariates in $\mathbf{C}_1$ in order to ensure that all variables are included, which potentially mediate the effects of exposure on survival and outcome simultaneously. As a result, our primary interest concerns settings in which $\mathbf{C}_1$ potentially includes a large number of covariates, a subset of which is possibly continuous, such that nonparametric methods for estimating the aforementioned density and regression models, such as smoothing techniques, may be of limited value. Consequently, next, we present three simple estimation strategies based on low-dimensional models. Let $\{\widehat{\mathbb{E}}\left(Y|A,S=1,\mathbf{C}_1\right),\widehat{\Pr}(S=1|A,\mathbf{C}_1),d\widehat{F}(\mathbf{C}_1|A)\}$ denote estimates obtained using parsimonious parametric working models for the unknown conditional mean and the two unknown conditional densities. Our first strategy entails direct substitution of unknown quantities in (9) by their corresponding estimate, which gives

$$\widehat{\mu}_0^1=\frac{\int \widehat{\mathbb{E}}\left(Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}\right)\widehat{\Pr}(S=1|A=1,\mathbf{C}_1=\mathbf{c})d\widehat{F}(\mathbf{c}|A=0)}{\int \widehat{\Pr}(S=1|A=1,\mathbf{C}_1=\mathbf{c})d\widehat{F}(\mathbf{c}|A=0)}$$

This estimator depends heavily on correct specification of all three models. An alternative estimator that makes fewer assumptions is based on the following equivalent representation of μ_0

$$\frac{\int \mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}\right)\Pr\left(S=1|A=1,\mathbf{C}_1=\mathbf{c}\right)dF(\mathbf{c}|A=0)}{\int \Pr\left(S=1|A=1,\mathbf{C}_1=\mathbf{c}\right)dF(\mathbf{c}|A=0)}$$
$$=\frac{\mathbb{E}\left\{\mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1\right)ASdF(\mathbf{C}_1|A=0)/dF\left(\mathbf{C}_1|A=1\right)\right\}}{\mathbb{E}\left\{ASdF\left(\mathbf{C}_1|A=0\right)/dF\left(\mathbf{C}_1|A=1\right)\right\}}$$

which gives the estimator

$$\widehat{\mu}_0^2=\frac{\mathbb{P}_n\left[\widehat{\mathbb{E}}\left(Y|A=0,S=1,\mathbf{C}_1\right)AS\left\{d\widehat{F}(\mathbf{C}_1|A=0)/d\widehat{F}(\mathbf{C}_1|A=1)\right\}\right]}{\mathbb{P}_n\left[AS\left\{d\widehat{F}(\mathbf{C}_1|A=0)/d\widehat{F}(\mathbf{C}_1|A=1)\right\}\right]}$$

where $\mathbb{P}_n(\cdot)=n^{-1}\sum_i(\cdot)_i$. This second approach improves over the first in terms of robustness, because it does not directly involve an estimate of the survival process. Finally, consider yet another representation of μ_0 :

$$\frac{\int \mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}\right)\Pr\left(S=1|A=1,\mathbf{C}_1=\mathbf{c}\right)dF(\mathbf{c}|A=0)}{\int \Pr\left(S=1|A=1,\mathbf{C}_1=\mathbf{c}\right)dF(\mathbf{c}|A=0)}$$
$$=\frac{\mathbb{E}\left\{Y(1-A)S\Pr\left(S=1|A=1,\mathbf{C}_1\right)/\Pr\left(S=1|A=0,\mathbf{C}_1\right)\right\}}{\mathbb{E}\left\{(1-A)S\Pr\left(S=1|A=1,\mathbf{C}_1\right)/\Pr\left(S=1|A=0,\mathbf{C}_1\right)\right\}} \quad (10)$$

which motivates the estimator

$$\widehat{\mu}_0^3=\frac{\mathbb{E}\left\{Y(1-A)S\widehat{\Pr}\left(S=1|A=1,\mathbf{C}_1\right)/\widehat{\Pr}\left(S=1|A=0,\mathbf{C}_1\right)\right\}}{\mathbb{E}\left\{(1-A)S\widehat{\Pr}\left(S=1|A=1,\mathbf{C}_1\right)/\widehat{\Pr}(S=1|A=0,\mathbf{C}_1)\right\}}$$

This last estimator has the advantage that it only requires an estimate of the survival process, and both the outcome regression and the covariates distribution are left unrestricted. From the three approaches presented above earlier, the last is most appealing, because it involves fitting a single model, whereas the other two approaches involve multiple models. In the next section, we show that the last strategy readily extends to a general longitudinal design with arbitrary follow-up. We do not further consider the first two strategies on above grounds, although in Section 5, a doubly robust approach is given, which combines all three mentioned strategies such that consistent estimation of μ_0 remains possible even under partial model misspecification, that is, when only some but not all required models are correctly specified.

3. Longitudinal studies of arbitrary length

3.1. Longitudinal nonparametric structural equations model and identification of survivor average causal effect

We turn to the more general context of a longitudinal study with arbitrary follow-up $j=0,\ldots J$, with $J\geqslant 2$ is fixed, and at each occasion j, one observes $\left(S_j,S_j\mathbf{C}_j\right)$, where S_j indicates survival status at time j, and $\mathbf{C}_j$ includes covariates measured at time j. We suppose that $S_0=S_1=1$, and therefore, a vector of pre-exposure covariates $\mathbf{C}_0$ is measured on all individuals in the target population, and exposure A is measured concurrently with covariates $\mathbf{C}_1$ on all individuals in the target population. The variable $\mathbf{C}_J=Y$ encodes the outcome measured at the end of follow-up. We consider the general NPSEM:

$$\begin{aligned}&\text{For } j=0\\&\quad\mathbf{C}_0=\mathbf{g}_{\mathbf{C}_0}\left(\boldsymbol{\varepsilon}_{\mathbf{C}_0}\right)\end{aligned} \quad (11)$$

$$\begin{aligned}&\text{for } j=1\\&\quad\begin{cases}\mathbf{C}_1=\mathbf{g}_{\mathbf{C}_1}\left(\boldsymbol{\varepsilon}_{\mathbf{C}_1},\mathbf{C}_0\right)\\A=g_A\left(\varepsilon_A,\mathbf{C}_0\right)\end{cases}\end{aligned} \quad (12)$$

and for $j = 2, \ldots, J$

$$S_j = \begin{cases} g_{S_j}\left(A, \overline{\mathbf{C}}_{j-1}, \varepsilon_{S_j}\right) & \text{if } S_{j-1} = 1 \\ 0 & \text{if } S_{j-1} = 0 \end{cases} \tag{13}$$

$$\mathbf{C}_j = \begin{cases} \mathbf{g}_{\mathbf{C}_j}\left(A, \overline{\mathbf{C}}_{j-1}, \boldsymbol{\varepsilon}_{\mathbf{C}_j}\right) & \text{if } S_j = 1 \\ \text{undefined} & \text{if } S_j = 0 \end{cases} \tag{14}$$

We assume that

$$\varepsilon_A \perp\!\!\!\perp \left\{\varepsilon_{s_j} : j = 2, \ldots, J\right\}$$

and we also assume that

$$\left\{\varepsilon_A, \varepsilon_{s_j} : j = 2, \ldots, J\right\} \perp\!\!\!\perp \left\{\boldsymbol{\varepsilon}_{\mathbf{C}_j} : j \geqslant 2\right\}$$

However, $\boldsymbol{\varepsilon}_{\mathbf{C}_j}$ and $\boldsymbol{\varepsilon}_{\mathbf{C}_{j'}}$ may be dependent, and ε_{s_j} and $\varepsilon_{s_{j'}}$ may be dependent, $j \neq j'$. The causal diagram of Figure 2 depicts the observed data, generated under such an NPSEM for an individual alive at the end of follow-up, in the special case where $J = 4$. We allow all error distributions to otherwise remain arbitrary.

This more general NPSEM extends the previous model, to accommodate, both confounding by pre-exposure covariates $\mathbf{C}_0$, and longitudinal data $\overline{\mathbf{C}}_J$, where $\overline{\mathbf{C}}_j$ denotes the history $\left(\mathbf{C}_0, \ldots, \mathbf{C}_j\right)$. Note that because $\mathbf{C}_0$ and $\mathbf{C}_1$ are respectively prior to and concurrent with exposure A (now defined to occur at time $j = 1$ to ensure proper temporal ordering with $\mathbf{C}_0$), they cannot be affected by exposure, and thus $\mathbf{C}_0(a) = \mathbf{C}_0$ and $\mathbf{C}_1(a) = \mathbf{C}_1$. Technically, $\mathbf{C}_0$ confounds the effects of A, but $\mathbf{C}_1$ is not considered a confounder even though it may be correlated with A and may be used to account for survival bias. Crucially, independence of $\boldsymbol{\varepsilon}_{\mathbf{C}_{j-1}}$ and ε_{s_j} implies that for individuals alive at time $j-1$, $\left(\overline{\mathbf{C}}_{j-2}, A\right)$ intercepts or blocks all noncausal pathways between $\mathbf{C}_{j-1}$ and S_j; in the language of causal graphs, $\left(\overline{\mathbf{C}}_{j-2}, A\right)$ is said to block all back-door paths from $\mathbf{C}_{j-1}$ to S_j.

The SACE is defined to be the causal effect of exposure on an outcome measured at the end of follow-up, among individuals that would survive whether exposed or not, and with the covariate history, they would have under possibly conflicting exposure status:

$$\psi_J = \mathbb{E}\left\{Y(a=1) - Y(a=0) | S_J\left(a, \overline{\mathbf{C}}_{J-1}(a^*)\right) = 1;\ a, a^* \in \{0,1\}\right\}$$

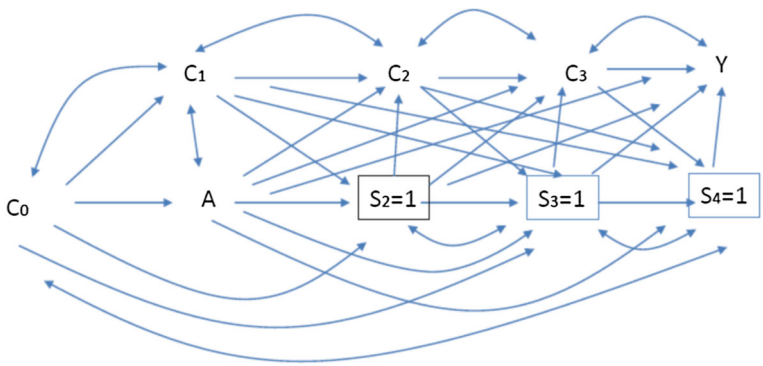

Figure 2. Causal diagram associated with the nonparametric structural equations model Equations (11)–(14) at $J = 4$.

As was the case in the previous section with three time points, we likewise have that if $S_J(0,\overline{\mathbf{C}}_{J-1}(0))S_J(1,\overline{\mathbf{C}}_{J-1}(1))=1$ implies $S_J(0,\overline{\mathbf{C}}_{J-1}(1))S_J(1,\overline{\mathbf{C}}_{J-1}(0))=1$ almost surely, then

$$\psi_J=\mathbb{E}\left\{Y(a=1)-Y(a=0)|S_J\left(a,\overline{\mathbf{C}}_{J-1}(a)\right)=S_J(a)=1;\ a\in\{0,1\}\right\}$$

which corresponds to the traditional definition of SACE.

Identification requires an extension of the monotonicity and concordant survivorship assumptions:

Sequential monotonicity assumption for point exposure:

$$\text{if } S_{j-1}\left(1,\overline{\mathbf{C}}_{j-2}(a)\right)S_{j-1}\left(0,\overline{\mathbf{C}}_{j-2}(a)\right)=1 \text{ then } S_j\left(1,\overline{\mathbf{C}}_{j-1}(a)\right)\leqslant S_j\left(0,\overline{\mathbf{C}}_{j-1}(a)\right)$$
$$\text{almost surely; } a=0,1,\ j=1,\ldots,J$$

Let $\mathcal{P}_{0,J}$ denote the subset of individuals that would survive until the end of follow-up regardless of smoking, in a hypothetical situation where $\overline{\mathbf{C}}_J$ would behave as if they did not smoke, and let $\mathcal{P}_{1,J}$ denote the subset of individuals that would survive irrespective of smoking, in a hypothetical situation where $\overline{\mathbf{C}}_J$ would behave as if they smoked.

Concordant survivorship assumption for point exposure:

$$\mathcal{P}_{0,J}=\mathcal{P}_{1,J} \text{ almost surely}$$

The sequential monotonicity assumption states that for a person that would survive up to time $j-1$ irrespective of smoking status, in the hypothetical situation in which his covariate history behaves as if smoking status were fixed to a, if the person were to survive at time j when exposed and with covariate history as if smoking status were equal to a, the person would also survive at time j, with similar covariate history, if he were not to smoke. The concordant survivorship assumption essentially states that

$$S_J\left(1,\overline{\mathbf{C}}_{J-1}(0)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(0)\right)=1\iff S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(1)\right)=1 \text{ almost surely}$$

which is the natural extension of the previous concordant survivorship assumption. Let

$$\begin{aligned}
\pi_j\left(a,\overline{\mathbf{c}}_{j-1}\right)&=\Pr\left(S_j=1|A=a,\overline{\mathbf{C}}_{j-1}=\overline{\mathbf{c}}_{j-1},S_{j-1}=1\right),\ j=2,\ldots J\\
G_{a,j}\left(\mathbf{c}_j;\overline{\mathbf{c}}_{j-1}\right)&=F\left(\mathbf{c}_j|A=a,\overline{\mathbf{C}}_{j-1}=\overline{\mathbf{c}}_{j-1},S_j=1\right),\ j=2,\ldots,J-1\\
G_{0,1}\left(\mathbf{c}_1;\overline{\mathbf{c}}_0\right)&=G_{1,1}\left(\mathbf{c}_1;\overline{\mathbf{c}}_0\right)=F\left(\mathbf{c}_1|\mathbf{c}_0\right)\\
G_{0,0}\left(\mathbf{c}_0;\overline{\mathbf{c}}_{-1}\right)&=G_{1,0}\left(\mathbf{c}_0;\overline{\mathbf{c}}_{-1}\right)=F(\mathbf{c}_0)
\end{aligned}$$

Throughout, we further assume that

$$\prod_{k=2}^{J}\pi_k\left(1,\overline{\mathbf{C}}_{k-1}\right)G_{1,k-1}\left(\mathbf{C}_{k-1};\overline{\mathbf{C}}_{j-2}\right)/\prod_{k=2}^{J}\pi_k\left(0,\overline{\mathbf{C}}_{k-1}\right)G_{0,k-1}\left(\mathbf{C}_{k-1};\overline{\mathbf{C}}_{k-2}\right)<\infty$$

almost surely, where π_k is defined in Theorem 2, and $0<p\left(\mathbf{C}_0\right)<1$ almost surely.

Then, we have the following result.

Theorem 2

Under the NPSEM given by Equations (11)–(14), the sequential monotonicity assumption, and the concordant survivorship assumption for point exposure, we have that ψ_J is nonparametrically identified and is given by $\mu_{1,J}-\mu_{0,J}$, where

$$\begin{aligned}
\mu_{1,J}&=\mathbb{E}\left\{Y(a=1)|S_J\left(a,\overline{\mathbf{C}}_{J-1}\left(a^*\right)\right)=1;\ a,a^*\in\{0,1\}\right\}\\
&=\mathbb{E}\left(Y|A=1,S_J=1\right)
\end{aligned}$$

and

$$\mu_{0,J} = \mathbb{E}\left\{ Y(a=0) | S_J\left(a, \overline{\mathbf{C}}_{J-1}(a^*)\right) = 1;\ a, a^* \in \{0,1\} \right\}$$
$$= \frac{\int \ldots \int \mathbb{E}\left(Y | A=0, S_J=1, \overline{\mathbf{C}}_{J-1} = \overline{\mathbf{c}}_{J-1}\right) \prod_{j=2}^{J} \pi_j\left(1, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=0}^{J-1} dG_{0,j}\left(\mathbf{c}_j; \overline{\mathbf{c}}_{j-1}\right)}{\int \ldots \int \prod_{j=2}^{J} \pi_j\left(1, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=0}^{J-1} dG_{0,j}\left(\mathbf{c}_j; \overline{\mathbf{c}}_{j-1}\right)} \quad (15)$$

According to the theorem, as was the case in the three-occasion example, estimation of the survivor average outcome for smoking likewise presents no particular difficulty and can be achieved by a simple average of the outcome of exposed individuals who survived until the end of follow-up. The theorem states that the survivor average outcome for nonsmoking is given by the functional in Equation (15). This formula is a generalization of Equation (9) accounting for arbitrary follow-up and pre-exposure confounding by $\mathbf{C}_0$. Instead of a marginal SACE parameter, a conditional SACE parameter might be of interest, say

$$\psi_J(v) = \mathbb{E}\left\{ Y(a=1) - Y(a=0) | \mathbf{V} = \mathbf{v}, S_J\left(a, \overline{\mathbf{C}}_{J-1}(a^*)\right) = 1;\ a, a^* \in \{0,1\} \right\}$$

where $\mathbf{V} \subset \mathbf{C}_0$ is a vector of pre-exposure confounders, such that $\mathbf{C}_0 = (\mathbf{V}, \mathbf{W})$ with $\mathbf{W}$ all components of $\mathbf{C}_0$ not in $\mathbf{V}$. Then, it is straightforward to verify that a corollary of Theorem 2 gives under the same set of assumptions:

$$\psi_J(\mathbf{v}) = \mu_{1,J}(\mathbf{v}) - \mu_{0,J}(\mathbf{v})$$

where

$$\mu_{1,J}(\mathbf{v}) = \mathbb{E}\left\{ Y | A=1, \mathbf{V}=\mathbf{v}, S_J = 1 \right\}$$
$$\mu_{0,J}(v) = \frac{\int \ldots \int \mathbb{E}\left(Y | A=0, S_J=1, \overline{\mathbf{C}}_{J-1} = \overline{\mathbf{c}}_{J-1}\right) \prod_{j=2}^{J} \pi_j\left(1, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=1}^{J-1} dG_{0,j}\left(\mathbf{c}_j; \overline{\mathbf{c}}_{j-1}\right) dG_{0,0}(\mathbf{w}; \mathbf{v})}{\int \ldots \int \prod_{j=2}^{J} \pi_j\left(1, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=1}^{J-1} dG_{0,j}\left(\mathbf{c}_j; \overline{\mathbf{c}}_{j-1}\right) dG_{0,0}(\mathbf{w}; \mathbf{v})}$$

and

$$G_{0,0}(\mathbf{w}; \mathbf{v}) = \begin{cases} F(\mathbf{w}|\mathbf{v}) \text{ if } \mathbf{W} \text{ is not empty} \\ 1 \text{ if } \mathbf{W} \text{ is empty} \end{cases}$$

3.2. Weighted estimation of models of survivor average causal effect

A simple weighted approach is given, which is consistent in the absence of model misspecification. Let $\left\{\widehat{\pi}_j(\cdot), j = 2, \ldots J\right\}$ denote a fitted model for the survival process, using a standard parametric approach; likewise, let $p(\mathbf{C}_0) = \Pr(A=1|\mathbf{C}_0)$ denote the propensity score for exposure, and $\widehat{p}(\mathbf{C}_0)$ its corresponding estimator. Define the following estimated weights for individuals who survive until the end of follow-up:

$$\widehat{W}_S = \frac{\prod_{k=2}^{J} \widehat{\pi}_k\left(1, \overline{\mathbf{C}}_{k-1}\right)}{\prod_{k=2}^{J} \widehat{\pi}_k\left(A, \overline{\mathbf{C}}_{k-1}\right)}$$
$$\widehat{W}_A = \left[\widehat{p}\left(\mathbf{C}_0\right)^A \left\{1 - \widehat{p}\left(\mathbf{C}_0\right)\right\}^{1-A}\right]^{-1}$$

Then, let $\widehat{\psi}_J$ denote the weighted ordinary least-squares estimator of the marginal effect of A on Y using only data on individuals who survived until the end of follow-up, with individual weight equal to $\widehat{W}_S \times \widehat{W}_A$. Then, we have the following result:

Theorem 3
Suppose that the assumptions of Theorem 2 hold and that $\{\widehat{\pi}_j(\cdot), j = 2, \ldots J\}$ and $\widehat{P}(\cdot)$ are consistent, then $\widehat{\psi}_J$ is consistent for ψ_J.

The theorem states that ψ_J may be estimated consistently, by weighted least-squares estimation of a regression of Y on A among survivors at time J, with weight $\widehat{W}_S \times \widehat{W}_A$. Furthermore, the estimator $\widehat{\psi}_J$ is asymptotically normal under standard regularity conditions. The weights component given by $\widehat{W}_A$ corresponds to standard inverse probability-of-treatment weighting a well-known propensity score technique to control for confounding [25, 26]. Intuitively, treatment weights create possibly fractional copies of each individual with complete follow-up, such that in the weighted sample, $\mathbf{C}_0$ no longer predicts A and therefore is not a confounder. The other component of the weights $\widehat{W}_S$ corrects for selective survival of unexposed individuals. Intuitively, monotonicity of the effects of exposure on survival implies that unexposed survivors may be over-represented relative to the exposed, and thus if $\mathbf{C}_0$ were empty so that $\widehat{W}_A$ could be set to 1, then because $0 < W_S = \prod_{k=2}^{j} \pi_k\left(1, \overline{\mathbf{C}}_{k-1}\right) / \prod_{k=2}^{j} \pi_k\left(A, \overline{\mathbf{C}}_{k-1}\right) \leqslant 1$, we would have that the survival weight essentially adjusts the contribution of unexposed survivors downwards and does so continuously as a function of $\overline{\mathbf{C}}_J$.

Although the theorem identifies SACE on the additive scale, estimation using $\widehat{W}_S \times \widehat{W}_A$ as weight is a universal strategy for estimating SACE on a variety of scales. For instance, the approach could be used to estimate SACE on the multiplicative scale, or for other choice of link function, such as logit or probit link functions. This could be achieved by simply replacing the normal equations with the corresponding set of estimating equations one would have used in the absence of selective survivorship and by multiplying each survivor's contribution by the weight. The approach could also be used for quantile regression or for weighting other standard likelihood or quasi-likelihood methods. Finally, suppose that instead of marginal causal effects, a conditional causal effect, say $\psi_J(\mathbf{C}_0)$, was in view. Then, suppose that one were to use a parametric or semiparametric model $\mathbb{E}\left\{Y(a)|\mathbf{C}_0 = \mathbf{c}_0, S_J(a', \overline{\mathbf{C}}_{J-1}(a^*)) = 1;\ \text{for all } a', a^* \in \{0, 1\}; \theta\right\}$ to describe the survivor average causal effect of A within levels of $\mathbf{C}_o$. Then, assuming the model was correct, θ could then be estimated via a weighted approach, using only the survival component of the weight $\widehat{W}_S$, because the regression model would already account for baseline confounding by conditioning on $\mathbf{C}_0$. This strategy is in fact adopted in the data illustration of the next section.

For inference about ψ_J or $\psi_J(\mathbf{C}_0)$, in general, one could use the nonparametric bootstrap such that the extra variation due to first-stage estimation of the weights is appropriately accounted for. Alternatively, one could use a consistent estimate of the large sample variance of the weighted estimator of SACE to construct Wald-type CIs; such an estimator of the large sample variance can be computed in a manner similar to the variance estimator given in Section 5.

4. A data application

We illustrate the new methodology in an evaluation of the effects of smoking on cognitive decline in an aging population subject to substantial attrition due to death and dropout for other reasons [27]. In their paper, Weuve *et al.* [27] noted that selective attrition in this population may introduce bias into analyses of the effects of smoking status measured at the start of follow-up on cognitive decline, for the following reasons.

(1) An individual's evolving health status is likely to be a common cause for attrition and cognitive decline among survivors who do not drop out.
(2) An individual's evolving health status is likely to mediate the causal effect of smoking on cognitive decline.

To account for (1) and (2), Weuve *et al.* used inverse probability-of-attrition weights and examined the influence of selective attrition on the estimated association of current smoking (versus never smoking) with cognitive decline in participants of the Chicago Health and Aging Project (n = 3713), aged 65–109 years, who were current smokers or never smokers, and underwent cognitive assessments up to five times at 3-year intervals. Only 20% of the original sample remained at the fourth follow-up, and mortality accounted for most (~70%) of the attrition. Weuve *et al.* used separate pooled logistic regressions to fit predictive models of attrition due to death or study dropout across follow-up visits using both baseline and time-updated data to construct inverse probability-of-attrition weights. We refer the

reader to Weuve *et al.* for additional details on their design and analysis of the study, also see [11, 24] for additional discussion. Similar to Weuve *et al.* [27], we estimated a linear mean regression model contrasting rates of change in cognitive scores in current versus never smokers, adjusting for the following pre-exposure confounders in the regression: age, sex, race, education, and alcohol consumption. As recommended by Tchetgen Tchetgen *et al.* [28], we assumed an independence correlation structure for the five serial measurements of cognitive function (coded as z-scores). The two main sources of attrition, death and dropout for other reasons (denoted D_j), were modeled separately as discrete-time PH models via pooled logistic regression. Each model included main effects for the following baseline and time-updated variables: age, race (African American versus White), sex (male versus female), education (0–8, 9–12 referent, 13–16, and 17–30 years), alcohol consumption at the previous visit (none referent, up to 1 drink/day, 1 drink/day), social network score at the previous visit, cognitive activity at the previous visit, disability score at the previous visit, self-rated health at the previous visit (per unit worsening in rating), chronic cardiovascular conditions, diabetes, global cognitive score at the previous visit, and smoking status (current versus never). A logistic model for nondeath-related censoring was also estimated using only baseline variables. These predictive models were combined as in the study of Weuve *et al.* to account for selective censoring other than death via stabilized weights [27]:

$$\widehat{W}_j^{\cdot\cdot} = \frac{\prod_{k=2}^{j} \widehat{\Pr}\left(D_k = 0 | \mathbf{C}_0, A, S_k = 1\right)}{\prod_{k=2}^{j} \widehat{\Pr}\left(D_k = 0 | \overline{\mathbf{C}}_{k-1}, A, S_k = 1\right)}$$

An additional set of weights was estimated to account for truncation by death using the approach developed in Section 3.2

$$\widehat{W}_{S,j} = \frac{\prod_{k=2}^{j} \widehat{\pi}_k\left(1, \overline{\mathbf{C}}_{k-1}\right)}{\prod_{k=2}^{j} \widehat{\pi}_k\left(A, \overline{\mathbf{C}}_{k-1}\right)}$$

and the final weight $\widehat{W}_j^{\cdot\cdot} \times \widehat{W}_{S,j}$ was applied at the level of observations within individuals, such that for each person's contribution to our analysis at time j, the weight was the product of censoring weights and survival weights.

The CIs were obtained via the bootstrap. In unweighted analyses, current smokers' cognitive scores declined 0.11 standard units per decade more rapidly than never smokers' cognitive scores (95% CI = −0.20 to −0.02). Weighting for attrition due to dropout or death using $\widehat{W}_j^{\cdot\cdot} \times \widehat{W}_{S,j}$ for weight gave an estimate that was considerably larger, with smoking's estimated 10-year rate of decline compared with nonsmoking 55% larger than in the unweighted analysis (95% CI = −0.27 to −0.07). Under the assumptions of Theorem 2, this latter estimate may be interpreted as the survivor average causal effect of smoking on cognitive decline conditional on pre-exposure covariates. Monotonicity in the current setting essentially states that smoking does not offer any direct individual survival benefit not mediated by the time-varying factors included in our analysis. We emphasize that the assumption does not rule out that there may be certain pathways through which smoking could offer survival benefits (e.g., by alleviating depressive symptoms in an individual with suicidal tendencies); however, according to the assumption, such pathways are assumed to be outweighed for each individual by the harmful health effects associated with smoking through other pathways. We find this assumption to be credible given the overwhelming amount of evidence of the harmful effects of smoking. The concordant survivorship assumption here essentially states that survivors who experienced no natural direct effect of smoking (not mediated by time-updated factors $\overline{\mathbf{C}}_{J-1}$) would do so irrespective of the reference exposure value used in defining the direct effect. This assumption may be reasonable, if one can safely assume that $\overline{\mathbf{C}}_{J-1}$ captures the most important pathways by which smoking affects survival and that the remaining pathways not captured by $\overline{\mathbf{C}}_{J-1}$ do not have significant additive interactions with the latter. Our interpretation of our estimated effect as identifying SACE is thus contingent on the extent to which this last assumption may be reasonable. Finally, we should note that similar results were obtained for SACE when dropout for other reasons was simply ignored by redefining the weight as $\widehat{W}_{S,j}$, suggesting that most of the selection bias due to attrition was related to death.

5. Results on double robustness and sensitivity analysis

5.1. Double robustness

Consider the simple three-occasion setting described in Section 2. The following result gives a doubly robust estimator of μ_0 in the three-occasion setting, which essentially combines $\widehat{\mu}_0^2$ and $\widehat{\mu}_0^3$ of Section 2.4, such that consistency is obtained under a union model where either $\widehat{\mathbb{E}}\left(Y|A=0,S=1,\mathbf{C}_1\right)$ and $\widehat{dF}(\mathbf{C}_1|A=0)$ are both consistent or $\widehat{\Pr}(S=1|A=0,\mathbf{C}_1)$ is consistent, but all models are not necessarily consistent. To state the result, consider the following estimating function:

$$\begin{aligned}U(\mu_0) = (1-A)S\frac{\Pr(S=1|A=1,\mathbf{C}_1)}{\Pr(S=1|A=0,\mathbf{C}_1)}\left\{Y-\mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1\right)\right\}\\ +A\mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1\right)\frac{dF(\mathbf{C}_1|A=0)}{dF(\mathbf{C}_1|A=1)}\left\{S-\Pr\left(S=1|A=1,\mathbf{C}_1\right)\right\}\\ +\left\{(1-A)\mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1\right)\Pr\left(S=1|A=1,\mathbf{C}_1\right)-\mu_0\right\}\end{aligned}$$

and define $\widehat{U}(\mu_0)$ similarly, evaluated under $\{\widehat{\mathbb{E}}\left(Y|A,S=1,\mathbf{C}_1\right),\widehat{\Pr}(S=1|A,\mathbf{C}_1),\widehat{dF}(\mathbf{C}_1|A)\}$.

Theorem 4
Under the assumptions of Theorem 1, $\widehat{\mu}_0^{dr}$ is doubly robust and therefore converges to μ_0 and is asymptotically normal if one but not necessarily both of the following conditions hold:

(1) $\widehat{\mathbb{E}}\left(Y|A,S=1,\mathbf{C}_1\right)$ and $\widehat{dF}(\mathbf{C}_1|A=0)/\widehat{dF}(\mathbf{C}_1|A=1)$ are both consistent.
(2) $\widehat{\Pr}(S=1|A,\mathbf{C}_1)$ is consistent.

where $\widehat{\mu}_0^{dr}$ satisfies the estimating equation $\mathbb{P}_n\left\{\widehat{U}(\widehat{\mu}_0^{dr})\right\}=0$. Furthermore, at the intersection submodel where all estimators are consistent, $\widehat{\mu}_0^{dr}$ is semiparametric efficient in the nonparametric model where no model assumption is made, at the intersection submodel where both of the aforementioned conditions (1) and (2) hold.

The theorem gives an estimator of μ_0 that is doubly robust and semiparametric efficient in the nonparametric model where no modeling assumption is made, at the intersection submodel where all working models are correct. This last property is sometimes called semiparametric local efficiency. At the intersection submodel, the asymptotic variance of $\widehat{\mu}_0^{dr}$ can be estimated by the simple expression $\mathbb{P}_n\left\{\widehat{U}(\widehat{\mu}_0^{dr})^2\right\}^{-1}$. Interestingly, one may note that this expression is invariant to the choice of working models and corresponding estimators. This property does not apply outside of the intersection submodel; nonetheless, it remains possible to estimate the asymptotic variance of $\widehat{\mu}_0^{dr}$ outside the intersection submodel. To do so, let $\widehat{\boldsymbol{\gamma}}_Y,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1}$, and $\widehat{\boldsymbol{\gamma}}_S$ denote the estimates of $\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1}$, and $\boldsymbol{\gamma}_S$, the parameters indexing models for $\mathbb{E}\left(Y|A,S=1,\mathbf{C}_1\right)$, $F(\mathbf{C}_1|A=0)$, and $\Pr(S=1|A,\mathbf{C}_1)$, respectively. Suppose that such estimates are obtained by solving a set of score equations with respective scores $\mathbf{M}_Y\left(\boldsymbol{\gamma}_Y\right),\mathbf{M}_{\mathbf{C}_1}\left(\boldsymbol{\gamma}_{\mathbf{C}_1}\right)$, and $\mathbf{M}_S\left(\boldsymbol{\gamma}_S\right)$. Let $\mathbf{M}\left(\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1},\boldsymbol{\gamma}_S\right)=\left(\mathbf{M}_Y^T\left(\boldsymbol{\gamma}_Y\right),\mathbf{M}_{\mathbf{C}_1}^T\left(\boldsymbol{\gamma}_{\mathbf{C}_1}\right),\mathbf{M}_S^T\left(\boldsymbol{\gamma}_S\right)\right)^T$ and define $U(\mu_0,\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1},\boldsymbol{\gamma}_S)$ to equal $U(\mu_0)$ under the parametric model, such that $\widehat{U}(\mu_0)=U(\mu_0,\widehat{\boldsymbol{\gamma}}_Y,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1},\widehat{\boldsymbol{\gamma}}_S)$. Then, a standard Taylor series expansion can be used to show that the large sample variance of $\widehat{\mu}_0^{dr}$ is consistently estimated by $\widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{\Gamma}}^{-1}$, where

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}^{-1} &= \mathbb{P}_n\left(\frac{\partial U(\mu_0,\widehat{\boldsymbol{\gamma}}_Y,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1},\widehat{\boldsymbol{\gamma}}_S)}{\partial\mu_0}|_{\widehat{\mu}_0^{dr}}\right)\\ \widehat{\boldsymbol{\Omega}} &= \mathbb{P}_n\left(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T\right)\\ \widehat{\mathbf{L}} &= U\left(\widehat{\mu}_0^{dr},\widehat{\boldsymbol{\gamma}}_Y,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1},\widehat{\boldsymbol{\gamma}}_S\right)-\mathbb{P}_n\left(\frac{\partial U\left(\widehat{\mu}_0^{dr},\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1},\boldsymbol{\gamma}_S\right)}{\partial\left(\boldsymbol{\gamma}_Y^T,\boldsymbol{\gamma}_{\mathbf{C}_1}^T,\boldsymbol{\gamma}_S^T\right)^T}|_{\left(\widehat{\boldsymbol{\gamma}}_Y^T,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1}^T,\widehat{\boldsymbol{\gamma}}_S^T\right)^T}\right)\\ &\quad\times\mathbb{P}_n\left(\frac{\partial\mathbf{M}\left(\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1},\boldsymbol{\gamma}_S\right)}{\partial\left(\boldsymbol{\gamma}_Y,\boldsymbol{\gamma}_{\mathbf{C}_1},\boldsymbol{\gamma}_S\right)^t}|_{\left(\widehat{\boldsymbol{\gamma}}_Y^T,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1}^T,\widehat{\boldsymbol{\gamma}}_S^T\right)^T}\right)^{-1}\mathbf{M}\left(\widehat{\boldsymbol{\gamma}}_Y,\widehat{\boldsymbol{\gamma}}_{\mathbf{C}_1},\widehat{\boldsymbol{\gamma}}_S\right)\end{aligned}$$

5.2. *Sensitivity analysis*

A key assumption we have made in order to identify SACE is that there is no unmeasured common cause of S and Y, which is clearly encoded in the causal diagram of Figure 1 and its associated NPSEM (1)–(4). In the current section, a sensitivity analysis technique is developed to assess the extent to which a violation of the assumption might affect results. Unlike previous sensitivity analysis techniques for truncation by death and related contexts [7, 8, 12, 13], the proposed sensitivity analysis technique makes explicit use of post-exposure covariates and therefore extends previous methods to the current more general longitudinal context. We begin by describing the approach in the simple three-occasion study. Define the selection bias function as follows:

$$t(\mathbf{c}_1) = \mathbb{E}\left\{Y(a=0)|A=0, \mathbf{C}_1(a=0)=\mathbf{c}_1, S(a=0)=S\left(a=1, \mathbf{C}_1(a=0)=\mathbf{c}_1\right)=1\right\} - \mathbb{E}\left\{Y(a=0)|A=0, \mathbf{C}_1(a=0)=\mathbf{c}_1, S(a=0)=1, S\left(a=1, \mathbf{C}_1(a=0)=\mathbf{c}_1\right)=0\right\}$$

We have that $t(\cdot)=0$ under the independence assumptions encoded in the NPSEM (1)–(4); however, if there was an unmeasured common cause of S and Y, such that ε_Y and ε_S were no longer independent, then we would expect that $t(\mathbf{c}_1) \neq 0$ for some value of $\mathbf{c}_1$, even if all other independencies encoded in the NPSEM continued to hold. In the smoking example, if the selection bias function is not null, we would expect that persons with worse health status and therefore worse cognitive decline would be more likely not to survive if they smoked, and therefore, we expect $t(\mathbf{C}_1) \leqslant 0$. We propose to recover inferences about SACE by assuming that the selection bias function $t(\mathbf{c}_1)$ is known, which encodes the magnitude and direction of the unmeasured common cause of S and Y. Specifically, the sensitivity analysis considers relaxation of the assumption that $\mathbf{C}_1$ includes all common causes of Y and S. To motivate the proposed approach, suppose for the moment for the sake of exposition that $\pi\left(a, \mathbf{C}_1\right) = \Pr(S|a, \mathbf{C}_1)$ were known, then we show in the Appendix (as a special case of the more general expression (17) given in the following text) that

$$\begin{aligned}
&\mathbb{E}\left\{Y(a=0)|A=0, \mathbf{C}_1(a=0)=\mathbf{c}_1, S(a=0)=S(a=1, \mathbf{C}_1(a=0)=\mathbf{c}_1)=1\right\} \\
&= \mathbb{E}\left\{Y(a=0)|A=0, \mathbf{C}_1(a=0)=\mathbf{c}_1, S(a=0)=1\right\} + t(\mathbf{c}_1) \times \left\{1 - \frac{\pi\left(1, \mathbf{c}_1\right)}{\pi\left(0, \mathbf{c}_1\right)}\right\} \\
&= \mathbb{E}\left\{Y|A=0, \mathbf{C}_1=\mathbf{c}_1, S=1\right\} + t(\mathbf{c}_1) \times \left\{1 - \frac{\pi\left(1, \mathbf{c}_1\right)}{\pi\left(0, \mathbf{c}_1\right)}\right\}
\end{aligned} \tag{16}$$

therefore, knowing $t(\mathbf{c}_1)$ allows one to recover the average potential outcome when unexposed in the principal strata of survivors $\{S(a=0)=S(a=1, \mathbf{C}_1(a=0)=\mathbf{c}_1)=1\}$, by adjusting the average observed outcome in the unexposed who survived, using the aforementioned expression. The expression in the above display can then be combined with the representation of $\mu_0 = \mathbb{E}\left\{Y(a=0)|\, S(a, \mathbf{C}_1(a'))=1; a, a'=0, 1\right\}$ given by Equation (10), to produce the following modified estimator:

$$\widehat{\mu}_0(t) = \frac{\mathbb{P}_n\left\{(1-A)\frac{\widehat{\pi}(1,\mathbf{C}_1)}{\widehat{\pi}(0,\mathbf{C}_1)}\left[Y + t(\mathbf{C}_1) \times \left\{1 - \frac{\widehat{\pi}(1,\mathbf{C}_1)}{\widehat{\pi}(0,\mathbf{C}_1)}\right\}\right] S\right\}}{\mathbb{P}_n\left\{(1-A)S\frac{\widehat{\pi}(1,\mathbf{C}_1)}{\widehat{\pi}(0,\mathbf{C}_1)}\right\}}$$

A formal sensitivity analysis can be obtained by repeating this process and reporting inferences about $\mu_1 - \mu_0$ using $\widehat{\mu}_1 - \widehat{\mu}_0(t)$ and a corresponding CI for each choice of $t(\cdot)$, say in a finite set of user-specified functions $\mathcal{T} = \{t_\lambda(\cdot) : \lambda\}$ indexed by a finite dimensional parameter λ with $t_0(\cdot) \in \mathcal{T}$ corresponding to the assumption of no unmeasured common cause of Y and S, that is, $t_0(\cdot) \equiv 0$.

As is shown next, this sensitivity analysis technique readily extends to a longitudinal study with $J > 3$ occasions. Define the selection bias function:

$$t_J\left(\overline{\mathbf{c}}_{J-1}\right) = \mathbb{E}\left\{Y(a=0)|A=0, \overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}, S_J(a=0)=S_J\left(a=1, \overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\} - \mathbb{E}\left\{Y(a=0)|A=0, \overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}, S_J(a=0)=1, S_J\left(a=1, \overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0\right\}$$

Similar to the three-occasion setting, $t_J(\cdot) = 0$ under the independence assumptions encoded in the NPSEM (11)–(14); however, if there were an unmeasured common cause of the survival process and Y among individuals alive at the end of follow-up, such that ε_Y and $\{\varepsilon_{S_j} : j\}$ were no longer independent, then we would expect that $t_J(\overline{\mathbf{c}}_{J-1}) \neq 0$ for some value of $\overline{\mathbf{c}}_{J-1}$, even if all other independencies of the NPSEM were to continue to hold. Then, similar to the three-occasion derivation, one obtains the following relation, which is derived in the Appendix:

$$\begin{aligned}
&\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1})=1\right\}\\
&=\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right\}+t_J(\overline{\mathbf{c}}_{J-1})\times\left\{1-\frac{\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)}{\prod_{j=2}^{J}\pi_j\left(0,\overline{\mathbf{c}}_{j-1}\right)}\right\}\\
&=\mathbb{E}\left\{Y|A=0,\overline{\mathbf{C}}_{J-1}=\overline{\mathbf{c}}_{J-1},S_J=1\right\}+t_J(\overline{\mathbf{c}}_{J-1})\times\left\{1-\frac{\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)}{\prod_{j=2}^{J}\pi_j\left(0,\overline{\mathbf{c}}_{j-1}\right)}\right\}
\end{aligned} \tag{17}$$

This result combines with the representation of $\mu_{0,J}$ from Section 3, to produce

$$\mu_{0,J}(t)=\frac{\mathbb{E}\left\{(1-A)W_SW_A\left[Y+t_J\left(\overline{\mathbf{c}}_{J-1}\right)\times\left(1-W_S\right)\right]S_J\right\}}{\mathbb{E}\left\{(1-A)S_JW_SW_A\right\}}$$

where as we recall

$$\begin{aligned}
W_S&=\frac{\prod_{k=2}^{J}\pi_k\left(1,\overline{\mathbf{C}}_{k-1}\right)}{\prod_{k=2}^{J}\pi_k\left(A,\overline{\mathbf{C}}_{k-1}\right)}\\
W_A&=\left[p\left(\mathbf{C}_0\right)^A\left\{1-p\left(\mathbf{C}_0\right)\right\}^{1-A}\right]^{-1}
\end{aligned}$$

which in turn can be used to obtain a consistent estimator. A sensitivity analysis then simply proceeds similar to the three-occasion setting described previously.

6. Survivor marginal structural models

It is now well known that in the context of a time-varying exposure with time-varying confounding, standard confounding adjustment techniques, such as stratification or standard regression analysis in general cannot appropriately account for time-varying confounding and therefore can be biased for the joint causal effects of the exposure, even under the causal null hypothesis of no exposure effect over time. In fact, the standard use of regression models to estimate the causal effect of a time-varying exposure can be biased even in the absence of unmeasured confounders whether or not one adjusts further for the past history of measured covariates in the analysis, when (i) there exists a time-dependent risk factor for the outcome, which also predicts subsequent exposure; and (ii) past exposure history predicts subsequent risk factor level. The reason is, when both conditions (i) and (ii) hold, an analysis that does not adjust for past covariates is biased because of the uncontrolled confounding, yet an analysis that includes current covariates is also biased as it adjusts for a variable affected by past exposure.

The MSMs were introduced by Robins [29] to estimate the joint causal effect of a time-dependent exposure in the presence of time-dependent confounders that are themselves intermediate variables, affected by previous exposure. MSMs were proposed as an alternative approach to the semiparametric g-computation algorithm estimator [1] and to g-estimation of structural nested models [29]. Robins [29] and subsequently Hernan *et al.* [30] described inverse probability-of-treatment-weighted estimation of MSMs, a method which in contrast to standard methods provides consistent estimates of causal effects when unmeasured confounding, model misspecification, and truncation by death are absent. We extend the results of previous sections to the context of MSMs.

First, we redefine the NPSEM to allow for time-updated exposure, and we assume no other form of loss to follow-up is present. Let $A_0 = 0$, such that individuals are assumed to be unexposed at start of follow-up, and let $\mathbf{C}_{-1} = 0$, then for $j = 0, \ldots, J$:

$$\mathbf{C}_j = \begin{cases} \mathbf{g}_{\mathbf{C}_j}\left(\overline{A}_j, \overline{\mathbf{C}}_{j-1}, \boldsymbol{\varepsilon}_{\mathbf{C}_j}\right) & \text{if } S_j = 1 \\ \text{undefined} & \text{if } S_j = 0 \end{cases} \tag{18}$$

$$S_{j+1} = \begin{cases} g_{S_{j+1}}\left(\overline{A}_j, \overline{\mathbf{C}}_j, \varepsilon_{S_{j+1}}\right) & \text{if } S_j = 1 \\ 0 & \text{if } S_j = 0 \end{cases} \tag{19}$$

$$A_{j+1} = \begin{cases} g_{A_{j+1}}\left(\overline{A}_j, \overline{\mathbf{C}}_j, \varepsilon_{A_{j+1}}\right) & \text{if } S_{j+1} = 1 \\ \text{undefined} & \text{if } S_{j+1} = 0 \end{cases} \tag{20}$$

Let $Y = \mathbf{C}_J$ denote the outcome. We assume that

$$\left\{\varepsilon_{A_j} : j \geqslant 1\right\} \perp\!\!\!\perp \left\{\varepsilon_{s_j} : j \geqslant 1\right\}$$

and we also assume that

$$\left\{\varepsilon_{A_j}, \varepsilon_{s_j} : j \geqslant 1\right\} \perp\!\!\!\perp \left\{\boldsymbol{\varepsilon}_{\mathbf{C}_j} : j \geqslant 0\right\}$$

However, as before, $\boldsymbol{\varepsilon}_{\mathbf{C}_j}$ and $\boldsymbol{\varepsilon}_{\mathbf{C}_{j'}}$ may be dependent, and likewise, for ε_{s_j} and $\varepsilon_{s_{j'}}$, $j \neq j'$. The NPSEM (18)–(20) with the associated error independencies essentially states that $\overline{\mathbf{C}}_{j-1}$ is sufficiently rich such that $\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{j-1}\right)$ account for any association between S_j and $\mathbf{C}_j$, and likewise, $\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{j-1}\right)$ accounts for confounding of the effects of A_j.

To account for truncation by death, consider the average potential outcome of survivors:

$$\mu_J(\overline{a}) = \mathbb{E}\left\{Y\left(\overline{a}\right) | S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^*\right)\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^*\right)\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) = 1;\ \text{for all } \overline{a}^* \in \{0,1\}^J\right\}$$

The conditioning event of the above expectation would be satisfied if an individual would survive irrespective of exposure history, including under certain cross-world situations where the covariate history behaves as if under an exposure history that possibly conflicts with that influencing the outcome. Thus, SMSMs give a natural generalization of standard MSMs to account for truncation by death. In the special case where $S_J(\overline{a}^*, \overline{\mathbf{C}}_{J-1}(\overline{a}^*))S_J(\overline{0}, \overline{\mathbf{C}}_{J-1}(\overline{0})) = 1 \Rightarrow S_J(\overline{0}, \overline{\mathbf{C}}_{J-1}(\overline{a}^*))S_J(\overline{a}^*, \overline{\mathbf{C}}_{J-1}(\overline{0})) = 1$ almost surely, then $\mu_J\left(\overline{a}\right)$ simplifies and may be written

$$\mathbb{E}\left\{Y\left(\overline{a}\right) | S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^*\right)\right) = S_J\left(\overline{a}^*\right) = 1;\ \overline{a}^* \in \{0,1\}^J\right\}$$

which extends the standard definition of SACE to time-updated exposure settings.

Identification of $\mu_J\left(\overline{a}\right)$ requires a modification of the monotonicity and concordant survivorship assumptions:

Sequential monotonicity assumption for time-dependent exposure: *For any treatment history* $\overline{a}_{j-1} \in \{0,1\}^{J-1}$, *if*

$$S_{j-1}\left(\overline{a}_{j-2}, \overline{\mathbf{C}}_{j-2}\left(\overline{a}_{j-3}\right)\right) S_{j-1}\left(\overline{0}_{j-2}, \overline{\mathbf{C}}_{j-2}\left(\overline{a}_{j-3}\right)\right) = 1 \textit{ almost surely}$$

then

$$S_j\left(\overline{0}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right)\right) \leqslant S_j\left(\overline{a}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right)\right) \textit{ almost surely}$$

where $S_0(\cdot,\cdot) = 1$ *almost surely; and if*

$$S_{j-1}\left(\overline{a}_{j-2}, \overline{\mathbf{C}}_{j-2}\left(\overline{0}_{j-3}\right)\right) S_{j-1}\left(\overline{0}_{j-2}, \overline{\mathbf{C}}_{j-2}\left(\overline{0}_{j-3}\right)\right) = 1 \text{ almost surely}$$

then

$$S_j\left(\overline{0}_{j-1}, \overline{\mathbf{C}}_{j-1}(0)\right) \leqslant S_j\left(\overline{a}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{0}_{j-2}\right)\right) \text{ almost surely.}$$

Concordant survivorship assumption for time-dependent exposure:

$$\text{if either } S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^*\right)\right) S_J\left(\overline{0}_{J-1}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^*\right)\right) = 1 \text{ or } S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) = 1$$

almost surely, for some exposure history $\overline{a}^*$, *then*

$$S_J\left(\overline{a}^{**}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^{**}\right)\right) S_J\left(\overline{0}_{J-1}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}^{**}\right)\right) = 1$$
$$\text{and } S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) S_J\left(\overline{a}^{**}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) = 1 \text{ for all } \overline{a}^{**}, \text{ almost surely.}$$

We should note that the sequential monotonicity assumption essentially states that receiving a dose of exposure can never be harmful for survival relative to remaining unexposed over time, and therefore in contrast with previous sections where exposure was smoking history and therefore was assumed to have a harmful effect on survival, exposure is now assumed to have a protective survival effect, for example, highly active antiretroviral therapy taken by HIV patients [23]. Specifically, the condition states that a person who would survive up to time $j-1$ if either untreated or with treatment history $\overline{a}$, in the hypothetical situation in which his covariate history behaves as if his treatment history was set to $\overline{a}$, and the person were to survive at time j when untreated and with covariate history under treatment regime $\overline{a}$, then the person would also survive at time j, with similar covariate history, if his treatment history was set to $\overline{a}$.

The concordant survivorship assumption essentially states that a person who would survive under a given treatment history $\overline{a}^*$, as well as if he were never exposed, in the hypothetical situation in which his covariate history behaves as if his exposure history was set to $\overline{a}^*$, then he would also survive under any other treatment history $\overline{a}^{**}$, and he would likewise survive if he were never exposed, in the hypothetical situation in which his covariate history behaves as if his exposure history was set to $\overline{a}^{**}$. A similar assumption is made for an individual for whom $S_J(\overline{0}, \overline{\mathbf{C}}_{J-1}(\overline{0}^*))S_J(\overline{a}^*, \overline{\mathbf{C}}_{J-1}(\overline{0})) = 1$ for some $\overline{a}^*$. Let

$$\pi_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = \Pr\left(S_j = 1 | \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{\mathbf{C}}_{j-1} = \overline{\mathbf{c}}_{j-1}, S_{j-1} = 1\right), j = 2, \ldots J$$
$$G_j\left(\mathbf{c}_j; \overline{a}_j, \overline{\mathbf{c}}_{j-1}\right) = F\left(\mathbf{c}_j | \overline{A}_j = \overline{a}_j, \overline{\mathbf{C}}_{j-1} = \overline{\mathbf{c}}_{j-1}, S_j = 1\right), \; j = 0, \ldots, J-1$$

Throughout, we make the following positivity assumptions

$$\frac{\prod_{j=2}^{J} \pi_j\left(\overline{0}_{j-1}, \overline{\mathbf{C}}_{j-1}\right) G_{j-1}\left(\mathbf{C}_{j-1}; \overline{0}_{j-1}, \overline{\mathbf{C}}_{j-2}\right)}{\prod_{j=1}^{J} \pi_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{j-1}\right) G_j\left(\mathbf{C}_{j-1}; \overline{A}_{j-1}, \overline{\mathbf{C}}_{j-2}\right)} < \infty \text{ almost surely}$$

and if

$$f\left(\overline{A}_j = \overline{a}_j, \overline{\mathbf{C}}_j = \overline{\mathbf{c}}_j, S_{j+1} = 1\right) > 0 \text{ then } f\left(A_{j+1} = a_{j+1} | \overline{A}_j = \overline{a}_j, \overline{\mathbf{C}}_j = \overline{\mathbf{c}}_j, S_{j+1} = 1\right) > 0 \text{ almost surely}$$

Then, we have the following result.

Theorem 5
Under the NPSEM given by Equations (18)–(20), the sequential monotonicity assumption for time-dependent exposure, the concordant survivorship assumption for time-dependent exposure, and the positivity assumptions, we have that $\mu_J\left(\overline{a}\right)$ is nonparametrically identified and is given by

$$\mu_J\left(\overline{a}\right) = \frac{\int \ldots \int \mathbb{E}\left(Y|\overline{A}=\overline{a}, S_J=1, \overline{\mathbf{C}}_{J-1}=\overline{\mathbf{c}}_{J-1}\right) \prod_{j=2}^{J} \pi_j\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=0}^{J-1} dG_j\left(\mathbf{c}_j; \overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right)}{\int \ldots \int \prod_{j=2}^{J} \pi_j\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) \prod_{j=0}^{J-1} dG_j\left(\mathbf{c}_j; \overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right)}$$

In practice, to estimate $\mu_J\left(\overline{a}\right)$, one may proceed as in Hernan *et al.* [23] and specify a model $\mu_J\left(\overline{a};\lambda\right)$ indexed by an unknown parameter λ. We refer to such a model as a SMSM, because it is an MSM for individuals that would survive under any treatment history, including under certain cross-world situations where the covariate history behaves as if under an exposure history, which conflicts with that influencing the outcome. For instance, consider the simple linear SMSM:

$$\mu_J\left(\overline{a};\lambda\right) = (1, cum(\overline{a}))\lambda$$

where $cum(\overline{a})) = \sum_{j=1}^{J} a_j$. Then, similar to the weighted least-squares approach developed in Section 3.2, one can likewise show that a consistent weighted least-squares estimator of the SMSM can be obtained by using the following modified weight $\widehat{W}_S^{\#} \times \widehat{W}_A^{\#}$, which accounts for time-dependent exposure and confounding and death-related attrition:

$$\widehat{W}_S^{\#} = \frac{\prod_{j=2}^{J} \widehat{\pi}_j\left(\overline{0}_{j-1}, \overline{\mathbf{C}}_{j-1}\right)}{\prod_{j=1}^{J} \widehat{\pi}_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{j-1}\right)}$$

$$\widehat{W}_A^{\#} = \prod_{j=1}^{J}\left[\widehat{p}_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{J-1}\right)^{A_j}\left\{1-\widehat{p}_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{J-1}\right)\right\}^{1-A_j}\right]^{-1}$$

where $\widehat{\pi}_j(\cdot,\cdot)$ is a consistent estimator of π_j and $\widehat{p}_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{J-1}\right)$ is a consistent estimator of $p_j\left(\overline{A}_{j-1}, \overline{\mathbf{C}}_{J-1}\right) = \Pr(A_j|\overline{A}_{j-1}, \overline{\mathbf{C}}_{J-1})$. For inference, it is possible to derive a sandwich variance estimator similar to the one provided in previous sections; alternatively, one may use the nonparametric bootstrap.

7. Discussion

In this paper, we have developed a general framework for identification and estimation of causal effects when the outcome in view is subject to truncation by death. The proposed approach is shown to equally apply in the context of a point exposure but also if joint effects of a time-updated exposure are in view. A simple weighted approach is described for estimation, which readily scales with follow-up of increasing length and applies irrespective of whether the exposure is time-updated or occurs at a single point in time. Doubly robust estimation is shown to be possible a simple three-occasion study. However, it is unclear that a similar doubly robust estimator is available beyond this simple setup, say if one has more than three follow-up visits in a longitudinal study. Although such generalizations are of definite interest and deserve further investigation. A sensitivity analysis technique is developed for a general longitudinal study of arbitrary length, which may be used to evaluate the extent to which a violation of the assumption that one has observed sufficient post-exposure covariates to account for an association between the outcome and survival, may bias the results. In the future, we plan to further develop the sensitivity analysis approach and to implement these techniques to the longitudinal smoking data application illustrated in the paper.

Appendix

Proof of Theorem 1
We can write

$$\begin{aligned}\psi &= \mathbb{E}\left\{Y(a=1)-Y(a=0)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}\\ &= \mathbb{E}\left\{Y(a=1)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}\\ &\quad -\mathbb{E}\left\{\ Y(a=0)|S(a=0,\mathbf{C}_1(a^*))=S(a=1,\mathbf{C}_1(a^*))=1,\ a^*=0,1\right\}\end{aligned}$$

Then, note that by the concordant survivorship assumption

$$\begin{aligned}&\mathbb{E}\left\{Y(a=1)|S\left(a=0,\mathbf{C}_1(a^*)\right)=S\left(a=1,\mathbf{C}_1(a^*)\right)=1,\ a^*=0,1\right\}\\&=\mathbb{E}\left\{Y(a=1)|S\left(a=0,\mathbf{C}_1(1)\right)=S\left(a=1,\mathbf{C}_1(1)\right)=1\right\}\\&=\frac{\mathbb{E}\left\{Y(a=1)S\left(a=0,\mathbf{C}_1(1)\right)S\left(a=1,\mathbf{C}_1(1)\right)\right\}}{\mathbb{E}\left\{S\left(a=0,\mathbf{C}_1(1)\right)S\left(a=1,\mathbf{C}_1(1)\right)\right\}}\end{aligned}$$

and by the monotonicity assumption

$$\begin{aligned}&\mathbb{E}\left\{Y(a=1)S(a=0,\mathbf{C}_1(1))S(a=1,\mathbf{C}_1(1))\right\}\\&=\mathbb{E}\left\{Y(a=1)S\left(a=1,\mathbf{C}_1(1)\right)\right\}\\&=\int\mathbb{E}\left\{Y(a=1)|S\left(a=1,\mathbf{C}_1(1)=\mathbf{c}_1\right)=1,\mathbf{C}_1(1)=\mathbf{c}_1)\right\}\\&\quad\times\Pr\left(S\left(a=1,\mathbf{C}_1(1)=\mathbf{c}_1\right)=1|\mathbf{C}_1(1)=\mathbf{c}_1\right)dF_{\mathbf{C}_1(1)}\left(\mathbf{c}_1\right)\\&=\int\mathbb{E}\left\{Y|A=1,S=1,\mathbf{C}_1=\mathbf{c}_1\right)\}\Pr\left(S=1|\mathbf{C}_1=\mathbf{c}_1,A=1\right)dF_{\mathbf{C}_1|A=1}\left(\mathbf{c}_1|A=1\right)\end{aligned}$$

by the independence assumptions associated with the NPSEM, and similarly,

$$\begin{aligned}&\mathbb{E}\left\{S\left(a=0,\mathbf{C}_1(1)\right)S\left(a=1,\mathbf{C}_1(1)\right)\right\}\\&=\int\Pr\left(S=1|\mathbf{C}_1=\mathbf{c}_1,A=1\right)dF_{\mathbf{C}_1|A=1}\left(\mathbf{c}_1|A=1\right)\end{aligned}$$

and therefore,

$$\begin{aligned}&\mathbb{E}\left\{Y(a=1)|S\left(a=0,\mathbf{C}_1(a^*)\right)=S\left(a=1,\mathbf{C}_1(a^*)\right)=1,\ a^*=0,1\right\}\\&=\mathbb{E}\left\{Y|A=1,S=1)\right\}\end{aligned}$$

Next consider

$$\begin{aligned}&\mathbb{E}\left\{Y(a=0)|S\left(a=0,\mathbf{C}_1(a^*)\right)=S\left(a=1,\mathbf{C}_1(a^*)\right)=1,\ a^*=0,1\right\}\\&=\mathbb{E}\left\{Y(a=0)|S\left(a=0,\mathbf{C}_1(0)\right)=S\left(a=1,\mathbf{C}_1(0)\right)=1\right\}\\&=\frac{\mathbb{E}\left\{Y(a=0)S\left(a=0,\mathbf{C}_1(0)\right)S\left(a=1,\mathbf{C}_1(0)\right)\right\}}{\mathbb{E}\left\{S\left(a=0,\mathbf{C}_1(0)\right)S\left(a=1,\mathbf{C}_1(0)\right)\right\}}\end{aligned}$$

Note that

$$\begin{aligned}&\mathbb{E}\left\{Y(a=0)S(a=0,\mathbf{C}_1(0))S(a=1,\mathbf{C}_1(0))\right\}\\&=\int\mathbb{E}\left\{Y(a=0)|S\left(a=0,\mathbf{C}_1(0)=\mathbf{c}_1\right)=S\left(a=1,\mathbf{C}_1(0)=\mathbf{c}_1\right)=1,\mathbf{C}_1(0)=\mathbf{c}_1\right)\right\}\\&\quad\times\Pr\left(S\left(a=1,\mathbf{C}_1(0)=\mathbf{c}_1\right)S\left(a=0,\mathbf{C}_1(0)=\mathbf{c}_1\right)=1|\mathbf{C}_1(0)=\mathbf{c}_1\right)dF_{\mathbf{C}_1(0)}\left(\mathbf{c}_1\right)\\&=\int\mathbb{E}\left\{Y(a=0)|S\left(a=0,\mathbf{c}_1\right)=1,\mathbf{C}_1(0)=\mathbf{c}_1\right)\right\}\\&\quad\times\Pr\left(S\left(a=1,\mathbf{c}_1\right)=1|A=1,\mathbf{C}_1(0)=c_1\right)dF_{\mathbf{C}_1|A}(\mathbf{c}_1|A=0)\text{ (monotonicity NPSEM independence)}\\&=\int\mathbb{E}\left\{Y(a=0)|S\left(a=0,\mathbf{c}_1\right)=1,\mathbf{C}_1(0)=\mathbf{c}_1\right)\right\}\\&\quad\times\Pr\left(S\left(a=1,\mathbf{c}_1\right)=1|A=1,\mathbf{C}_1(1)=\mathbf{c}_1\right)dF_{\mathbf{C}_1|A}\left(\mathbf{c}_1|A=0\right)\text{ (NPSEM independence)}\\&=\int\mathbb{E}\left\{Y|A=0,S=1,\mathbf{C}_1=\mathbf{c}_1\right)\}\text{ (NPSEM independence)}\\&\quad\times\Pr(S=1|A=1,\mathbf{C}_1=\mathbf{c}_1)dF_{\mathbf{C}_1|A}\left(\mathbf{c}_1|A=0\right).\end{aligned}$$

Likewise,

$$\mathbb{E}\left\{S(a=0,\mathbf{C}_1(0))S(a=1,\mathbf{C}_1(0))\right\}$$
$$=\int \Pr(S=1|A=1,\mathbf{C}_1=\mathbf{c}_1)dF_{\mathbf{C}_1|A}(\mathbf{c}_1|A=0), \text{ which proves the theorem}$$

□

Proof of Theorem 2
Similarly to the proof of Theorem 1, note that

$$\mu_{1,J}=\mathbb{E}\left\{Y(a=1)|S_J\left(a,\overline{\mathbf{C}}_{J-1}(a^*)\right)=1;\ a,a^*\in\{0,1\}\right\}$$
$$=\frac{\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}{\mathbb{E}\left\{S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}$$

by the concordant survivorship assumption, furthermore, by the sequential monotonicity assumption:

$$\frac{\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}{\mathbb{E}\left\{S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)S_J\left(0,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}$$
$$=\frac{\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}{\mathbb{E}\left\{S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}}$$

Then, note that

$$\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)\right\}$$
$$=\mathbb{E}\left[\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)|\overline{\mathbf{C}}_1\right\}\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left\{\mathbb{E}\left\{Y(a=1)S_J\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)|S_2\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)=1,\overline{\mathbf{C}}_1\right\}\right.\right.$$
$$\left.\left.\Pr\left(S_2\left(1,\overline{\mathbf{C}}_{J-1}(1)\right)=1|\overline{\mathbf{C}}_1\right)|\overline{\mathbf{C}}_1\right\}\right]$$
$$\vdots$$
$$=\int\cdots\int\mathbb{E}\left\{Y\left(a=1,\overline{\mathbf{c}}_{J-1}(1)\right)|S_J\left(1,\overline{\mathbf{c}}_{J-1}\right)=1,\overline{\mathbf{C}}_{J-1}(1)=\overline{\mathbf{c}}_{J-1}\right\}$$
$$\times\prod_{j=2}^{J}\Pr\left(S_j\left(1,\overline{\mathbf{c}}_{j-1}\right)=1,\overline{\mathbf{C}}_{j-1}(1)=\overline{\mathbf{c}}_{j-1},S_{j-1}\left(1,\overline{\mathbf{c}}_{j-2}\right)=1\right)$$
$$\times\prod_{j=0}^{J-1}dF_{\mathbf{C}_j\left(1,\overline{\mathbf{c}}_{j-1}\right)|\overline{\mathbf{C}}_{j-1}(1),S_j\left(1,\overline{\mathbf{C}}_{j-1}(1)\right)=1}\left(\mathbf{c}_j|S_j\left(1,\overline{\mathbf{c}}_{j-1}\right)=1,\overline{\mathbf{C}}_{j-1}(1)=\overline{\mathbf{c}}_{j-1}\right)$$
$$=\int\cdots\int\mathbb{E}\left\{Y\left(a=1,\overline{\mathbf{c}}_{J-1}(1)\right)|A=1,S_J\left(1,\overline{\mathbf{c}}_{J-1}\right)=1,\overline{\mathbf{C}}_{J-1}(1)=\overline{\mathbf{c}}_{J-1}\right\}$$
$$\prod_{j=2}^{J}\Pr\left(S_j\left(1,\overline{\mathbf{c}}_{j-1}\right)=1|A=1,\overline{\mathbf{C}}_{j-1}(1)=\overline{\mathbf{c}}_{j-1},S_{j-1}\left(1,\overline{\mathbf{c}}_{j-2}\right)=1\right)$$
$$\times\prod_{j=0}^{J-1}dF_{\mathbf{C}_j\left(1,\overline{\mathbf{c}}_{j-1}\right)|A,\overline{\mathbf{C}}_{j-1}(1),S_j\left(1,\overline{\mathbf{C}}_{j-1}(1)\right)=1}\left(\mathbf{c}_j|A=1,S_j\left(1,\overline{\mathbf{c}}_{j-1}\right)=1,\overline{\mathbf{C}}_{j-1}(1)=\overline{\mathbf{c}}_{j-1}\right)$$

$$= \int \ldots \int \mathbb{E}\left\{Y \,|A=1, S_J \;=1, \bar{\mathbf{c}}_{J-1}\right\} \prod_{j=2}^{J} \Pr\left(S_j = 1 | A=1, \bar{\mathbf{c}}_{j-1}, S_{j-1}=1\right)$$
$$\times \prod_{j=0}^{J-1} dF\left(\mathbf{c}_j | A=1, S_j=1, \bar{\mathbf{c}}_{j-1}\right) = \mathbb{E}\left\{Y\, S_J | A=1\right\}$$

A similar argument shows that

$$\mathbb{E}\left\{S_J\left(1, \overline{\mathbf{C}}_{J-1}(1)\right)\right\} = \mathbb{E}\left\{S_J | A=1\right\}$$

and therefore,

$$\frac{\mathbb{E}\left\{Y(a=1) S_J\left(1, \overline{\mathbf{C}}_{J-1}(1)\right)\right\}}{\mathbb{E}\left\{S_J\left(1, \overline{\mathbf{C}}_{J-1}(1)\right)\right\}} = \mathbb{E}\left\{Y \,|S_J, A=1\right\}$$

Next, note that

$$\mu_{0,J} = \mathbb{E}\left\{Y(a=0) | S_J\left(a, \overline{\mathbf{C}}_{J-1}(a^*)\right) = 1;\; a, a^* \in \{0,1\}\right\}$$
$$= \frac{\mathbb{E}\left\{Y(a=0) S_J\left(0, \overline{\mathbf{C}}_{J-1}(0)\right) S_J\left(1, \overline{\mathbf{C}}_{J-1}(0)\right\}\right.}{\mathbb{E}\left\{S_J\left(0, \overline{\mathbf{C}}_{J-1}(0)\right) S_J\left(1, \overline{\mathbf{C}}_{J-1}(0)\right\}\right.}$$

by the concordant survivorship assumption. We then have that

$$\mathbb{E}\left\{Y(a=0) S_J\left(0, \overline{\mathbf{C}}_{J-1}(0)\right) S_J\left(1, \overline{\mathbf{C}}_{J-1}(0)\right\}\right.$$
$$= \mathbb{E}\left[\mathbb{E}\left\{Y(a=0) S_J\left(0, \overline{\mathbf{C}}_{J-1}(0)\right) S_J\left(1, \overline{\mathbf{C}}_{J-1}(0) | \overline{\mathbf{C}}_1\right\}\right.\right]$$
$$\vdots$$
$$= \int \ldots \int \mathbb{E}\left\{Y\left(a=0, \bar{\mathbf{c}}_{j-1}\right) | S_J\left(1, \bar{\mathbf{c}}_{J-1}\right) = S_J\left(0, \bar{\mathbf{c}}_{J-1}\right) = 1, \overline{\mathbf{C}}_{J-1}(0) = \bar{\mathbf{c}}_{J-1}\right\}$$
$$\times \prod_{j=2}^{J} \Pr\left(S_j\left(1, \bar{\mathbf{c}}_{j-1}\right) = S_j\left(0, \bar{\mathbf{c}}_{j-1}\right) = 1 \overline{\mathbf{C}}_{j-1}(0) = \bar{\mathbf{c}}_{j-1}, S_{j-1}\left(1, \bar{\mathbf{c}}_{j-2}\right) = S_{j-1}\left(0, \bar{\mathbf{c}}_{j-2}\right) = 1\right)$$
$$\times \prod_{j=0}^{J-1} dF_{\mathbf{C}_j\left(0, \bar{\mathbf{c}}_{j-1}\right) | \overline{\mathbf{C}}_{j-1}(0), S_j\left(1, \overline{\mathbf{C}}_{j-1}(0)\right), S_j\left(0, \overline{\mathbf{C}}_{j-1}(0)\right)=1}\left(\mathbf{c}_j | S_j\left(1, \bar{\mathbf{c}}_{j-1}\right) = S_j\left(0, \bar{\mathbf{c}}_{j-1}\right) = 1, \overline{\mathbf{C}}_{j-1}(0) = \bar{\mathbf{c}}_{j-1}\right)$$
$$= \int \ldots \int \mathbb{E}\left\{Y\left(a=0, \bar{\mathbf{c}}_{j-1}\right) | S_J\left(0, \bar{\mathbf{c}}_{J-1}\right) = 1, \overline{\mathbf{C}}_{J-1}(0) = \bar{\mathbf{c}}_{J-1}\right\} \text{ (NPSEM independence)}$$
$$\times \prod_{j=2}^{J} \Pr\left(S_j\left(1, \bar{\mathbf{c}}_{j-1}\right) = 1 | \overline{\mathbf{C}}_{j-1}(0) = \bar{\mathbf{c}}_{j-1}, S_{j-1}\left(1, \bar{\mathbf{c}}_{j-2}\right) = 1\right) \text{ (sequential monotonicity)}$$
$$\times \prod_{j=0}^{J-1} dF_{\mathbf{C}_j\left(0, \bar{\mathbf{c}}_{j-1}\right) | \overline{\mathbf{C}}_{j-1}(0), S_j\left(0, \overline{\mathbf{C}}_{j-1}(0)\right)=1}\left(\mathbf{c}_j | S_j\left(0, \bar{\mathbf{c}}_{j-1}\right) = 1, \overline{\mathbf{C}}_{j-1}(0) = \bar{\mathbf{c}}_{j-1}\right) \text{ (NPSEM independence)}$$
$$= \int \ldots \int \mathbb{E}\left\{Y(a=0, \bar{c}_{j-1}) | A=0, S_J(0, \bar{\mathbf{c}}_{J-1}) = 1, \overline{\mathbf{C}}_{J-1}(0) = \bar{\mathbf{c}}_{J-1}\right\} \text{ (NPSEM independence)}$$
$$\times \prod_{j=2}^{J} \Pr\left(S_j\left(1, \bar{\mathbf{c}}_{j-1}\right) = 1 | A=1, \overline{\mathbf{C}}_{j-1}(1) = \bar{\mathbf{c}}_{j-1}, S_{j-1}\left(1, \bar{\mathbf{c}}_{j-2}\right) = 1\right) \text{ (NPSEM independence)}$$

$$\times \prod_{j=0}^{J-1} dF_{\mathbf{C}_j\left(0,\overline{\mathbf{c}}_{j-1}\right)|A,\overline{\mathbf{C}}_{j-1}(0),S_j\left(0,\overline{\mathbf{C}}_{j-1}(0)\right)=1}\left(\mathbf{c}_j|A=0,S_j\left(0,\overline{\mathbf{c}}_{j-1}\right)=1,\overline{\mathbf{C}}_{j-1}(0)=\overline{\mathbf{c}}_{j-1}\right)$$

(NPSEM independence)

$$=\int\ldots\int \mathbb{E}\left(Y|A=0,S_J=1,\overline{\mathbf{C}}_{J-1}=\overline{\mathbf{c}}_{J-1}\right)\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)\prod_{j=0}^{J-1}dG_{0,j}\left(\mathbf{c}_j;\overline{\mathbf{c}}_{j-1}\right)$$

One can show using similar arguments that

$$\mathbb{E}\left\{S_J\left(0,\overline{\mathbf{C}}_{J-1}(0)\right)S_J\left(1,\overline{\mathbf{C}}_{J-1}(0)\right\}\right.$$
$$=\int\ldots\int\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)\prod_{j=0}^{J-1}dG_{0,j}\left(\mathbf{c}_j;\overline{\mathbf{c}}_{j-1}\right)$$

proving the result. □

Proof of Theorem 3

$\widehat{\psi}_J$ converges in probability to the solution of the following population weighted normal equations:

$$\mathbb{E}\left\{S_J W_S W_A(1,A)^T\left(Y-\mu_{0,J}-\overline{\psi}_J A\right)\right\}=0$$

where $W_A=\left[p\left(\mathbf{C}_0\right)^A\left\{1-p\left(\mathbf{C}_0\right)\right\}^{1-A}\right]^{-1}$. It is straightforward to verify that under the assumptions of Theorem 3, the left hand-side of the equation in the aforementioned display is equal to

$$\sum_{a=0}^{1}\mathbb{E}_a^*\left\{S_J(1,a)^T\left(Y-\mu_{0,J}-\overline{\psi}_J a\right)\right\}=0$$

where $\mathbb{E}_a^*$ is the expectation with respect to the law

$$\begin{cases} f\left(Y|A=0,S_J=1,\overline{\mathbf{C}}_{J-1}\right)\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{C}}_{j-1}\right)\prod_{j=0}^{J-1}dG_{0,j}\left(\mathbf{C}_j;\overline{\mathbf{C}}_{j-1}\right) & \text{if } a=0 \\ f\left(Y,A=1,S_J=1\right) & \text{if } a=1 \end{cases}$$

giving the result. □

Proof of Theorem 4

Let $\pi_1\left(a,\mathbf{C}_1\right)=\Pi(a)=\Pr(S=1|A=0,\mathbf{C}_1), B_1=\mathbb{E}\left(Y|A=0,S=1,\mathbf{C}_1\right), G_a(\mathbf{C}_1)=F(\mathbf{C}_1|A=a)$. To prove the theorem, it suffice to show that $\mathbb{E}\left\{U(\mu_0,\pi_1^\dagger,B_1^\dagger)\right\}=0$ if either $\pi_1^\dagger=\pi_1$, or $B_1^\dagger=B_1$ and $G_a^\dagger(\mathbf{c}_1)=G_a(\mathbf{c}_1)$, but not necessarily both hold, where

$$U\left(\mu_0,\pi_1^\dagger,B_1^\dagger\right)=(1-A)S\frac{\Pi^\dagger(1)}{\Pi^\dagger(0)}\left\{Y-B_1^\dagger\right\}$$
$$+AB_1^\dagger\frac{dG_0^\dagger}{dG_1^\dagger}\left\{S-\Pi^\dagger(1)\right\}$$
$$+\left\{(1-A)B_1^\dagger\Pi^\dagger(1)-\mu_0\right\}$$

First assume that $\pi_1^\dagger = \pi_1$, then

$$\begin{aligned}
&\mathbb{E}\left\{U\left(\mu_0, \pi_1, B_1^\dagger\right)\right\} \\
&= \mathbb{E}\left[(1-A)S\frac{\Pi(1)}{\Pi(0)}\left\{Y - B_1^\dagger\right\} + AB_1^\dagger\frac{dG_0^\dagger}{dG_1^\dagger}\left\{S - \Pi(1)\right\} + \left\{(1-A)B_1^\dagger\Pi(1) - \mu_0\right\}\right] \\
&= \mathbb{E}\left[(1-A)S\frac{\Pi(1)}{\Pi(0)}\left\{Y - B_1^\dagger\right\} + \left\{(1-A)B_1^\dagger\Pi(1) - \mu_0\right\}\right] \\
&= \mathbb{E}\left[(1-A)S\frac{\Pi(1)}{\Pi 0(0)}Y - \mu_0\right] \\
&= 0
\end{aligned}$$

Next, suppose that $B_1^\dagger = B_1$ and $G_a^\dagger(\mathbf{c}_1) = G_a(\mathbf{c}_1)$, then

$$\begin{aligned}
&\mathbb{E}\left\{U\left(\mu_0, \pi_1, B_1^\dagger\right)\right\} \\
&= \mathbb{E}\left[(1-A)S\frac{\Pi^\dagger(1)}{\Pi^\dagger(0)}\left\{Y - B_1\right\} + AB_1\frac{dG_0}{dG_1}\left\{S - \Pi^\dagger(1)\right\} + \left\{(1-A)B_1\Pi^\dagger(1) - \mu_0\right\}\right] \\
&= \mathbb{E}\left[AB_1\frac{dG_0}{dG_1}S - \mu_0\right] \\
&= 0
\end{aligned}$$

proving the result. □

Proof of Theorem 5
The proof is similar to that of Theorem 2; suppose that $\overline{a} \neq \overline{0}$, then

$$\begin{aligned}
\mu_J\left(\overline{a}\right) &= \mathbb{E}\left\{Y\left(\overline{a}\right)|S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right) = 1\right\} \text{ (concordant survivorship)} \\
&= \frac{\mathbb{E}\left\{Y\left(\overline{a}\right)S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)\right\}}{\mathbb{E}\left\{S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)\right\}}
\end{aligned}$$

and we have

$$\begin{aligned}
&\mathbb{E}\left\{Y\left(\overline{a}\right)S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{Y\left(\overline{a}\right)S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right)|\mathbf{C}_0\right\}\right] \\
&\vdots \\
&= \int\cdots\int \mathbb{E}\left\{Y\left(\overline{a}, \overline{\mathbf{c}}_J\right)|S_J\left(\overline{a}, \overline{\mathbf{c}}_{J-1}\right) = S_J\left(\overline{0}, \overline{\mathbf{c}}_{J-1}\right) = 1, \overline{\mathbf{C}}_J\left(\overline{a}\right) = \overline{\mathbf{c}}_J\right\} \\
&\quad\times\prod_{j=2}^{J}\Pr\left(S_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = S_j\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = 1|\overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right) = \overline{\mathbf{c}}_{j-1}, S_{j-1}\left(\overline{a}_{j-2}, \overline{\mathbf{c}}_{j-2}\right)\right. \\
&\qquad\qquad\left. = S_{j-1}\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-2}\right) = 1\right) \\
&\quad\times\prod_{j=0}^{J} dF_{\mathbf{C}_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right)|\overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right), S_j\left(\overline{a}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right)\right)S_j\left(\overline{0}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right)\right)=1}\left(\mathbf{c}_j|S_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&= S_j\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = 1, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right) = \overline{\mathbf{c}}_{j-1}\Big) \\
&= \int \ldots \int \mathbb{E}\left\{ Y\left(\overline{a}, \overline{\mathbf{c}}_J\right) | S_J\left(\overline{a}, \overline{\mathbf{c}}_{J-1}\right) = 1, \overline{\mathbf{C}}_J\left(\overline{a}\right) = \overline{\mathbf{c}}_J \right\} \text{ (NPSEM independence)} \\
&\quad \times \prod_{j=2}^{J} \Pr\left(S_j\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = 1 | \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right) = \overline{\mathbf{c}}_{j-1}, S_{j-1}\left(\overline{0}_{j-1}, \overline{\mathbf{c}}_{j-2}\right) = 1 \right) \text{ (sequential monotonicity)} \\
&\quad \times \prod_{j=0}^{J} dF_{\mathbf{C}_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) | \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right), S_j\left(\overline{a}_{j-1}, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right)\right)=1} \left(\mathbf{c}_j | S_j\left(\overline{a}_{j-1}, \overline{\mathbf{c}}_{j-1}\right) = 1, \overline{\mathbf{C}}_{j-1}\left(\overline{a}_{j-2}\right) = \overline{\mathbf{c}}_{j-1} \right) \\
&\quad \text{(NPSEM independence)} \\
&= \int \ldots \int \mathbb{E}\left\{ Y | \overline{A}_J = \overline{a}_J, S_J = 1, \overline{\mathbf{C}}_J = \overline{\mathbf{c}}_J \right\} \text{ (NPSEM independence)} \\
&\quad \times \prod_{j=2}^{J} \Pr\left(S_j = 1 | \overline{A}_{j-1} = \overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}, S_{j-1} = 1 \right) \text{ (NPSEM independence)} \\
&\quad \times \prod_{j=0}^{J} dF_{\mathbf{C}_j | \overline{C}_{j-1}, \overline{A}_{j-1}, S_j=1} \left(\mathbf{c}_j | \overline{A}_{j-1} = \overline{a}_{j-1}, S_j = 1, \overline{\mathbf{c}}_{j-1} \right) \text{ (NPSEM independence)}
\end{aligned}
$$

giving the result for

$$\mathbb{E}\left\{ Y\left(\overline{a}\right) S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right) \right\}$$

The expression for

$$\mathbb{E}\left\{ S_J\left(\overline{a}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{a}\right)\right) \right\}$$

is similarly obtained. The result for $\mu_J\left(\overline{0}\right)$ is obtained by noting that

$$
\begin{aligned}
\mu_J\left(\overline{0}\right) &= \mathbb{E}\left\{ Y\left(\overline{0}\right) | S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) S_J\left(\overline{a}^*, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) = 1 \text{ for all } \overline{a}^* \right\} \text{ (concordant survivorship)} \\
&= \frac{\mathbb{E}\left\{ Y\left(\overline{0}\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) \right) = 1 \right\}}{\mathbb{E}\left\{ S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) \right) = 1 \right\}} \text{ (sequential monotonicity)}
\end{aligned}
$$

and it is straightforward to verify that

$$
\begin{aligned}
\mathbb{E}\left\{ Y\left(\overline{0}\right) S_J\left(\overline{0}, \overline{\mathbf{C}}_{J-1}\left(\overline{0}\right)\right) \right) = 1 \right\} &= \int \ldots \int \mathbb{E}\left\{ Y | \overline{A}_J = \overline{0}, S_J = 1, \overline{\mathbf{C}}_J = \overline{\mathbf{c}}_J \right\} \\
&\quad \times \prod_{j=2}^{J} \Pr\left(S_j = 1 | \overline{A}_{j-1} = \overline{0}_{j-1}, \overline{\mathbf{c}}_{j-1}, S_{j-1} = 1 \right) \\
&\quad \times \prod_{j=0}^{J} dF_{\mathbf{C}_j | \overline{\mathbf{C}}_{j-1}, \overline{A}_{j-1}, S_j=1} \left(\mathbf{c}_j | \overline{A}_{j-1} = \overline{0}_{j-1}, S_j = 1, \overline{\mathbf{c}}_{j-1} \right)
\end{aligned}
$$

and the expression for $\mathbb{E}\left\{ S_J(\overline{0}, \overline{\mathbf{C}}_{J-1}(\overline{0})) \right) = 1 \right\}$ is similarly derived. □

Derivation of Equation (17)

$$
\begin{aligned}
&\mathbb{E}\left\{Y|A=0,\overline{\mathbf{C}}_{J-1}=\overline{\mathbf{c}}_{J-1},S_J=1\right\}\\
&=\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\}\\
&\quad\times\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&\quad+\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1,S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0\right\}\\
&\quad\times\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&=\left[\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1,S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0\right\}\right.\\
&\quad\left.-\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\}\right]\\
&\quad\times\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&\quad+\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\}\\
&=\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\}\\
&\quad-t_J\left(\overline{\mathbf{c}}_{J-1}\right)\times\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=0|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)
\end{aligned}
$$

Then, note that

$$
\begin{aligned}
&\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&=\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&=\frac{\Pr\left(S_J\left(a=1,\overline{\mathbf{c}}_{J-1}\right)S_J(a=0)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_{J-1}\left(a=1,\overline{\mathbf{c}}_{J-2}\right)S_{J-1}\left(a=0,\overline{\mathbf{c}}_{J-2}\right)=1\right)}{\Pr\left(S_J(a=0)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_{J-1}(a=0)=1\right)}\\
&\quad\times\Pr\left(S_{J-1}\left(a=1,\overline{\mathbf{C}}_{J-2}(a=0)=\overline{\mathbf{c}}_{J-2}\right)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_{J-1}(a=0)=1\right)\\
&=\frac{\Pr\left(S_J\left(a=1,\overline{\mathbf{c}}_{J-1}\right)=1|A=1,\overline{\mathbf{C}}_{J-1}(a=1)=\overline{\mathbf{c}}_{J-1},S_{J-1}\left(a=1,\overline{\mathbf{c}}_{J-2}\right)=1\right)}{\Pr\left(S_J(a=0)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_{J-1}(a=0)=1\right)}\\
&\qquad\qquad\text{(sequential monotonicity and NPSEM independence)}\\
&\quad\times\Pr\left(S_{J-1}\left(a=1,\overline{\mathbf{C}}_{J-2}(a=0)=\overline{\mathbf{c}}_{J-2}\right)=1|A=0,\overline{\mathbf{C}}_{J-2}(a=0)=\overline{\mathbf{c}}_{J-2},S_{J-1}(a=0)=1\right)\\
&\qquad\qquad\text{(NPSEM independence)}
\end{aligned}
$$

thus by iterating, one obtains

$$
\begin{aligned}
&\Pr\left(S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=1\right)\\
&=\prod_{j=2}^{J}\frac{\Pr\left(S_j\left(a=1,\overline{\mathbf{c}}_{j-1}\right)=1|A=1,\overline{\mathbf{C}}_{j-1}(a=1)=\overline{\mathbf{c}}_{j-1},S_{j-1}\left(a=1,\overline{\mathbf{c}}_{j-2}\right)=1\right)}{\Pr\left(S_j(a=0)=1|A=0,\overline{\mathbf{C}}_{j-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_{j-1}(a=0)=1\right)}\\
&=\frac{\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)}{\prod_{j=2}^{J}\pi_j\left(0,\overline{\mathbf{c}}_{j-1}\right)}
\end{aligned}
$$

which gives

$$\begin{aligned}
&\mathbb{E}\left\{Y|A=0,\overline{\mathbf{C}}_{J-1}=\overline{\mathbf{c}}_{J-1},S_J=1\right\}\\
&\quad=\mathbb{E}\left\{Y(a=0)|A=0,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1},S_J(a=0)=S_J\left(a=1,\overline{\mathbf{C}}_{J-1}(a=0)=\overline{\mathbf{c}}_{J-1}\right)=1\right\}\\
&\qquad-t_J\left(\overline{\mathbf{c}}_{J-1}\right)\times\left\{1-\frac{\prod_{j=2}^{J}\pi_j\left(1,\overline{\mathbf{c}}_{j-1}\right)}{\prod_{j=2}^{J}\pi_j\left(0,\overline{\mathbf{c}}_{j-1}\right)}\right\}
\end{aligned}$$

proving the result. Equation (16) is obtained by setting $J=2$.

Acknowledgements

This work was supported by Award Number U54GM088558 from the National Institute Of General Medical Sciences and Award Numbers R01AI104459-01A1 from the National Institute of Allergy and Infectious Diseases and R01ES020337 from the National Institute of Environmental Health Sciences. The author would like to thank Denis Evans, Jennifer Weuve, and George Dombrowski for providing access to data from the Chicago Health and Aging Project (CHAP). The author is particularly grateful to Todd L. Beck for his invaluable assistance with the data analysis of CHAP.

References

1. Robins JM. A new approach to causal inference in mortality studies with sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986; **7**:1393–1512.
2. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**(1):21–29.
3. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by 'death'. *Journal of Educational and Behavioral Statistics* 2003; **28**(4):353–368.
4. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with 'censoring due to death' (with discussion). *Statistical Science* 2006; **21**(3):299–321.
5. Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine* 2003; **22**(14):2281–2298.
6. Hudgens MG, Halloran ME. Causal vaccine effects on binary postinfection outcomes. *Journal of the American Statistical Association* 2006; **101**(473):51–64.
7. Shepherd BE, Gilbert PB, Jemiai Y, Rotnitzky A. Sensitivity analyses comparing outcomes only existing in a subset selected postrandomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics* 2006; **62**(2):332–342.
8. Jemiai Y, Rotnitzky A, Shepherd BE, Gilbert PB. Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs. *Journal of the Royal Statistical Society Series B-Statistical Methodology 1* 2007; **69**(5):879–901.
9. Pearl J. *Causality: Models, Reasoning, and Inference* 2nd ed. Cambridge University Press: New York, 2009.
10. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics* 2007; **39**:17–23.
11. Tchetgen Tchetgen EJ, Glymour M, Weuve J, Shpitser I. To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology* 2012a; **23**(4):644–646.
12. Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* 2003; **59**(3):531–540.
13. Shepherd BE, Gilbert PB, Lumley T. Sensitivity analyses comparing time-to-event outcomes existing only in a subset selected postrandomization. *Journal of the American Statistical Association* 2007; **102**(478):573–582.
14. Shepherd B, Gilbert P, Dupont C. Sensitivity analyses for comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity. *Biometrics* 2011; **67**:1100–1110.
15. Gilbert P, Jin Y. Semiparametric estimation of the average causal effect of treatment on an outcome measured after a post-randomization event, with missing outcome data. *Biostatistics* 2010; **11**:34–47.
16. Chiba Y, VanderWeele TJ. A simple methd for principal strata effects when the outcome has been truncated due to death. *American Journal of Epidemiology* 2011; **173**:745–751.
17. Imai K. Sharp bounds on the causal effects in randomized experiments with 'truncation-by-death'. *Statistics and Probability Letters* 2008; **78**:144–149.
18. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* 2005; **61**(1):305–310.
19. Zhang JL, Rubin DB, Mealli F. 'Likelihood-based analysis of causal effects via principal stratification: new approach to evaluating job training programs'. *Journal of the American Statistical Association* 2009; **104**:166–176.
20. Ding P, Geng Z, Yan W, Zhou X. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association* 2011; **106**(496):1578–1591.
21. Nolen T, Hudgens M. Randomization-based inference within principal strata. *Journal of the American Statistical Association* 2011; **106**(494):581–593.

22. Dai JY, Gilbert PB, Mâsse BR. Partially hidden Markov model for time-varying principal stratification in HIV prevention trials. *Journal of the American Statistical Association* 2012; **107**(497):52–65.
23. Tchetgen Tchetgen EJ. On causal mediation analysis with a survival outcome. *International Journal of Biostatistics* 2011; **7**(1):1–38.
24. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *The Annals of Statistics* 2012; **40**(3):1816–1845.
25. Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.
26. Robins JM. Marginal structural models. *Proceedings of the American Statistical Association*, Section on Bayesian Statistical Science, 1997, 1–10.
27. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, Evans DE, Mendes de Leon CF. Accounting for bias due to selective attrition in analyses of cognitive decline: the example of smoking and cognitive decline in older adults.(with discussion and rejoinder). *Epidemiology* 2012; **23**(1):119–128.
28. Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. A cautionary note on specification of the correlation structure in inverse-probability-weighted estimation for repeated measures. *Epidemiology* 2012b; **23**(4):644–646.
29. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, Halloran ME, Berry D (eds), IMA, vol. 116. Springer-Verlag: NY, 1999; 95–134.
30. Hernán M, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**(5):561–570.
31. Hernán M, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* 2001; **96**(454):440–448.

