Statistica Sinica **33** (2023), 2495-2517
doi:https://doi.org/10.5705/ss.202021.0167

DEFINING AND ESTIMATING PRINCIPAL STRATUM SPECIFIC NATURAL MEDIATION EFFECTS WITH SEMI-COMPETING RISKS DATA

Fei Gao, Fan Xia and K. C. G. Chan

Fred Hutchinson Cancer Center,
University of California at San Francisco and University of Washington

Abstract: In many medical studies, an ultimate failure event, such as death, is likely to be affected by the occurrence and timing of other intermediate clinical events. Both event times are subject to censoring by loss-to-follow-up, but the nonterminal event may be further censored by the occurrence of the primary outcome, but not vice versa. To study the effect of an intervention on both events, the intermediate event may be viewed as a mediator. However, the conventional definitions of direct and indirect effects do not apply, because of the semi-competing risks data structure. We define three principal strata based on whether the potential intermediate event occurs before the potential failure event. This allows us to properly define direct and indirect effects in one stratum, and define total effects for all strata. We discuss the identification conditions for the stratum-specific effects, and propose a semiparametric estimator based on a multivariate logistic stratum membership model and within-stratum proportional hazards models for the event times. By treating the unobserved stratum membership as a latent variable, we propose an expectation-maximization algorithm for the computation. We study the asymptotic properties of the estimators using modern empirical process theory and examine the performance of the estimators in numerical studies.

Key words and phrases: Illness-death model, missing data, principal stratification, proportional hazards model, survival analysis.

1. Introduction

Evaluating the causal effects of an intervention on a clinical outcome is a common theme in many medical studies. After an overall relationship between an intervention and an outcome is established, it is often of further interest to understand the biological or mechanistic pathways that contribute to the causal treatment effect. Causal mediation analysis is often used to disentangle the total treatment effect by decomposing it into an indirect effect, that is, the effect exerted by intermediate variables (mediators), and a direct effect, that is, the effect

Corresponding author: Fei Gao, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA. E-mail: fgao@fredhutch.org.

involving pathways independent of the hypothesized mediators. Intuitively, to evaluate a direct effect, the mediators need to be somehow fixed. A variety of mediation effects can be defined by different ways of fixing the mediators. Natural effects are most relevant when studying treatment effect mechanisms in public health, because they compare the outcome under which the mediators are set to the value that occurs naturally under different treatment assignments. A number of methods have been proposed for causal mediation analysis with survival outcomes, for a single mediator measured at study entry (Lange and Hansen (2011); VanderWeele (2011); Tchetgen Tchetgen (2011); Lange, Vansteelandt and Bekaert (2012)), and for longitudinal mediators (Lin et al. (2017); Zheng and van der Laan (2017); Didelez (2019); Vansteelandt et al. (2019); Aalen et al. (2020)).

In many biomedical studies, intermediate nonterminal landmark events are recorded in addition to the primary failure event, because they are important when evaluating the prognosis. Because of the ordering of the two events, the nonterminal event is subject to censoring by the occurrence of the terminal event, but not vice versa, such that semi-competing risks data are observed (Fine, Jiang and Chappell (2001)). In this paper, we consider a setting in which a nonterminal event may serve as a mediator for individuals to whom the event occurs before the terminal event. An example is a multi-center trial of allogeneic bone marrow transplants in patients with acute leukemia (Copelan et al. (1991); Klein and Moeschberger (2006)), where the primary interest is on the effect of different treatment regimens (methotrexate + cyclosporine vs methylprednisolone + cyclosporine) on the survival time. The event time of an intermediate endpoint, chronic graft-versus-host disease (GVHD), is a major side effect of the transplant that can be lethal. However, some patients died without experiencing GVHD, such that the GVHD event time is subject to censoring by the death time.

Causal mediation analysis with semi-competing risks data is particularly challenging. First, the mediator is only well defined for those who experience the nonterminal event before the occurrence of the primary event. Therefore, the conventional definitions of natural indirect and direct effects, based on replacing the counterfactual of the mediator under one treatment with that under the other, do not apply to the entire population. Moreover, the semi-competing risks data structure, that is, the primary event may censor the intermediate event but not vice versa, posts additional challenges for the identifiability of the parameters relating to natural indirect and direct effects.

There has been an increase in research on various causal inference problems with semi-competing risk data. However, most of these works differ from ours

because they only consider subgroup average treatment effects (Comment et al. (2019); Xu et al. (2022); Nevo and Gorfine (2022)), and not mediation effects. Upon finishing this paper, we became aware of the newly accepted paper by Huang (2021), who estimates natural mediation effects for semi-competing risk data using a counting process framework. However, the problem formulation, estimand, and assumptions all differ significantly from those in our work. For instance, we do not make sequential ignorability assumptions on surviving subpopulations at an arbitrary post-treatment time, as assumed in Huang (2021), because evolving subpopulations are, in general, healthier than the baseline study population before the treatment is assigned.

In this paper, we consider a novel principal stratification approach to define the causal mediation effects in the subgroup in which the intermediate event happens when given either treatment, that is, those susceptible to the intermediate event under both treatments. The notation and settings are given in Sections 2.1 and 2.2, respectively. We discuss the identification conditions needed to estimate the stratum-specific natural indirect and direct effects in Section 2.3, and, in Section 2.4, we propose a semiparametric estimator based on a multivariate logistic stratum membership model and within-stratum proportional hazards models for the event times. In Section 2.5, by treating the unobserved stratum membership as a latent variable, we propose an expectation-maximization (EM) algorithm for the computation of the nonparametric maximum likelihood estimator. We also study the asymptotic properties of the estimators using modern empirical process theory in Section 2.6, and examine the performance of the estimators in simulation studies in Section 3. An analysis of data from a clinical trial is given in Section 4. Section 5 concludes the paper. All proofs, detailed derivations, and additional numerical results are given in the online Supplementary Material. The computation code for the simulation studies is available at `https://github.com/feigao1/Med_Semi_Comp`.

2. Methods

2.1. Notation for observed data

Let A be a binary treatment, T be the time to a primary event of interest, and M be the time to an intermediate nonterminal event. The intermediate event time M may be censored by the occurrence of the primary event, but not vice versa, such that we observe semi-competing risks data. For example, A is a treatment that prolongs the survival time, T is the time to death, and M is the time to cancer progression. The occurrence of death may censor the cancer

progression onset, but not vice versa.

Let $\boldsymbol{X}$ be a collection of baseline covariates that may be associated with either or both events. Let C denote a censoring time for the primary event, for example, the end of the follow-up time. Then, we observe $Y \equiv \min(T, C)$ and $\Delta^T = I(T \leq C)$ for the primary event, and $Z \equiv \min(M, Y)$ and $\Delta^M = I(M \leq Y)$ for the intermediate event. The observations are versions of the counterfactual variables, which we define in the next section.

2.2. Counterfactuals and causal estimands

To define causal mediation effects of interest, we adopt the potential outcomes framework. Let the variables $M(a)$ and $T(a) \equiv T(a, M(a))$ denote the counterfactual nonterminal and terminal event times, respectively, when the treatment is set to a. These quantities are called single-world variables, because the intervention is set to a single realizable value. Cross-world variables, defined later, involve interventions set to different values. In conventional settings, $M(a)$ and $T(a)$ can be defined separately as the values of M and T that would be observed had the treatment been set to a. However, in a semi-competing risk setting, the value for $M(a)$ cannot vary arbitrarily. When the potential primary event happens before the potential intermediate event, the value of the mediator is not well defined (and is often set to ∞, by convention). In such a case, the potential primary event time shall not be dependent on an arbitrary m greater than the potential primary event time. Therefore, to be consistent with the semi-competing risks data structure, we have the order invariance that either $M(a) < \infty$ and $T(a, M(a)) \geq M(a)$, or $M(a) = \infty$ and therefore $T(a, M(a)) < M(a)$.

In conventional mediation analysis, a comparison between potential outcomes with the mediator set to different values would define an indirect effect. Furthermore, if the mediators are set to counterfactual values under different interventions, the indirect effect is called the natural indirect effect. On the other hand, if the mediators are both set to the same value, a comparison between potential outcomes with different interventions defines a direct effect. Like the natural indirect effect, if the mediators are set to the counterfactual value under the same intervention, the direct effect is called the natural direct effect. Both the natural indirect and natural direct effects involve the term $T(a^*, M(a))$, that is, the potential outcome for the terminal event time when the treatment is set to a^* and the nonterminal event time is set to $M(a)$, which is the counterfactual nonterminal event time when the treatment is set to a. This quantity is a cross-world variable. In order to respect the semi-competing risks data structure, we need to restrict our attention to scenarios in which, if the nonterminal event happens, it

happens before the terminal event, such that $T(a^*, M(a))$ is well defined. This is formally given as the following cross-world ordering invariance assumption.

Assumption 1. *For any* $a, a^* \in \{0,1\}$*, either (i)* $M(a) < \infty$ *and* $T(a^*, M(a)) \geq M(a)$*, or (ii)* $M(a) = \infty$ *and* $T(a^*, M(a)) < M(a)$*.*

Note that without Assumption 1, $T(a^*, M(a))$ may not be well defined, which would make a mediation analysis pointless.

Furthermore, the potential nonterminal event may or may not occur before the potential primary event time under different treatment assignments, which motivates us to examine the causal effects based on our proposed principal stratification approach, extended from Frangakis and Rubin (2002). Intuitively, we stratify the study population into latent classes identified by U, with three categories, based on whether they are susceptible to the nonterminal event under different treatment assignments:

1. $U = 1$: $M(0) \leq T(0)$ and $M(1) \leq T(1)$ (always susceptible).
2. $U = 2$: $M(0) \leq T(0)$ and $M(1) = \infty$ (prevented).
3. $U = 3$: $M(0) = \infty$ and $M(1) = \infty$ (always nonsusceptible).

Here, we do not have a fourth stratum, $M(0) = \infty$ and $M(1) \leq T(1)$, such that the treatment never converts a subject from nonsusceptible to susceptible. This restriction is along the same line as the "no defier" assumption commonly adopted in instrumental variables methods, suggesting that the treatment effect is "monotone" and that there is no reversed effect for the subjects (Angrist, Imbens and Rubin (1996)).

Remark 1. The defined strata (and associated stratum-specific effects) are substantially different from the survivors' principal stratum (and the survivor average causal effect (SACE)) commonly defined in the "truncation by death" literature (Zhang and Rubin (2003); Comment et al. (2019)). In particular, the survivors' principal stratum is defined as $\{T(0) \geq t, T(1) \geq t\}$, for some fixed time t, in Comment et al. (2019), whereas our definition does not depend on an arbitrary post-treatment time t.

Remark 2. Lin et al. (2017) explain the difficulties in defining natural mediation effects in a survival context with longitudinal mediators. They define interventional effects in a discrete-time setting, where the mediators and past survival status are subject to a hypothetical intervention. They mention principal stratification as an alternative framework to avoid such a hypothetical intervention,

but do not explore this further. We consider a different setting that shares some of the difficulties, but also with a unique data structure so that the principal strata can be defined.

Under suitable assumptions (to be made clear in Section 2.3), for $U = 1$, the joint distribution of $(T(a), M(a))$, for $a = 0, 1$, can be nonparametrically identified on the upper wedge of the positive quadrant, and by cross-world invariance, $T(1, M(0)) \geq M(0)$ is well defined in the same region. Therefore, we can define and estimate the stratum-specific natural indirect and direct effects, as follows.

$$\begin{aligned} NIE_1(t; \boldsymbol{x}) = &\Pr\{T(1, M(1)) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 1\} \\ &- \Pr\{T(1, M(0)) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 1\} \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} NDE_1(t; \boldsymbol{x}) = &\Pr\{T(1, M(0)) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 1\} \\ &- \Pr\{T(0, M(0)) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 1\}. \end{aligned} \tag{2.2}$$

In the stratum with $U = 2$, although the pair $(T(1, M(0)), M(0))$ is technically defined on the upper wedge, $(T(1), M(1))$ is not defined in that region, because $M(1) = \infty$. Hence, there is no common support when the mediator is being considered, and we do not seek to estimate the indirect and direct effects. However, the stratum-specific total effect can still be estimated:

$$TE_2(t; \boldsymbol{x}) = \Pr\{T(1) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 2\} - \Pr\{T(0) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 2\} \ .$$

In the stratum with $U = 3$, $M(0) = M(1) = \infty$ and $T(1, M(0)) = T(1, M(1))$, so there is no indirect effect. Here, the stratum-specific total effect is defined as

$$TE_3(t; \boldsymbol{x}) = \Pr\{T(1) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 3\} - \Pr\{T(0) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 3\}.$$

Remark 3. In principle, a mediator satisfies temporal precedence, that is, it occurs before the primary event. Therefore, the mediator is technically absent in $U = 3$, and an attempt to define mediation effects would be futile. In $U = 2$, the presence of the mediator before the primary event only happens in one treatment level with certainty. As a result, one cannot fix the mediator level at a different treatment level, and the mediation effects cannot be defined. Note that when $U = 2$, TE_2 can be interpreted as the treatment effect in survival among individuals whose mediating events are prevented by the treatment.

2.3. Identification

To identify the stratum-specific natural indirect and direct effects and stratum-specific total effects, we impose the following assumptions.

Assumption 2. *If $A = a$, then $M = M(a)$ and $T = T(a)$ with probability one.*

Assumption 3. *For $a, a^* \in \{0, 1\}$ and $u \in \{1, 2, 3\}$,*

$$\{T(a, M(a^*)), M(a^*)\} \perp A | \boldsymbol{X}, U = u \tag{2.3}$$

and

$$\begin{aligned} &\Pr(T(a, M(a^*))|M(a^*) = m, A = a, \boldsymbol{X}, U = u) \\ &= \Pr(T(a, M(a))|M(a) = m, A = a, \boldsymbol{X}, U = u). \end{aligned} \tag{2.4}$$

Assumption 2 is the standard consistency assumption for causal inference. Assumption 3 serves a similar purpose to that of the sequential ignorability assumption (Imai, Keele and Yamamoto (2010)), but it is weaker, so that the assumption holds within a stratum and requires only that $T(a, M(a^*))$ be well defined. Based on Assumptions 2 and 3, we are able to connect the stratum-specific natural indirect and direct effects and stratum-specific total effects with the distribution of the observed data, given stratum membership, as follows.

Theorem 1. *Under Assumptions 2 and 3, for the stratum with $U = 1$, the stratum-specific natural indirect effect $NIE_1(t; \boldsymbol{x})$ is equal to*

$$\begin{aligned} &\int_0^t \{1 - \Pr(T < t|M = m, \boldsymbol{X} = \boldsymbol{x}, A = 1, U = 1)\} \\ &\times \left\{dF_{M|\boldsymbol{X}=\boldsymbol{x},A=1,U=1}(m) - dF_{M|\boldsymbol{X}=\boldsymbol{x},A=0,U=1}(m)\right\} \\ &+ \Pr(M \le t|\boldsymbol{X} = \boldsymbol{x}, A = 0, U = 1) - \Pr(M \le t|\boldsymbol{X} = \boldsymbol{x}, A = 1, U = 1), \end{aligned}$$

and the stratum-specific natural direct effect $NDE_1(t; \boldsymbol{x})$ is equal to

$$\begin{aligned} &\int_0^t \{\Pr(T < t|M = m, \boldsymbol{X} = \boldsymbol{x}, A = 0, U = 1) \\ &- \Pr(T < t|M = m, \boldsymbol{X} = \boldsymbol{x}, A = 1, U = 1)\} \, dF_{M|\boldsymbol{X}=\boldsymbol{x},A=0,U=1}(m). \end{aligned}$$

Under Assumptions 2 and 3, for the stratum with $U = 2$, the stratum-specific total effect $TE_2(t; \boldsymbol{x})$ is equal to

$$\Pr(T \ge t|A = 1, \boldsymbol{X} = \boldsymbol{x}, U = 2) - \Pr(T \ge t|A = 0, \boldsymbol{X} = \boldsymbol{x}, U = 2),$$

and for the stratum with $U = 3$, the stratum-specific total effect $TE_3(t; \boldsymbol{x})$ is equal to

$$\Pr(T \geq t | A = 1, \boldsymbol{X} = \boldsymbol{x}, U = 3) - \Pr(T \geq t | A = 0, \boldsymbol{X} = \boldsymbol{x}, U = 3).$$

The proof of Theorem 1 is given in Section S1.1 of the online Supplementary Material. Because U is unobserved, we cannot use Theorem 1 directly to identify the stratum-specific effects from the observed data. To do so, we require the following assumptions.

Assumption 4. *With probability one, U is conditional independent of A, given $\boldsymbol{X}$.*

Assumption 5. *With probability one,*

$$\begin{aligned}
&\Pr(M(0) = m | \boldsymbol{X} = \boldsymbol{x}, U = 2) = g_1 \left\{ \Pr(M(0) = m | \boldsymbol{X} = \boldsymbol{x}, U = 1); \boldsymbol{x} \right\}, \\
&\Pr(T(0) \geq t | M(0) = m, \boldsymbol{X} = \boldsymbol{x}, U = 2) \\
&= g_2 \left\{ \Pr(T(0) \geq t | M(0) = m, \boldsymbol{X} = \boldsymbol{x}, U = 1); \boldsymbol{x} \right\}, \\
&\Pr(T(1) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 2) = g_3 \left\{ \Pr(T(1) \geq t | \boldsymbol{X} = \boldsymbol{x}, U = 3); \boldsymbol{x} \right\},
\end{aligned}$$

for some known functions $g_k(\cdot; \boldsymbol{x})$ $(k = 1, 2, 3)$ that satisfy the conditions given in Section S1.2 *of the online Supplementary Material.*

Assumption 6. *(M, T, U) is conditionally independent of C, given A and $\boldsymbol{X}$, and the upper bound of the support of T is no larger than that of C.*

Assumption 4 requires that the stratum membership not be affected by the treatment assignment A, given the covariates $\boldsymbol{X}$. Assumption 5 requires a known relationship of stratum-specific event time distributions that guarantees the identification of the distributions. The assumption may be relaxed by assuming unknown g_k that can be parameterized and estimated from the observed data. For example, we may assume a proportional hazards models for the stratum-specific event times, as in Section 2.4. The first part of Assumption 6 is a standard assumption for noninformative censoring times. The second part of Assumption 6 is an extension of the independent censoring and sufficient follow-up assumption in Maller and Zhou (1992) for the nonparametric estimation of a cured proportion in censored data. The assumption on the upper bounds of the supports ensures that we can observe sufficient data to infer the tail behavior of the event times in order to identify the stratum membership. By further assuming Assumptions 4−6, we obtain the identification results in Theorem 2, the proof of which is given in Section S1.2 of the Supplementary Material.

Theorem 2. *Under Assumptions* 2–6*, the stratum-specific effects can be identified using the identification formulas given in Section* S1.2 *of the Supplementary Material.*

Theorem 2 gives the identification result based on nonparametric models for U, and for (M, T) given U. In particular, Assumption 4 requires several modeling assumptions. In practice, we may consider additional model assumptions for U and (M, T) to gain power in estimating the causal effects. In the next section, we extend the multistate modeling idea for semi-competing risks data to form such a model.

2.4. Modeling assumptions

One way to model semi-competing risks data is to use a multistate framework (Xu, Kalbfleisch and Tai (2010)). In a multistate analysis of semi-competing risks data, usually three states are involved, corresponding to healthy (state 1), illness (state 2), and death (state 3) in an illness-death model. All subjects start at state 1. A subject enters state 2 if he/she develops the intermediate event, and enters state 3 if he/she develops the primary event. In a traditional illness-death model for semi-competing risks data, three processes that move from one state to another are modeled: (1) healthy to illness, (2) illness to death, and (3) healthy to death.

Here, we extend the idea to model processes that move from one state to another in different strata, as defined in Section 2.2. Subjects with $U = 1$ and subjects with $U = 2$ receiving $A = 0$ involve the processes of healthy to illness and illness to death, and we model the time to the nonterminal event M and the residual time $R \equiv T - M$. We assume that M and R are conditionally independent given $A, \boldsymbol{X}$, and U. This serves two purposes: to obtain a tractable EM algorithm, and to avoid the problem of induced informative censoring caused by residual dependence between M and R (Wang and Wells (1998); Lin, Sun and Ying (1999)). Subjects with $U = 2$ receiving $A = 1$ and subjects with $U = 3$ involve the process of healthy to death. This proposed model is related to, but different from the illness-death model, in that subjects in different principal strata may experience a different transition structure in the proposed model.

Suppose that for a subject with $U = 1$, the nonterminal event time follows a proportional hazards model, with the hazard function given by

$$\lambda_M^{(1)}(t|A = a, \boldsymbol{X} = \boldsymbol{x}) = \lambda_1(t) \exp\left(\beta_{M1} a + \boldsymbol{\gamma}_{M1}^{\mathrm{T}} \boldsymbol{x}\right).$$

Furthermore, the gap time between the occurrences of the nonterminal and ter-

minal events, R, follows a proportional hazards model, with the hazard function given by

$$\lambda_R^{(1)}(r|A=a,\boldsymbol{X}=\boldsymbol{x})=\lambda_2(r)\exp\left(\beta_{R1}a+\boldsymbol{\gamma}_{R1}^{\mathrm{T}}\boldsymbol{x}\right).$$

Suppose that for a subject with $U=2$ and not exposed to treatment ($A=0$), the nonterminal event time follows a proportional hazards model, with the hazard function given by

$$\lambda_M^{(2)}(t|A=0,\boldsymbol{X}=\boldsymbol{x})=\lambda_1(t)\exp\left(\beta_{M2}+\boldsymbol{\gamma}_{M2}^{\mathrm{T}}\boldsymbol{x}\right),$$

and the gap time between the occurrences of the nonterminal and terminal events follows another proportional hazards model, with the hazard function given by

$$\lambda_R^{(2)}(r|A=0,\boldsymbol{X}=\boldsymbol{x})=\lambda_2(r)\exp\left(\beta_{R2}+\boldsymbol{\gamma}_{R2}^{\mathrm{T}}\boldsymbol{x}\right).$$

Here, subjects with $U=1$ and subjects with $U=2$ and not exposed to treatment share the same baseline hazard functions, although the hazard ratios for the covariates may be different. The parameters β_{M1} and β_{R1} are the log hazard ratios of the treatment on the nonterminal event time and gap time, respectively, for subjects with $U=1$; the parameters β_{M2} and β_{R2} are the log hazard ratios on the nonterminal event time and gap time, respectively, comparing subjects with $U=1$ and $U=2$, neither of whom were exposed to treatment, with a baseline covariate value $\boldsymbol{X}=\mathbf{0}$.

For subjects with $U=2$ who were exposed to treatment ($A=1$), we assume that the terminal event time follows a proportional hazards model, with the hazard function given by

$$\lambda_T^{(2)}(t|A=1,\boldsymbol{X}=\boldsymbol{x})=\lambda_3(t)\exp\left(\beta_{T2}+\boldsymbol{\gamma}_{T2}^{\mathrm{T}}\boldsymbol{x}\right).$$

For subjects with $U=3$, we suppose that the terminal event time follows a proportional hazards model, with the hazard function given by

$$\lambda_T^{(3)}(t|A=a,\boldsymbol{X}=\boldsymbol{x})=\lambda_3(t)\exp\left(\beta_{T3}a+\boldsymbol{\gamma}_{T3}^{\mathrm{T}}\boldsymbol{x}\right).$$

Note that the terminal event times for subjects with $U=3$ and subjects with $U=2$ who were exposed to treatment share the same baseline hazard function. The parameter β_{T3} is the log hazard ratio of treatment on the terminal event time for subjects with $U=3$, and β_{T2} is the log hazard ratio of the terminal event time comparing subjects with $U=3$ and $A=0$ with subjects with $U=2$ and $A=1$, with the same covariate value $\boldsymbol{X}=\mathbf{0}$.

Using the results in Theorem 1, we obtain the expression of the stratum-

specific effects by replacing the probabilities with the expressions under the proportional hazards models. The natural indirect and direct effects in the stratum with $U = 1$ can be presented as

$$\begin{aligned} NIE_1(t|\boldsymbol{X}=\boldsymbol{x}) &= \int_0^t \exp\left\{-\Lambda_2(t-m)e^{\beta_{R1}+\boldsymbol{\gamma}_{R1}^{\mathrm{T}}\boldsymbol{x}}\right\}\lambda_1(m)e^{\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}} \\ &\times\left[e^{\beta_{M1}}\exp\left\{-\Lambda_1(m)e^{\beta_{M1}+\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right\}-\exp\left\{-\Lambda_1(m)e^{\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right\}\right]dm \\ &+\exp\left\{-\Lambda_1(t)e^{\beta_{M1}+\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right\}-\exp\left\{-\Lambda_1(t)e^{\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right\} \end{aligned}$$

and

$$\begin{aligned} NDE_1(t|\boldsymbol{X}=\boldsymbol{x}) = \int_0^t &\left[\exp\left\{-\Lambda_2(t-m)e^{\beta_{R1}+\boldsymbol{\gamma}_{R1}^{\mathrm{T}}\boldsymbol{x}}\right\}-\exp\left\{-\Lambda_2(t-m)e^{\boldsymbol{\gamma}_{R1}^{\mathrm{T}}\boldsymbol{x}}\right\}\right] \\ &\times\lambda_1(m)e^{\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\exp\left\{-\Lambda_1(m)e^{\boldsymbol{\gamma}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right\}dm, \end{aligned}$$

respectively, where $\Lambda_1(t) = \int_0^t \lambda_1(s)ds$ and $\Lambda_2(t) = \int_0^t \lambda_2(s)ds$. The total effects in strata with $U = 2$ and $U = 3$ are given by

$$\begin{aligned} &TE_2(t|\boldsymbol{X}=\boldsymbol{x}) \\ &=\exp\left\{-\Lambda_3(t)e^{\beta_{T2}+\boldsymbol{\gamma}_{T2}^{\mathrm{T}}\boldsymbol{x}}\right\}-1+\int_0^t \lambda_1(m)e^{\beta_{M2}+\boldsymbol{\gamma}_{M2}^{\mathrm{T}}\boldsymbol{x}} \\ &\quad\times\exp\left\{-\Lambda_1(m)e^{\beta_{M2}+\boldsymbol{\gamma}_{M2}^{\mathrm{T}}\boldsymbol{x}}\right\}\left[1-\exp\left\{-\Lambda_2(t-m)e^{\beta_{R2}+\boldsymbol{\gamma}_{R2}^{\mathrm{T}}\boldsymbol{x}}\right\}\right]dm \end{aligned}$$

and

$$TE_3(t|\boldsymbol{X}=\boldsymbol{x}) = \exp\left\{-\Lambda_3(t)e^{\beta_{T3}+\boldsymbol{\gamma}_{T3}^{\mathrm{T}}\boldsymbol{x}}\right\}-\exp\left\{-\Lambda_3(t)e^{\boldsymbol{\gamma}_{T3}^{\mathrm{T}}\boldsymbol{x}}\right\},$$

respectively, where $\Lambda_3(t) = \int_0^t \lambda_3(s)ds$.

As in Yu et al. (2015), we consider a multinomial logistic regression model on the stratum membership. In particular, we assume

$$\begin{aligned} w_1(\boldsymbol{x};\boldsymbol{\alpha}) &= \Pr(U=1|\boldsymbol{X}=\boldsymbol{x}) = \frac{\exp\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)}{1+\exp\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)+\exp\left(\boldsymbol{\alpha}_2^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)}, \\ w_2(\boldsymbol{x};\boldsymbol{\alpha}) &= \Pr(U=2|\boldsymbol{X}=\boldsymbol{x}) = \frac{\exp\left(\boldsymbol{\alpha}_2^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)}{1+\exp\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)+\exp\left(\boldsymbol{\alpha}_2^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)}, \end{aligned}$$

and $w_3(\boldsymbol{x};\boldsymbol{\alpha}) = \Pr(U=3|\boldsymbol{X}=\boldsymbol{x}) = \{1+\exp\left(\boldsymbol{\alpha}_1^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)+\exp\left(\boldsymbol{\alpha}_2^{\mathrm{T}}\widetilde{\boldsymbol{x}}\right)\}^{-1}$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}})^{\mathrm{T}}$ and $\widetilde{\boldsymbol{x}} = (1, \boldsymbol{x}^{\mathrm{T}})^{\mathrm{T}}$. Then, the marginalized stratum-specific

natural indirect and direct effects are given by

$$NIE_1(t) = \frac{\int NIE_1(t|\boldsymbol{X}=\boldsymbol{x})w_1(\boldsymbol{x};\boldsymbol{\alpha})dF(\boldsymbol{x})}{\int w_1(\boldsymbol{x};\boldsymbol{\alpha})dF(\boldsymbol{x})}$$

and

$$NDE_1(t) = \frac{\int NDE_1(t|\boldsymbol{X}=\boldsymbol{x})w_1(\boldsymbol{x};\boldsymbol{\alpha})dF(\boldsymbol{x})}{\int w_1(\boldsymbol{x};\boldsymbol{\alpha})dF(\boldsymbol{x})},$$

respectively, where $F(\cdot)$ is the cumulative distribution function of $\boldsymbol{X}$.

2.5. Nonparametric maximum likelihood estimation

For a random sample of n subjects, the observed semi-competing risks data are given by $\mathcal{O} = \{\mathcal{O}_i : i = 1, \ldots, n\}$, where $\mathcal{O}_i = \{\Delta_i^M, Z_i, \Delta_i^T, Y_i, A_i, \boldsymbol{X}_i\}$. The likelihood function for the observed data can be constructed by laying out the conditional probabilities of the observed data, given possible stratum membership. Specifically, if $\Delta_i^M = 1$, then $U_i = 3$ with probability zero, and $U_i = 1$ with probability one if further $A_i = 1$. If $\Delta_i^M = 0$ and $\Delta_i^T = 1$, then $U_i = 1$ with probability zero, and $U_i = 3$ with probability one if further $A_i = 0$. If $\Delta_i^M = \Delta_i^T = 0$, then there are positive probabilities for $U_i = 1, 2, 3$. The likelihood function for the observed data $\mathcal{O}$ is then given by

$$\prod_{i=1}^{n} \widetilde{L}_{i1}(\mathcal{O}_i)^{\Delta_i^M} \left\{ \widetilde{L}_{i2}(\mathcal{O}_i)^{\Delta_i^T} \widetilde{L}_{i3}(\mathcal{O}_i)^{1-\Delta_i^T} \right\}^{1-\Delta_i^M},$$

where

$$\begin{aligned}
\widetilde{L}_{i1}(\mathcal{O}_i) =& \Pr\left(U_i = 1|\boldsymbol{X}_i\right) \Pr\left(Z_i, Y_i, \Delta_i^T | U_i = 1, \boldsymbol{X}_i, A_i\right) \\
&+ I\left(A_i = 0\right) \Pr\left(U_i = 2|\boldsymbol{X}_i\right) \Pr\left(Z_i, Y_i, \Delta_i^T | U_i = 2, \boldsymbol{X}_i, A_i = 0\right), \\
\widetilde{L}_{i2}(\mathcal{O}_i) =& \Pr\left(U_i = 3|\boldsymbol{X}_i\right) \Pr\left(Y_i, \Delta_i^T | U_i = 3, \boldsymbol{X}_i\right) \\
&+ I\left(A_i = 1\right) \Pr\left(U_i = 2|\boldsymbol{X}_i\right) \Pr\left(Y_i, \Delta_i^T | U_i = 2, \boldsymbol{X}_i, A_i = 1\right),
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{L}_{i3}(\mathcal{O}_i) =& \Pr\left(U_i = 1|\boldsymbol{X}_i\right) \Pr\left(Z_i, \Delta_i^M, Y_i, \Delta_i^T | U_i = 1, \boldsymbol{X}_i\right) \\
&+ \Pr\left(U_i = 2|\boldsymbol{X}_i\right) \left\{ I\left(A_i = 0\right) \Pr\left(Z_i, \Delta_i^M, Y_i, \Delta_i^T | U_i = 2, \boldsymbol{X}_i, A_i = 0\right) \right. \\
&\left. + I\left(A_i = 1\right) \Pr\left(Y_i, \Delta_i^T | U_i = 2, \boldsymbol{X}_i, A_i = 1\right) \right\} \\
&+ \Pr\left(U_i = 3|\boldsymbol{X}_i\right) \Pr\left(Y_i, \Delta_i^T | U_i = 3, \boldsymbol{X}_i\right).
\end{aligned}$$

We consider the nonparametric maximum likelihood estimation such that the estimators for Λ_1, Λ_2, and Λ_3 are step functions, as in Zeng and Lin (2007). In particular, let $0 < t_{11} < \cdots < t_{1m_1} < \infty$ be the ordered sequence of event times Z_i with $\Delta_i^M = 1$; let $0 < t_{21} < \cdots < t_{2m_2} < \infty$ be the ordered sequence of gap times $V_i \equiv Y_i - Z_i$ with $\Delta_i^M = \Delta_i^T = 1$; and let $0 < t_{31} < \cdots < t_{3m_3} < \infty$ be the ordered sequence of event times Y_i with $\Delta_i^M = 0$ and $\Delta_i^T = 1$. Let λ_{kl} be the jump size for Λ_k at t_{kl}, for $k = 1, 2, 3$ and $l = 1, \ldots, m_k$. Write $\boldsymbol{\eta}_{M1} = (\beta_{M1}, \boldsymbol{\gamma}_{M1})^{\mathrm{T}}$, $\boldsymbol{\eta}_{R1} = (\beta_{R1}, \boldsymbol{\gamma}_{R1})^{\mathrm{T}}$, $\boldsymbol{\eta}_{M2} = (\beta_{M2}, \boldsymbol{\gamma}_{M2})^{\mathrm{T}}$, $\boldsymbol{\eta}_{R2} = (\beta_{R2}, \boldsymbol{\gamma}_{R2})^{\mathrm{T}}$, $\boldsymbol{\eta}_{T2} = (\beta_{T2}, \boldsymbol{\gamma}_{T2})^{\mathrm{T}}$, $\boldsymbol{\eta}_{T3} = (\beta_{T3}, \boldsymbol{\gamma}_{T3})^{\mathrm{T}}$, $\boldsymbol{\theta} = (\boldsymbol{\eta}_{M1}^{\mathrm{T}}, \boldsymbol{\eta}_{R1}^{\mathrm{T}}, \boldsymbol{\eta}_{M2}^{\mathrm{T}}, \boldsymbol{\eta}_{R2}^{\mathrm{T}}, \boldsymbol{\eta}_{T2}^{\mathrm{T}}, \boldsymbol{\eta}_{T3}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$, and $\mathcal{A} = (\Lambda_1, \Lambda_2, \Lambda_3)^{\mathrm{T}}$. We maximize the objective function

$$L_n(\boldsymbol{\theta}, \mathcal{A}) = \prod_{i=1}^{n} L_{i1}(\boldsymbol{\theta}, \mathcal{A})^{\Delta_i^M} \left\{ L_{i2}(\boldsymbol{\theta}, \mathcal{A})^{\Delta_i^T} L_{i3}(\boldsymbol{\theta}, \mathcal{A})^{1-\Delta_i^T} \right\}^{1-\Delta_i^M},$$

where

$$\begin{aligned}
L_{i1}(\boldsymbol{\theta}, \mathcal{A}) &= w_1(\boldsymbol{X}_i; \boldsymbol{\alpha}) \Lambda_1\{Z_i\} e^{\boldsymbol{\eta}_{M1}^{\mathrm{T}} \boldsymbol{W}_i} \exp\left(-e^{\boldsymbol{\eta}_{M1}^{\mathrm{T}} \boldsymbol{W}_i} \sum_{t_{1l} \le Z_i} \lambda_{1l} \right) \\
&\quad \times \left(\Lambda_2\{V_i\} e^{\boldsymbol{\eta}_{R1}^{\mathrm{T}} \boldsymbol{W}_i} \right)^{\Delta_i^T} \exp\left(-e^{\boldsymbol{\eta}_{R1}^{\mathrm{T}} \boldsymbol{W}_i} \sum_{t_{2l} \le V_i} \lambda_{2l} \right) \\
&\quad + I(A_i = 0) w_2(\boldsymbol{X}_i; \boldsymbol{\alpha}) \Lambda_1\{Z_i\} e^{\boldsymbol{\eta}_{M2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \exp\left(-e^{\boldsymbol{\eta}_{M2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \sum_{t_{1l} \le Z_i} \lambda_{1l} \right) \\
&\quad \times \left(\Lambda_2\{V_i\} e^{\boldsymbol{\eta}_{R2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \right)^{\Delta_i^T} \exp\left(-e^{\boldsymbol{\eta}_{R2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \sum_{t_{2l} \le V_i} \lambda_{2l} \right), \\
L_{i2}(\boldsymbol{\theta}, \mathcal{A}) &= I(A_i = 1) w_2(\boldsymbol{X}_i; \boldsymbol{\alpha}) \left(\Lambda_3\{Y_i\} e^{\boldsymbol{\eta}_{T2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \right)^{\Delta_i^T} \exp\left(- \sum_{t_{3l} \le Y_i} \lambda_{3l} e^{\boldsymbol{\eta}_{T2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \right) \\
&\quad + w_3(\boldsymbol{X}_i; \boldsymbol{\alpha}) \left(\Lambda_3\{Y_i\} e^{\boldsymbol{\eta}_{T3}^{\mathrm{T}} \boldsymbol{W}_i} \right)^{\Delta_i^T} \exp\left(- \sum_{t_{3l} \le Y_i} \lambda_{3l} e^{\boldsymbol{\eta}_{T3}^{\mathrm{T}} \boldsymbol{W}_i} \right), \\
L_{i3}(\boldsymbol{\theta}, \mathcal{A}) &= L_{i2}(\boldsymbol{\eta}, \mathcal{A}) + w_1(\boldsymbol{X}_i; \boldsymbol{\alpha}) \exp\left(- \sum_{t_{1l} \le Z_i} \lambda_{1l} e^{\boldsymbol{\eta}_{M1}^{\mathrm{T}} \boldsymbol{W}_i} \right) \\
&\quad + I(A_i = 0) w_2(\boldsymbol{X}_i; \boldsymbol{\alpha}) \exp\left(- \sum_{t_{1l} \le Z_i} \lambda_{1l} e^{\boldsymbol{\eta}_{M2}^{\mathrm{T}} \widetilde{\boldsymbol{X}}_i} \right),
\end{aligned}$$

$\boldsymbol{W}_i = (A_i, \boldsymbol{X}_i^{\mathrm{T}})^{\mathrm{T}}$, and $\Lambda_k\{t\}$ is the jump size of Λ_k at time t for $k = 1, 2, 3$.

By treating U_i $(i = 1, \ldots, n)$ as missing data, we propose an EM algorithm to maximize this objective function. In the E-step of the EM algorithm, we evaluate the conditional expectation of the terms related to the latent variable U_i, which follows a multinomial distribution with probabilities corresponding to the observed-data likelihood, given different values of U_i. In the M-step, we update the parameter values using partial-score-structured estimating equations, Breslow-type estimators, and logistic-regression-type estimating equations. The details of the EM algorithm are given in Section S2 of the Supplementary Material. We write $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ as the estimators. The indirect and direct effects in the stratum with $U = 1$ can then be estimated by

$$\begin{aligned}\widehat{NIE}_1(t;\boldsymbol{x}) = \sum_{t_{1j}\le t}\Bigg[&\exp\left(-\sum_{t_{2k}\le t-t_{1j}}\widehat{\lambda}_{2k}e^{\widehat{\boldsymbol{\theta}}_{R1}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)\widehat{\lambda}_{1j}\\ &\times\left\{e^{\widehat{\boldsymbol{\theta}}_{M1}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\exp\left(-\sum_{k=1}^{j}\widehat{\lambda}_{1k}e^{\widehat{\boldsymbol{\theta}}_{M1}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)-e^{\widehat{\boldsymbol{\gamma}}_{M1}^{\mathrm{T}}\boldsymbol{x}}\exp\left(-\sum_{k=1}^{j}\widehat{\lambda}_{1k}e^{\widehat{\boldsymbol{\gamma}}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right)\right\}\Bigg]\\ &+\exp\left(-\sum_{t_{1j}\le t}\widehat{\lambda}_{1j}e^{\widehat{\boldsymbol{\theta}}_{M1}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)-\exp\left(-\sum_{t_{1j}\le t}\widehat{\lambda}_{1j}e^{\widehat{\boldsymbol{\gamma}}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right)\end{aligned}\tag{2.5}$$

and

$$\begin{aligned}&\widehat{NDE}_1(t;\boldsymbol{x})\\ &=\sum_{t_{1j}\le t}\Bigg[\left\{\exp\left(-\sum_{t_{2k}\le t-t_{1j}}\widehat{\lambda}_{2k}e^{\widehat{\boldsymbol{\theta}}_{R1}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)-\exp\left(-\sum_{t_{2k}\le t-t_{1j}}\widehat{\lambda}_{2k}e^{\widehat{\boldsymbol{\gamma}}_{R1}^{\mathrm{T}}\boldsymbol{x}}\right)\right\}\\ &\quad\times\widehat{\lambda}_{1j}e^{\widehat{\boldsymbol{\gamma}}_{M1}^{\mathrm{T}}\boldsymbol{x}}\exp\left(-\sum_{k=1}^{j}\widehat{\lambda}_{1k}e^{\widehat{\boldsymbol{\gamma}}_{M1}^{\mathrm{T}}\boldsymbol{x}}\right)\Bigg].\end{aligned}\tag{2.6}$$

The total effects in strata with $U = 2$ and $U = 3$ can be estimated by

$$\begin{aligned}\widehat{TE}_2(t;\boldsymbol{x}) &= \exp\left(-\sum_{t_{3j}\le t}\widehat{\lambda}_{3j}e^{\widehat{\boldsymbol{\theta}}_{T2}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)-1\\ &+\sum_{t_{1j}\le t}\left[\widehat{\lambda}_{1j}e^{\widehat{\boldsymbol{\theta}}_{M2}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\exp\left(-\sum_{k=1}^{j}\widehat{\lambda}_{1k}e^{\widehat{\boldsymbol{\theta}}_{M2}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)\left\{1-\exp\left(-\sum_{t_{2k}\le t-t_{1j}}\widehat{\lambda}_{2k}e^{\widehat{\boldsymbol{\theta}}_{R2}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right)\right\}\right]\end{aligned}\tag{2.7}$$

and

$$\widehat{TE}_3(t;\boldsymbol{x}) = \exp\left(-\sum_{t_{3j}\le t}\widehat{\lambda}_{3j}e^{\widehat{\boldsymbol{\theta}}_{T3}^{\mathrm{T}}\widetilde{\boldsymbol{x}}}\right) - \exp\left(-\sum_{t_{3j}\le t}\widehat{\lambda}_{3j}e^{\widehat{\boldsymbol{\gamma}}_{T3}^{\mathrm{T}}\boldsymbol{x}}\right). \tag{2.8}$$

The marginalized stratum-specific indirect and direct effects in the stratum with $U = 1$ can be estimated by

$$\widehat{NIE}_1(t) = \frac{\sum_{i=1}^n w_1(\boldsymbol{X}_i;\widehat{\boldsymbol{\alpha}})\widehat{NIE}_1(t;\boldsymbol{X}_i)}{\sum_{i=1}^n w_1(\boldsymbol{X}_i;\widehat{\boldsymbol{\alpha}})} \tag{2.9}$$

and

$$\widehat{NDE}_1(t) = \frac{\sum_{i=1}^n w_1(\boldsymbol{X}_i;\widehat{\boldsymbol{\alpha}})\widehat{NDE}_1(t;\boldsymbol{X}_i)}{\sum_{i=1}^n w_1(\boldsymbol{X}_i;\widehat{\boldsymbol{\alpha}})}, \tag{2.10}$$

respectively.

2.6. Asymptotic properties

We study the asymptotic properties of the estimators under the semiparametric model in Section 2.4. Under suitable regularity conditions, the estimators $(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{A}})$ have the usual large-sample properties, including consistency and asymptotic normality, as given in Theorem 3 below. Let $\boldsymbol{\theta}_0$, Λ_{10}, Λ_{20}, and Λ_{30} be the true values of $\boldsymbol{\theta}$, Λ_1, Λ_2, and Λ_3, respectively, $\|\cdot\|$ be the Euclidean norm, and τ_k be the upper limit of the support of $\widehat{\Lambda}_k$, for $k = 1, 2, 3$.

Theorem 3. *Under Conditions* 1–5 *in Section* S3 *of the online Supplementary Material,*

$$\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\| + \sum_{k=1}^{3}\sup_{t\in[0,\tau_k]}\left|\widehat{\Lambda}_k(t) - \Lambda_{k0}(t)\right|$$

converges to zero almost surely. In addition, $\sqrt{n}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{\Lambda}_1(\cdot) - \Lambda_{10}(\cdot), \widehat{\Lambda}_2(\cdot) - \Lambda_{20}(\cdot), \widehat{\Lambda}_3(\cdot) - \Lambda_{30}(\cdot)\}$ *converges weakly to a zero-mean Gaussian process in the Banach space* $\mathbb{R}^m \times l^\infty(\mathcal{A}_1) \times l^\infty(\mathcal{A}_2) \times l^\infty(\mathcal{A}_3)$*, where* m *is the dimension of* $\boldsymbol{\theta}$ *and* $\mathcal{A}_k$ *is the unit ball in the space of functions on* $[0, \tau_k]$ *with bounded variation, for* $k = 1, 2, 3$*. The limiting covariance matrix of* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *attains the semiparametric efficiency bound.*

Theorem 4. *Under Conditions* $1 - 5$ *in Section* S3 *of the online Supplementary Material, the estimators for the stratum-specific effects given in* (2.5)–(2.10) *are consistent and asymptotically normal.*

The proofs of Theorems 3 and 4 are given in Section S3 of the online Supplementary Material. Because the form of the limiting variances of the stratum-specific effects is complicated, we estimate the variance of the estimators using a nonparametric bootstrap procedure in our numerical studies.

3. Simulation Studies

We conducted simulation studies to examine the performance of the proposed methods. We generated two covariates, $X_1 \sim N(0,1)$ and $X_2 \sim Unif(0,1)$, and generated the treatment indicator $A \sim Bin(0.5)$ to reflect 1:1 randomization. We set $\Lambda_1(t) = t$, $\Lambda_2(t) = 0.2t$, and $\Lambda_3(t) = \log(1+t)$; the true values of the other parameters are shown in Tables S1 and S2 of the online Supplementary Material. We generated a censoring time $C \sim Unif(0,15)$ to obtain approximately 51% and 26% censoring rates for the nonterminal and terminal events, respectively. The proportions of subjects with $U = 1, 2, 3$ are approximately 31%, 41%, and 28%, respectively.

We considered 1,000 replicates with sample sizes $n = 1000$ and 2000, where 100 bootstrap samples were used for the variance estimation. All replications that we examined converge with a 10^{-6} convergence criterion. The results for the parameter estimators are shown in Tables S1 and S2 and Figure S1 of Section S4.1 of the online Supplementary Material. The parameter estimators are virtually unbiased, and the bootstrap variance estimators become more accurate as the sample size increases.

Table 1 shows the performance of the estimated stratum-specific indirect and direct effects in a stratum with $U = 1$ and $\boldsymbol{X} = (0.5, 0.5)^{\mathrm{T}}$, as well as the estimated total effects for strata with $U = 2, 3$ and the same covariate values. Similarly, for any t, the average is taken over all replicates with estimators that have a last jump time of no less than t. The bias gets smaller as the sample size increases. The variance estimator is accurate and the coverage probability is close to the nominal level when the sample size is large.

In Section S4.2 of the online Supplementary Material, we evaluate the performance of the proposed approach when the baseline hazard functions differ by strata. Specifically, we change the baseline hazard functions for the event times with $U = 2$. Even though the total effects in the strata with $U = 2$ and $U = 3$ may be estimated with bias, the bias for the mediation effects in the stratum with $U = 1$ is relatively small, and the coverage probabilities for the 95% confidence intervals are close to the nominal level.

Table 1. Simulation results for the stratum-specific mediation effects and total effects.

			$n = 1000$				$n = 2000$			
	t	True Value	Bias	SE	SEE	CP	Bias	SE	SEE	CP
NDE_1	2	−0.11	0.014	0.056	0.070	0.97	0.007	0.031	0.041	0.97
	4	−0.17	0.018	0.090	0.106	0.96	0.010	0.052	0.066	0.97
	6	−0.18	0.020	0.096	0.107	0.94	0.010	0.057	0.069	0.96
NIE_1	2	−0.04	−0.001	0.022	0.025	0.97	0.000	0.015	0.017	0.97
	4	−0.03	−0.001	0.016	0.018	0.97	−0.001	0.011	0.012	0.96
	6	−0.02	−0.001	0.010	0.011	0.97	0.000	0.007	0.007	0.96
TE_2	2	−0.10	−0.036	0.158	0.175	0.97	−0.022	0.115	0.126	0.97
	4	0.10	−0.046	0.159	0.180	0.96	−0.026	0.113	0.127	0.97
	6	0.17	−0.047	0.136	0.156	0.97	−0.027	0.097	0.109	0.97
	8	0.18	−0.044	0.115	0.131	0.96	−0.025	0.084	0.093	0.96
TE_3	2	−0.07	0.021	0.139	0.147	0.97	0.015	0.117	0.114	0.95
	4	−0.06	0.031	0.119	0.126	0.98	0.022	0.101	0.099	0.96
	6	−0.06	0.033	0.101	0.107	0.97	0.024	0.086	0.085	0.96
	8	−0.05	0.033	0.088	0.093	0.97	0.024	0.075	0.074	0.96

4. Application

In this section, we apply the proposed methods to data from a prostate cancer clinical trial. NCIC Clinical Trials Group PR.3/Medical Research Council PR07/Intergroup T94-0110 is a randomized controlled trial of patients with locally advanced prostate cancer. The primary objective is to determine whether adding radiotherapy (RT) to androgen-deprivation therapy (ADT) prolongs overall survival, defined as the time from random assignment to death from any cause. The study recruited and randomly assigned 1,205 patients with locally advanced prostate cancer between 1995 and 2005, 602 to ADT alone and 603 to ADT + RT. The final report of the study (Mason et al. (2015)) stated that, at a median follow-up time of eight years, 465 patients had died. In addition, overall survival was significantly improved in patients allocated to ADT + RT (hazard ratio 0.70 with 95% CI, 0.57 to 0.85; P<0.001).

In addition to the primary outcome of death, the study also collected data on time to disease progression, which was defined as the first of any of the following events: biochemical progression, local progression, or development of metastatic disease. We analyzed the data to reveal the proportions of the treatment effect on overall survival that are mediated by disease progression. In particular, we adjusted for initial PSA level (< 20 vs. 20 to 50, vs. >50g/L) and Gleason score (8 vs. 8 to 10).

We analyzed the data using the proposed approach, with 1,000 bootstrap samples for the variance estimation. The parameter estimates for the regression

Table 2. Parameter estimates for the regression coefficients of the event times.

Process	$U = 1$					
	Health → Disease			Disease → Death		
	Est	SEE	p-value	Est	SEE	p-value
ADT + RT	−0.825	0.925	0.373	0.460	0.569	0.419
Initial PSA Level (20 to 50 g/L)	0.321	0.563	0.569	−0.097	0.322	0.763
Initial PSA Level (> 50 g/L)	1.607	0.614	0.009	0.065	0.337	0.846
Gleason Score (8-10)	−2.008	0.478	< 0.0001	−0.378	0.274	0.167
Process	$U = 2$, ADT					
	Health → Disease			Disease → Death		
	Est	SEE	p-value	Est	SEE	p-value
Intercept	−1.917	1.534	0.211	−0.557	2.490	0.823
Initial PSA Level (20 to 50 g/L)	0.663	0.985	0.501	−0.304	2.007	0.880
Initial PSA Level (> 50 g/L)	1.619	0.967	0.094	−0.104	1.693	0.951
Gleason Score (8-10)	0.674	0.932	0.469	−0.106	1.638	0.948
Process	$U = 2$, ADT + DT			$U = 3$		
	Health → Death			Health → Death		
	Est	SEE	p-value	Est	SEE	p-value
Intercept	−3.212	7.225	0.657	−0.446	0.827	0.590
Initial PSA Level (20 to 50 g/L)	1.146	4.886	0.815	−0.022	0.602	0.970
Initial PSA Level (> 50 g/L)	1.601	4.937	0.746	−0.883	0.641	0.168
Gleason Score (8-10)	1.624	5.266	0.758	−0.514	0.484	0.289

coefficients for the event time processes are shown in Table 2. For the stratum with $U = 1$, ADT + RT is associated with a decreased risk of disease progression, but is associated with an increased risk from disease progression to death. For the stratum with $U = 3$, ADT + RT is associated with a decreased risk of death. The effects are not significant at the 0.05 level. For the stratum with $U = 1$, a subject with an initial PSA level > 50 g/L is associated with a significantly increased risk of disease progression, compared with a similar subject with an initial PSA level < 20 g/L. Furthermore, a subject with a Gleason score $8-10$ is associated with a significantly decreased risk of disease progression, compared with a similar subject with a Gleason score < 8.

Table 3 shows the parameter estimators of the logistic regression model for the stratum membership. By averaging over the stratum membership probabilities over all subjects, given their covariate values, the average probabilities of belonging to strata $U = 1, 2$, and 3 are 40.1%, 25.7%, and 34.2%, respectively. To verify that the model is reasonable, we estimated the stratum-specific survival functions for every subject, and summarized the subject-specific survival function by weighting them by his/her stratum membership probabilities. We average

Table 3. Parameter estimates for the regression coefficients of the stratum membership

	$\boldsymbol{\alpha}_1$			$\boldsymbol{\alpha}_2$		
	Est	SEE	p-value	Est	SEE	p-value
Intercept	0.205	0.433	0.636	0.291	0.611	0.634
Initial PSA Level (20 to 50 g/L)	0.090	0.562	0.872	0.625	0.898	0.486
Initial PSA Level (> 50 g/L)	−0.684	0.541	0.206	0.004	0.816	0.996
Gleason Score (8-10)	0.134	0.408	0.742	−1.495	0.771	0.053

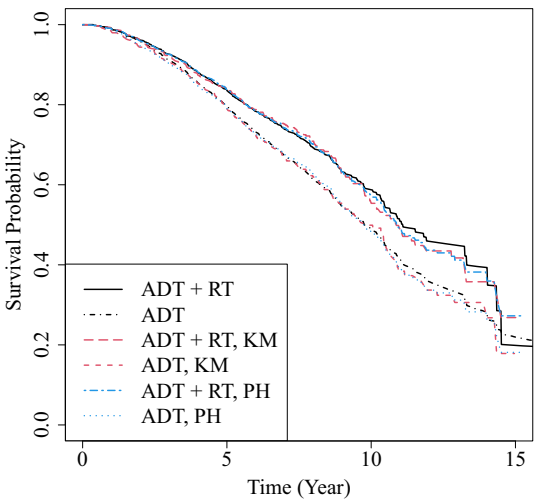

Figure 1. Estimated survival functions from the proposed, Kaplan-Meier, and proportional hazards model approaches.

the estimated survival functions for subjects assigned to ADT+RT versus ADT, and plot them against the survival function estimators from the Kaplan-Meier methods and the proportional hazards model. The results are shown in Figure 1. The estimated population-average survival functions for the ADT+RT and ADT groups are similar to those from the Kaplan-Meier methods and the proportional hazards model, especially up until 10 years, when the data are not sparse, indicating a proper fit of the proposed approach.

Figure 2 shows the estimated marginalized stratum-specific indirect and direct effects (with 95% confidence intervals) for the stratum with $U = 1$. The estimated natural indirect effect is positive and increasing over time, and the estimated natural direct effect is slightly negative over time. However, the 95% confidence intervals are wide, such that the stratum-specific natural indirect and direct effects are not significantly different from zero. The total effect in the

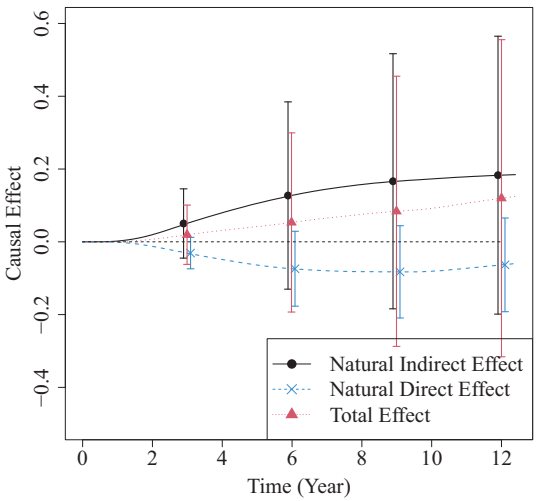

Figure 2. Estimated stratum-specific indirect and direct effects in the stratum with $U = 1$.

stratum with $U = 1$ is positive and increasing over time, corresponding to an increased survival probability assigned to ADT+RT versus ADT in the stratum with $U = 1$.

5. Discussion

Semi-competing risks data are frequently observed in medical studies, where the terminal event time may censor the intermediate event time, but not vice versa. To define and estimate the causal contrasts of the effect of a treatment on the terminal and intermediate events, we have introduced a novel principal stratification framework that distinguishes between susceptible and nonsusceptible subjects, given different treatments, and defined the natural indirect and direct effects in the stratum where the times to the intermediate and terminal events are well defined, given both treatments. We have provided reasonable assumptions to identify the stratum-specific natural indirect and direct effects, proposed a semiparametric model, and presented an EM algorithm to obtain the nonparametric maximum likelihood estimators of the model parameters. We have shown that the estimators are consistent and asymptotically efficient estimated under mild regularity conditions, and perform satisfactorily in finite-sample numerical studies.

In identifying the stratum-specific natural indirect and direct effects, we assumed that there are no subjects who are susceptible to the intermediate event

under the treatment ($A = 1$) and nonsusceptible under the control ($A = 0$). This assumption may need careful examination based on a scientific understanding of how the treatment may affect the intermediate event. In our data application, we assessed this assumption by fitting the proposed model with switched treatment indicator labels of ADT+RT and ADT. The estimated probability of belonging to the stratum with $U = 2$ (equivalent to the fourth stratum in the original labeling) is very low (6.9%), suggesting that the assumption of the nonexistence of the fourth stratum may be valid. In Section S4.3 of the online Supplementary Material, we include a sensitivity analysis that assesses the performance of the estimator for the stratum-specific effects when there is a fourth stratum with a small probability ($\sim$6.9%). Even though the assumption on the nonexistence of the fourth stratum fails, the stratum-specific effects in the first three strata can still be estimated with relatively small bias. In some applications, this fourth stratum may indeed exist. In the literature on principal stratification for uncensored data with four or more strata, the effect of interest often can only be interval identified. Interval identification with a regression model may result in a complicated solution manifold, with properties that are not well understood. We plan to explore this problem in future research.

The proposed nonparametric maximum likelihood estimation framework relies on the validity of the modeling assumptions, including the proportional hazards assumption on the hazard functions. Even though model checking for a mixture model for right-censored data has been considered in the literature (e.g., Peng and Taylor (2017)), model checking for semi-competing risks data may not be available for the illness-death model. A model checking procedure on the modeling assumptions may not be trivial and requires further research.

Supplementary Material

The online Supplementary Material includes proofs of Theorems 1 and 2 in Section 2.3, details of the EM algorithm in Section 2.5, other proofs of asymptotic results in Section 2.6, and additional simulation results in Sections 3 and 5.

Acknowledgments

This manuscript was prepared using data from Dataset NCT00002633-D1 from the NCTN Data Archive of the National Cancer Institute (NCI) National Clinical Trials Network (NCTN). Data were originally collected from clinical trial NCT00002633 Phase III Randomized Trial Comparing Total Androgen Blockade Versus Total Androgen Blockade Plus Pelvic Irradiation in Clinical Stage T3-

4, N0, M0 Adenocarcinoma of the Prostate. All analyses and conclusions in this manuscript are the sole responsibility of the authors and do not necessarily reflect the opinions or views of the clinical trial investigators, the NCTN, or the NCI. The authors were partially funded by the U.S. National Institutes of Health grants R01HL122212 and U01AG016976 and U.S. National Science Foundation grant DMS 1711952.

References

Aalen, O. O., Stensrud, M. J., Didelez, V., Daniel, R., Røysland, K. and Strohmaier, S. (2020). Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. *Biom. J.* **62**, 532–549.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91**, 444–455.

Comment, L., Mealli, F., Haneuse, S. and Zigler, C. (2019). Survivor average causal effects for continuous time: A principal stratification approach to causal inference with semicompeting risks. *arXiv preprint arXiv:1902.09304*.

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Szer, J., Klein, J. P. et al. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with BuCy2. *Blood* **78**, 838–843.

Didelez, V. (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime Data Anal.* **25**, 593–610.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Huang, Y.-T. (2021). Causal mediation of semicompeting risks. *Biometrics* **77**, 1143–1154.

Imai, K., Keele, L. and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25**, 51–71.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York.

Lange, T. and Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology* **22**, 575–581.

Lange, T., Vansteelandt, S. and Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *Am. J. Epidemiol.* **176**, 190–195.

Lin, D., Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika* **86**, 59–70.

Lin, S.-H., Young, J. G., Logan, R. and VanderWeele, T. J. (2017). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Stat. Med.* **36**, 4153–4166.

Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. *Biometrika* **79**, 731–739.

Mason, M. D., Parulekar, W. R., Sydes, M. R., Brundage, M., Kirkbride, P., Gospodarowicz, M. et al. (2015). Final report of the intergroup randomized study of combined androgen-deprivation therapy plus radiotherapy versus androgen-deprivation therapy alone in locally

advanced prostate cancer. *J. Clin. Oncol.* **33**, 2143–2150.

Nevo, D. and Gorfine, M. (2022). Causal inference for semi-competing risks data. *Biostatistics* **23**, 1115–1132.

Peng, Y. and Taylor, J. M. (2017). Residual-based model diagnosis methods for mixture cure models. *Biometrics* **73**, 495–505.

Tchetgen Tchetgen, E. J. (2011). On causal mediation analysis with a survival outcome. *Int. J. Biostat.* **7**, 1–38.

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology.* **22**, 582–585.

Vansteelandt, S., Linder, M., Vandenberghe, S., Steen, J. and Madsen, J. (2019). Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. *Stat. Med.* **38**, 4828–4840.

Wang, W. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561–572.

Xu, J., Kalbfleisch, J. D. and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics* **66**, 716–725.

Xu, Y., Scharfstein, D., Müller, P. and Daniels, M. (2022). A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks. *Biostatistics* **23**, 34-–49.

Yu, W., Chen, K., Sobel, M. E. and Ying, Z. (2015). Semiparametric transformation models for causal inference in time-to-event studies with all-or-nothing compliance. *J. R. Stat. Soc. Series B (Stat. Methodol.)* **77**, 397–415.

Zeng, D. and Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Series B (Stat. Methodol.)* **69**, 507–564.

Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J. Educ. Behav. Stat.* **28**, 353–368.

Zheng, W. and van der Laan, M. (2017). Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *J. Causal Inference* **5**, 1–24.

Fei Gao

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA.

E-mail: fgao@fredhutch.org

Fan Xia

Department of Epidemiology & Biostatistics, University of California at San Francisco, San Francisco, CA 94158, USA.

E-mail: Fan.Xia@ucsf.edu

K. C. G. Chan

Department of Biostatistics, University of Washington, Seattle, WA 98195, USA.

E-mail: kcgchan@u.washington.edu

(Received May 2021; accepted February 2022)